Voxel-based diktiometry: Combining convolutional neural networks with voxel-based analysis and its application in diffusion tensor imaging for Parkinson's disease

Alfonso Estudillo-Romero, Claire Haegelen, Pierre Jannin, John S H Baxter

Received: 23 December 2021 | Revised: 10 June 2022 | Accepted: 22 June 2022

DOI: 10.1002/hbm.26009

RESEARCH ARTICLE

WILEY

Voxel-based diktiometry: Combining convolutional neural networks with voxel-based analysis and its application in diffusion tensor imaging for Parkinson's disease

Alfonso Estudillo-Romero[1] | Claire Haegelen[1,2] | Pierre Jannin[1] | John S. H. Baxter[1]

[1]LTSI-INSERM UMR 1099, Université de Rennes 1, Rennes, France

[2]Département de Neurochirurgie, CHU Rennes, Rennes, France

Correspondence
John S. H. Baxter, LTSI-INSERM UMR 1099, Université de Rennes 1, 2 Avenue du Pr. Léon Bernard, Rennes, 35500 Rennes, France.
Email: jbaxter@univ-rennes1.fr

Funding information
Fondation Recherche Médicale, Grant/Award Number: DIC20161236441; Institut national de la santé et de la recherche médicale (INSERM); Institut des Neurosciences Cliniques de Rennes (INCR); SAD Région Bretagne

Abstract

Extracting population-wise information from medical images, specifically in the neurological domain, is crucial to better understanding disease processes and progression. This is frequently done in a whole-brain voxel-wise manner, in which a population of patients and healthy controls are registered to a common co-ordinate space and a statistical test is performed on the distribution of image intensities for each location. Although this method has yielded a number of scientific insights, it is further from clinical applicability as the differences are often small and altogether do not permit for a high-performing classifier. In this article, we take the opposite approach of using a high-performing classifier, specifically a traditional convolutional neural network, and then extracting insights from it which can be applied in a population-wise manner, a method we call *voxel-based diktiometry*. We have applied this method to diffusion tensor imaging (DTI) analysis for Parkinson's disease (PD), using the Parkinson's Progression Markers Initiative database. By using the network sensitivity information, we can decompose what elements of the DTI contribute the most to the network's performance, drawing conclusions about diffusion biomarkers for PD that are based on metrics which are not readily expressed in the voxel-wise approach.

KEYWORDS
convolutional neural networks, diffusion tensor imaging, Parkinson's disease, whole-brain voxel-based analysis

1 | INTRODUCTION

Deep convolutional neural networks (CNNs) have become a key element of modern medical image analysis. Traditional versions of the CNNs used for classification involve a series of linear convolutional layers intermixed with nonlinear activation and pooling layers. The convolutional layers act as simple image-processing operators, identifying particular features, with the nonlinear activation and pooling

layers providing both a source of nonlinearity at increasingly abstract and coarsely resolved images. At the most abstract level, the features identified may no longer be spatially localised, but encode some information about the content of the image as a whole, which are then used for classification. With the depth and complexity of these networks, it is often difficult to understand and communicate the network's processing and the "black-box" nature of the network has posed a number of issues for clinical acceptance and integration (F. Wang et al., 2020).

On the other hand, traditional methods of population-wide whole-brain voxel-based analysis (VBA) such as voxel-based morphometry (VBM) and voxel-based relaxometry (VBR) have become increasingly well-understood and validated in the neuroimaging domain. Arguably, the defining feature of these voxel-based population-wide analysis is their conceptual simplicity: a population is imaged and those images are deformably registered together into some common template space in which the quantitative intensity (in the case of VBR) or a derived characteristic (such as local scaling in the case of VBM) is used as a univariate distribution upon which one can directly and robustness measure the difference between two subpopulations, normally patients against healthy controls, or to correlate with a different clinically interesting variable derived from the patient's symptoms.

The issue with this approach also arises from its simplicity; it is designed to measure the correlation of a singular area with the clinical variable of interest, not measuring correlations *between* regions that may be of interest. That is, it only identifies regions that simply and strongly correlate with a clinical variable or sub-population, missing regions in which this correlation is weaker or conditioned on some other image feature. The second issue is that these voxel-based methods do not immediately provide a strong prospective method that makes use of their analysis, for example classifying new patient into a sub-population. The individual voxels on their own tend to offer relatively weak classifiers on their own as the statistical analysis only suggests they are better than chance, a relatively low bar for modern classification performance. Recently approaches have used VBA as a method for selecting features to use in machine learning classification, notably support vector machines (Chen et al., 2020; Prasuhn et al., 2020), which have had variable performance across different disease groups, but illustrate how additional, stronger classifiers would need to be appended to VBA in order to be clinically useful.

Broadly speaking, the general method underlying VBA methods is to use registration and statistical methods to identify potentially discriminative features of a disease; the diagnostic use of these features (including machine learning approaches) is applied afterwards. Depending on how this analysis is performed, larger or smaller regions of interest (ROI) can be used, but they are pre-specified, rather than determined empirically by classification utility. One possible approach to alleviate these issues is to invert the paradigm by starting with the creation of strong classifiers popularised by deep learning, and use population-wide analysis to understand how the patient images affect the outcome of these classifiers.

The contribution of this article is to use these techniques for voxel-based population-wide analysis to traditional CNNs for image classification, specifically the classification between Parkinsonian patients and healthy controls using solely diffusion tensor imaging (DTI). We call this method *voxel-based diktiometry* (VBD). Our goal is to show that traditional CNNs are sensitive to specific characteristic features of diffusion tensors in a nonlocal manner.

2 | THEORY AND PRIOR WORK

2.1 | Diffusion tensor imaging in Parkinson's disease

The DTI-based analysis of the brain white matter is a noninvasive imaging approach that has been widely used to measure the diffusivity of the water in the different tissues of the brain, thus allowing the characterisation of the integrity of the tissues associated with normal or abnormal diffusivity and anisotropy values.

In the context of the Parkinson's disease (PD), VBM, VBR, fixel-based analysis, and tractographic analysis (Cousineau et al., 2017; Y. Li et al., 2020; Xiao et al., 2021) are the more common approaches. Given the dimensionality of the data under analysis, some approaches have opted to perform the analysis on particular ROI, to obtain specific fibre bundles, for example (Wasserthal et al., 2019) and some of them have additionally identified the need to also perform correlations between the ROIs (Schuff et al., 2015). Fixel-based analysis follows a similar approach to other VBA methods and thus only a white-matter mask is used to limit the regions under investigation (Y. Li et al., 2020; Xiao et al., 2021).

On the PPMI data set (Marek et al., 2018) significant alterations between healthy controls and Parkinsonian patients located within the SN, the striatum and the subthalamic nucleus (STN), pallidum, putamen and thalamus have been previously reported (Cousineau et al., 2017; Schuff et al., 2015). (Xiao et al., 2021) have also found significant differences in the major white matter bundles specifically on the side of PD onset.

Although the simplest interpretation of PD is that it affects the dopaminergic components of the basal ganglia, the PD is a multisystem disorder involving several other neurotransmitters and pathways (Zhang & Burock, 2020). The distribution of abnormal changes in the DTI values not only at the early stages of the disease but also as possible consequences of neuroplasticity suggests the need to consider a model capable of taking advantage of the heterogeneity of the disease and find these complex correlations on non-predefined nonlocal regions.

2.2 | Salience in convolutional neural networks

CNNs have become a key tool in computer vision. Although originally considered to suffer from the "black-box" problem, where the reasoning of the machine learning tool is difficult or impossible to explain for any given case, CNNs have benefited from a large and early degree of attention towards their visualisation and explanation thanks in part to Simonyan's "saliency maps" (Simonyan et al., 2013) which led to the

development of more advanced methods in explaining the reasoning of CNNs, sometimes in creative ways, such as with DeepDream (Mordvintsev et al., 2015). These methods rely on propagating the gradient normally used to update the model weights into the input image, either directly visualising it (Simonyan et al., 2013) or using it to modify the underlying image (Mordvintsev et al., 2015). These have the benefit of being simple to implement as well as having an intuitive relationship to the notion of *sensitivity analysis* as the salience maps could be interpreted as the sensitivity of the network output towards a particular region in the image.

3 | METHODS

3.1 | Patient images

A total of 213 age-matched individuals with diffusion-weighted images were collected from the Parkinson's Progression Markers Initiative (PPMI) database. The PPMI data set contains PD patients with a diagnosis determined using clinical diagnostic criteria, requiring either at least two of resting tremor, bradykinesia and rigidity, or a single asymmetric resting tremor or asymmetric bradykinesia. Additionally, subjects were PD eligible if dopamine transporter (DAT) imaging demonstrated a dopaminergic deficit consistent with PD. Healthy control (HC) subjects were constrained to not display any of these symptoms or that of another clinically significant neurological disorder, had a Montreal Cognitive Assessment (MoCA) total score greater than 26, and had no first-degree family member with PD. (Marek et al., 2018) The age distribution for the two groups is shown in Figure 1. The data acquisition was conducted during three to four consecutive years within the majority of the PD patients whereas only during two consecutive years for the HC. The data is comprised of two diffusion-weighting (DWI) samples and one T1-weighted sample per session per year. By assuming a major progression of the disease on the PD group at the last screening, we have selected only the session from the last year for the 139 PD and all the sessions for the 74 HC. The clinical and demographic information as well as the total number of samples for each group is summarised in Table 1.

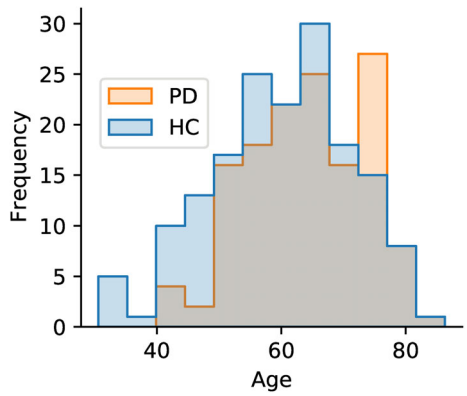

FIGURE 1 Age frequency distributions of the last screening session for the Parkinson's disease (PD) and the two sessions for the healthy control (HC)

The image acquisition protocol included a 3D magnetization prepared rapid gradient echo (MPRAGE) sequence for mapping anatomical details (repetition time [TR]/echo time [TE]/inversion time [TI] = 2300/3/900 ms; 1 mm isotropic resolution; twofold acceleration; sagittal-oblique angulation) and a cardiac-gated 2D single-shot echo-planar DTI sequence (TE = 88 ms, 2 mm isotropic resolution; 72 contiguous slices each 2 mm thick, twofold acceleration, axial-oblique aligned along the anterior–posterior commissure) with DWI gradients along 64 sensitization directions and a *b*-value of 1000 s/mm^2. TR was in the order of 8400–8800 ms, depending on the subject's heart rate (Marek et al., 2018; Schuff et al., 2015). MRI data were downloaded from the PPMI site https://ida.loni.usc.edu/ in DICOM format and converted to NIfTI format using the dcm2niix tool (X. Li et al., 2016).

3.2 | Preprocessing

The ROBEX (Iglesias et al., 2011) brain extraction tool was used to extract the brain mask before any other preprocessing step for the T1-weighted (T1w) images. The mask was subsequently used during the co-registration step. Noise removal from the T1w image was performed using the nonlocal means algorithm from the Dipy package (Descoteaux et al., 2008) followed by a bias field correction using the

TABLE 1 Demographic and clinical information about the cohorts used

	PD (N = 139)	HC (N = 74)
DTI samples	269	291
Sex *N* (%)		
F	49 (35.25%)	26 (35.14%)
M	90 (64.75%)	48 (64.86%)
Age		
Mean	64.20	60.31
(min, max)	(40, 86)	(31, 83)
MDS-UPDRS total (part III)		
Mean	25.10	0.63
(min, max)	(1, 80)	(0, 8)
Hoehn and Yahr *N* (%)		
Stage 0	1 (0.72%)	73 (98.65%)
Stage 1	30 (21.58%)	0
Stage 2	101 (72.66%)	1 (1.35%)
Stage 3–5	7 (5.03%)	0
MoCA total score		
Mean	27.22	28.28
(min, max)	(14, 30)	(26, 30)

Abbreviations: DTI, diffusion tensor imaging; HC, healthy controls; MoCA, Montreal Cognitive Assessment; PD, Parkinson's disease.

N4BiasFieldCorrection algorithm from the Advanced Normalisation Tools (ANTs) (Tustison et al., 2010). Noise and Gibbs ringing artefacts were removed from the DWI series using dwidenoise and mrdegibbs, respectively. Both tools can be found in the MRtrix3 suite (Tournier et al., 2019). The eddy_openmp algorithm from FSL (Andersson & Sotiropoulos, 2016) was used to correct for eddy currents and subject movement. The DWI image intensities were then fit to a tensor using the weighted least squares (WLS) method included in SlicerDMRI (Norton et al., 2017).

A deformable registration of the b_0 non-DWI with the T1w structural image was calculated using the BRAINSFit tool from 3D Slicer (Johnson et al., 2007) for each subject. The DTI was resampled in 3D Slicer, preserving the principal direction, through the previously calculated transformation and is then co-registered with the T1w in 3D Slicer (Kikinis et al., 2014).

All the images were normalised to have dimension $96 \times 112 \times 96$ (on the sagittal, coronal and axial planes, respectively) before entering the CNN by adding empty slices or by removing them when needed at each extreme of the volume. Therefore, most of the meaningful information in the centre of the image was retained to serve as the CNN input in a standardised size.

Finally, the images in the patient database were flipped in order to ensure consistent lateralisation of PD to the right side, similar to Xiao et al. (2021). Images from the healthy controls were not flipped (except as a form of data augmentation). This simplifies the network as it only needs to detect PD on the one side, rather than discriminate between the two lateralities of the disease and healthy controls.

3.3 | Convolutional neural network

Since the aim of the proposal is to explore the regions in the image where a CNN focuses its attention, we decided to start this exploratory task by implementing a simple CNN architecture (Figure 2a). The CNN was trained to classify a DTI into either HC or PD using a supervised approach. The same architecture was used for both the

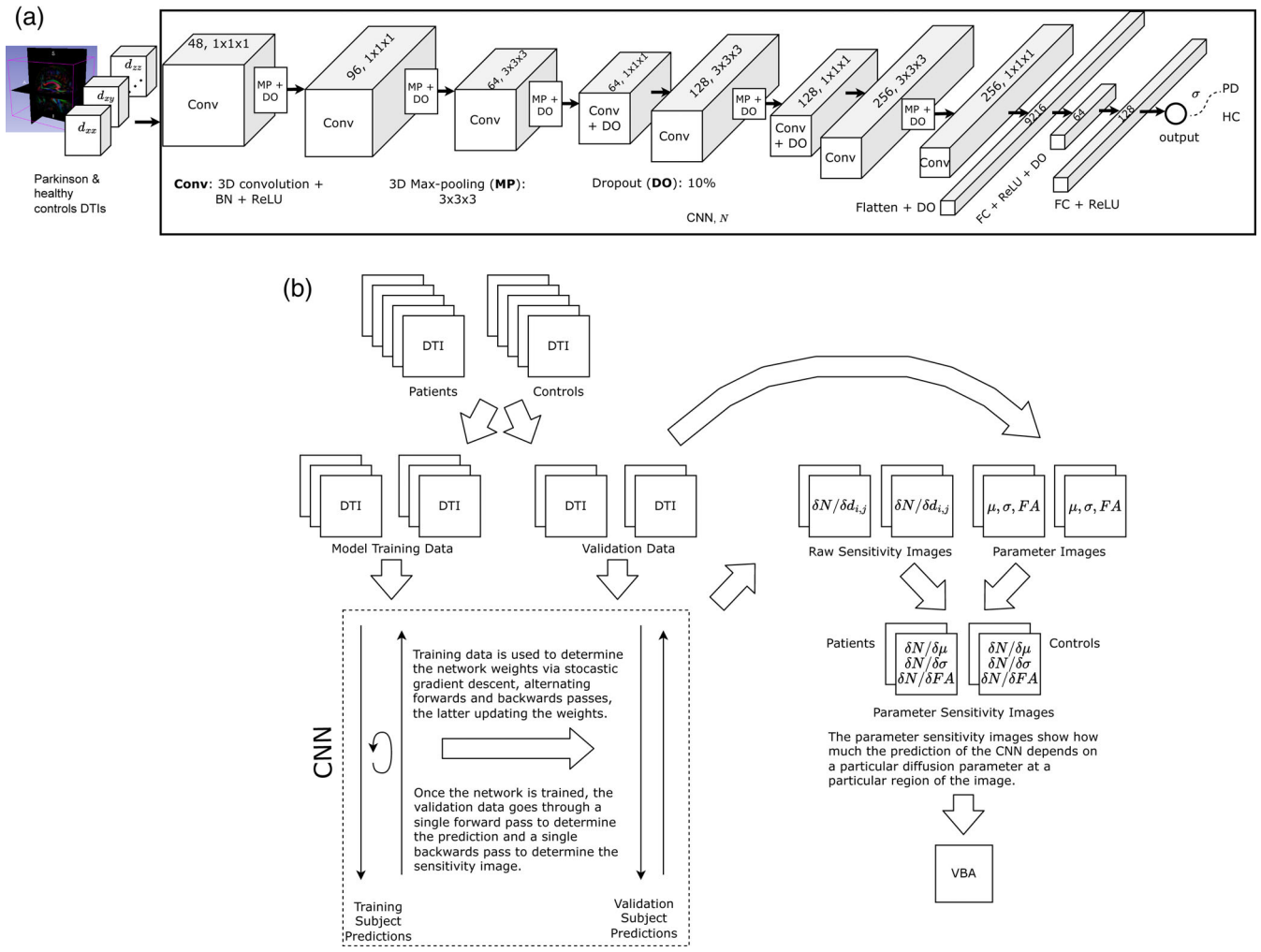

FIGURE 2 The proposed convolutional neural network (CNN) architecture with eight convolution and two fully connected layers, respectively (a) and, the overall method (b) used to train the CNN and compute the sensitivity images, which are used as the basis for voxel-based analysis

pre-registration and post-registration approaches (see Figure 3a,b) in order to keep the model complexity between the two approaches the same. The architecture was chosen to be a standard CNN with two convolution operators per resolution level, although the ones had to be removed in order to conserve GPU memory. The width of the layers was determined heuristically to use all available GPU memory given a batch size of 12.

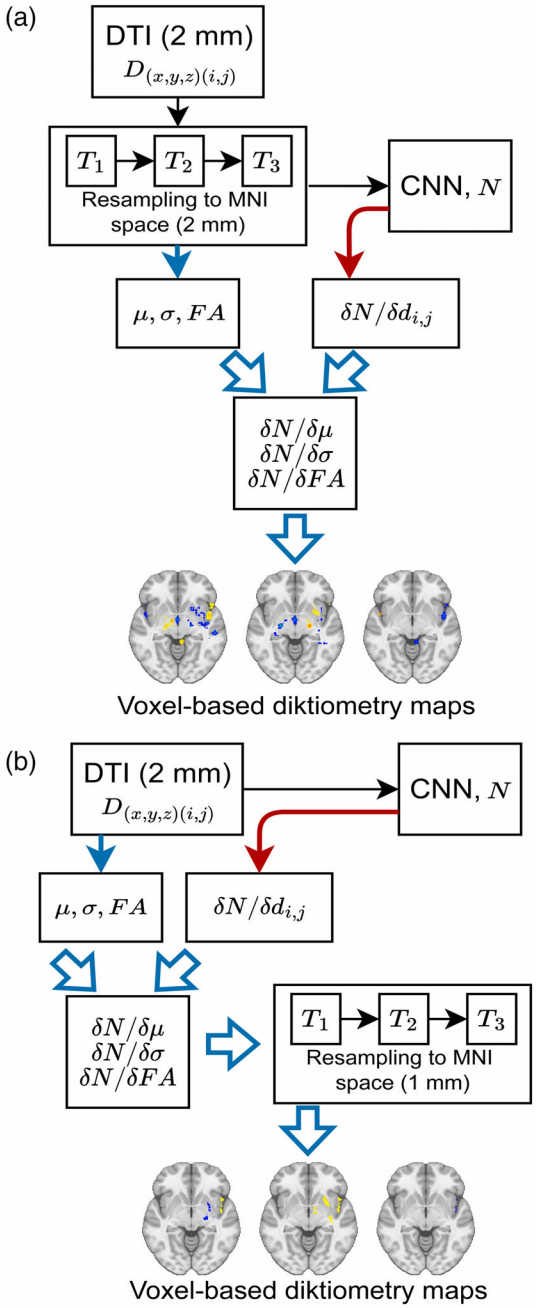

FIGURE 3 The two approaches to compute the sensitivity images (a) from registered nonaugmented diffusion tensor imaging (DTIs) and (b) from raw DTIs. The sensitivity images were computed only after the convolutional neural network (CNN) has been trained. The red arrow indicates the backward pass for the gradient computation. Voxel-based analysis (VBA) is performed over the registered sensitivity images to compute the final voxel-based diktiometry (VBD) maps.

We can consider our network to be some function, $N(\cdot)$, that takes an individual DTI, $D(p)$, as input and outputs the log-likelihood that the image is of a Parkinsonian patient, that is

$$P(p \text{ has PD}) \approx \frac{e^{N(D(p))}}{1+e^{N(D(p))}}, \tag{1}$$

where D is the DTI image. Note that increased values of $N(\cdot)$ reflects that the network believes the individual is more likely to be Parkinsonian and lower values imply the opposite.

In order to visualise the network's thought-process, we use a heuristic technique similar to Simonyan et al. (2013) in which the error gradients are propagated through the network also for testing images, not for modifying the network parameters, but for learning how sensitive the neural network is to particular characteristics in particular regions of the image. The overall training and application procedure is given in Figure 2b.

In order to ensure that the statistical power of later analyses are the same for the diktiometric and the traditional methods, we trained these networks in a 10-fold cross-validation manner which was repeated 10 times. Thus, the sensitivity maps for each individual can be interpreted as arising from an ensemble of 10 networks and each population-wide analysis (described in Section 3.5) is an aggregation of 10 networks. The training and testing patients for each fold were randomly selected from the 213 individuals in each repetition. We ensured that all the DTIs from the same individual were either entirely in the training or entirely in the testing subset.

The CNN was implemented in PyTorch and optimised using Adam optimiser with 0.01 L2 regularisation. The learning rate was set to 0.00001 and decreased it every 50 epochs by 4%. We also set a dropout rate of 10% between the convolutional and fully connected layers as shown in Figure 2a. We trained each CNN for a fixed 160 epochs.

Random left–right flipping on the HC class was performed during training for both the pre-registered and post-registered versions. For the post-registered version, we also added random in-plane rotations up to ±17° and random in-plane translations up to ±10 voxels for both classes.

3.4 | Tensor shape characteristics

In terms of the tensor shape characteristics, we used the *mean diffusivity* (μ), *anisotropy* (σ) and *pseudo-planarity* (θ):

$$\begin{aligned}\mu &= \tfrac{1}{3}(\lambda_1+\lambda_2+\lambda_3)\\ \sigma &= \sqrt{(\lambda_1-\mu)^2+(\lambda_2-\mu)^2+(\lambda_3-\mu)^2},\\ \theta &= \arctan\left(\lambda_1-\mu, \sqrt{\tfrac{1}{3}}(\lambda_2-\lambda_3)\right)\end{aligned} \tag{2}$$

which have the property of having orthogonal sensitivities (See Appendix A). The anisotropy in particular is highly related to another, more common metric, the *fractional anisotropy*:

$$\mathrm{FA}=\sqrt{\frac{3}{2}}\sqrt{\frac{\sum_i(\lambda_i-\mu)^2}{\sum_i\lambda_i{}^2}}=\frac{\sqrt{3/2}\sigma}{\sqrt{3\mu^2+\sigma^2}}, \tag{3}$$

but without the normalisation term in the denominator. This means that the anisotropy's bounds are $[0,1]$, but instead $\left[0,\sqrt{6}\mu\right]$. Because of this coupling with both μ and σ, the fractional anisotropy is not orthogonal and thus its sensitivity is coupled to that of the other two metrics:

$$\frac{\delta N}{\delta \mathrm{FA}}=\frac{3\mu^2+\sigma^2}{3\mathrm{FA}}\left(\sigma\frac{\delta N}{\delta\sigma}-\frac{1}{\mu}\frac{\delta N}{\delta\mu}\right). \tag{4}$$

After the image has been processed by the CNN, we performed an eigendecomposition of each voxel in the image to calculate μ, σ and FA to create patient-specific diffusion maps. In addition, we computed the sensitivity of the neural network with respect to μ, σ and FA ($\delta N/\delta\mu$, $\delta N/\delta\sigma$ and $\delta N/\delta \mathrm{FA}$ respectively).

3.5 | Population-wise registration

Unlike in traditional voxel-based analysis where a correspondence between voxels in different patients has to be found prior to the construction of the model, there is more flexibility with a CNN-based approach. Registration to find this correspondence can be done either prior to the training of the neural network, providing it with images with a standardised co-ordinate space, or afterwards solely for the population-wise aggregation of the results. The advantage of the first is that the network can learn particular spatially localised features more readily without having first to detect and localise them. However, this also removes the network's capability of using morphological information, that is, the sizes and shapes of the relevant anatomy have been standardised and their variability is no longer visible. The latter still has access to this morphological information and would be faster for prospective use, as registration would only be necessary for the analysis of a population rather than the individual patient. In addition, it can take advantage of more expressive data augmentation that affects this underlying co-ordinate system (e.g. random shifts and rotations) although it is more difficult to localise particular anatomical features.

The images were all deformably registered to the PD-specific template, ParkMedAtlis (Haegelen et al., 2013), and the Montreal Neurological Institute (MNI) template (Fonov et al., 2009; Fonov et al., 2011) using the BRAINSFit tool (Johnson et al., 2007).

3.6 | Voxel-based analysis statistical tests and filtering

The registered parameter or parameter-sensitivity maps can then be used for VBA in the common co-ordinate system. The values for each parameter map at corresponding locations through the patient and control database can then be rigorously compared. These maps were compared using SPM12 (rev. 7771) (Frackowiak et al., 1997) on MATLAB R2014a which assumes a linear relationship between the individual's status (i.e. PD vs. HC) and the value of the metrics, subject to Gaussian noise. We applied Gaussian smoothing (std. 8 mm isotropic) followed by spatial statistical correction with a family-wise error (FWE) rate of 1% and a minimum cluster size of 256 mm^3.

In order to determine the quality of our approach, we compared it to voxel-based diffusion analysis using the same patient database and registration procedures described in Sections 3.1 and 3.5 respectively.

4 | RESULTS

4.1 | Classification accuracy

In order to conclude that the underlying network correctly reflects PD, we first measured the performance of the network and ensured that it is well above that of random chance. Table 2 shows the classification results on the training and testing subsets, respectively. The overall accuracy was on the order of 70% for both the pre- and post-registration approaches. This greatly exceeds the accuracies on the order of 48%–58% found by Prasuhn et al. (2020) on the same data set using traditional VBA to perform feature extraction for a collection of different SVMs.

TABLE 2 Classification results for the training and testing subsets evaluated on the CNN ensemble

	Training		Testing	
	Ground truth		Ground truth	
	PD	HC	PD	HC
Post-registration				
Prediction				
PD	241	28	204	65
HC	33	258	102	189
Sensitivity	89.59%		75.84%	
Specificity	88.66%		64.95%	
Accuracy	89.11%		70.18%	
AUC	95.92%		76.88%	
Pre-registration				
Prediction				
PD	257	12	198	71
HC	2	289	102	189
Sensitivity	95.54%		73.61%	
Specificity	99.31%		64.95%	
Accuracy	97.50%		69.11%	
AUC	99.69%		75.25%	

Abbreviations: AUC, Area under the curve; HC, healthy controls; PD, Parkinson's disease; AU.

Regarding the two approaches, we observed marginally better classification performance by training the CNN directly on the nonregistered DTI images (i.e. the post-registration approach) than by registering them in advance (i.e. the pre-registration approach). This difference however is not significant, indicating that the two methods perform equivalently despite having different access to features such as spatial data augmentation, morphological information or a consistent co-ordinate space.

4.2 | Voxel-based diktiometry maps

Figures 4–7 show the qualitative results for our voxel-based diktiography approach and the comparative voxel-based diffusion approach using SPM and linear statistical analysis and a significance threshold of $p < 0.01$ FWE for the MD, A and FA metrics respectively. The majority of the results appear in the MD and A metrics for the CNN-based methods, and in the MD and FA methods for the traditional approach.

For the MD shown in Figure 4, both CNN-based approaches identified a decrease (blue to cyan) in diffusivity generally in the cerebellar white matter and the lenticular nucleus (composed of putamen and pallidum) (lateral to symptom onset). We can also observe an increase (yellow to red) in the diffusivity in the fourth ventricle (not lateralised) and the lateral fissure (lateral to symptom onset) which are likely to result from morphological changes. In the pre-registration version, several structures appeared on both sides of the brain, although with different signs, indicating an asymmetry. One of these structures appears to be the cortico-spinal tract, indicating a laterised effect that is easiest for the network to detect via an asymmetry check. In addition to these asymmetries, the pre-registration approach also found a decrease in diffusivity in the white matter of the temporal lobe.

In the traditional approach, an even slighter decrease is seen in the lenticular nucleus. This then extends in the superior direction until the white matter of the cortico-spinal tract. The traditional approach also shows numerous small clusters in cortical regions, especially grey matter regions on the boundary of the external CSF. Due to their small size and distribution, these are likely to be statistical artefacts arising from large patient variability in these regions. Thus, the methods largely agree in the cerebral regions, although the proposed method generally for cerebellar structures to be more indicative of PD.

For the anisotropy, Figure 5, many of the same regions are identified, although with different signs. For the cerebellum, there is a positive sensitivity to the anisotropy in the caudal region, whereas there is a negative sensitivity in the rostral region closer to the cerebellar grey matter. (This is lateralised for the post-registration approach, but bilateral for the pre-registration approach.) There is also a positive sensitivity to anisotropy in the area surrounding the putamen. In the pre-registration approach, there is a bilateral sensitivity in the cortico-spinal tract adjacent to the thalamus, but again with different signs on the two sides, indicating an asymmetry. One interesting finding is a negative sensitivity in the contralateral ventricle which is present in both CNN types. For the traditional approach, the results using the anisotropy appeared to be similar to those for the fractional anisotropy and will be discussed in the following paragraph.

For the fractional anisotropy, Figure 6, the traditional VBA approach showed an increase in FA in the cortico-spinal tract (lateral to symptom onset) and a decrease in FA in the putamen. There also appears to be a decrease in the FA in the area of the optical radiations, although that is more difficult to interpret in the context of PD and thus may be artefactual. Interpreting the maps for the CNNs is somewhat more difficult as the sensitivity with respect to FA is derived from those of MD and A as shown in Equation (4). In the post-registration image, only the fourth ventricle and lateral fissure were identified, suggesting a morphological change. In the pre-registration image, the lateral fissure showed an asymmetry and the contralateral ventricle was also identified, suggesting again a morphological origin.

The pseudo-planarity, Figure 7, did not yield any major results, possibly due to being the smallest source of variation in the diffusion tensor and only well-defined when the other metrics take on high values.

4.3 | Classification accuracy

Despite the numerous studies regarding the effects of PD on white matter and diffusion characteristics in the brain, Prasuhn et al. (2020) expressed doubt that DTI could be used for PD classification given the failure of their registration + SVM approach for particular subcortical structures, this approach having achieved some success in classifying other neurological disorders such as epilepsy (Chen et al., 2020). The successful classification between PD patients and HC by traditional CNNs indicates that the nonlocality and nonlinearity yielded by neural networks are important to PD classification and, relatedly, that descriptive diffusion characteristics of PD are not localised to a single voxel or even a local region. Nonlinear methods such as a combination of multi-modal MRI features (Talai et al., 2021) have yielded similar accuracies as our CNN approaches, although the approach (Talai et al., 2021) used for selecting the "optimal" method involved a degree of data leakage and thus they potentially overestimated their performance on their much smaller data set.

However, the purpose of the classifier in this article is indirect, as it is only meant to provide a basis for network sensitivity analysis and visualisation. Thus, we have chosen the simplest architecture in order to display the power of the diktiometric approach at extracting nonlocal biomarkers even from relatively simple strong classifiers. Even if the classification performance is not as high as that based on other images or diagnostic criteria, if it is significantly better than random chance, analysing this classifier can provide insight into how PD affects diffusion in the brain.

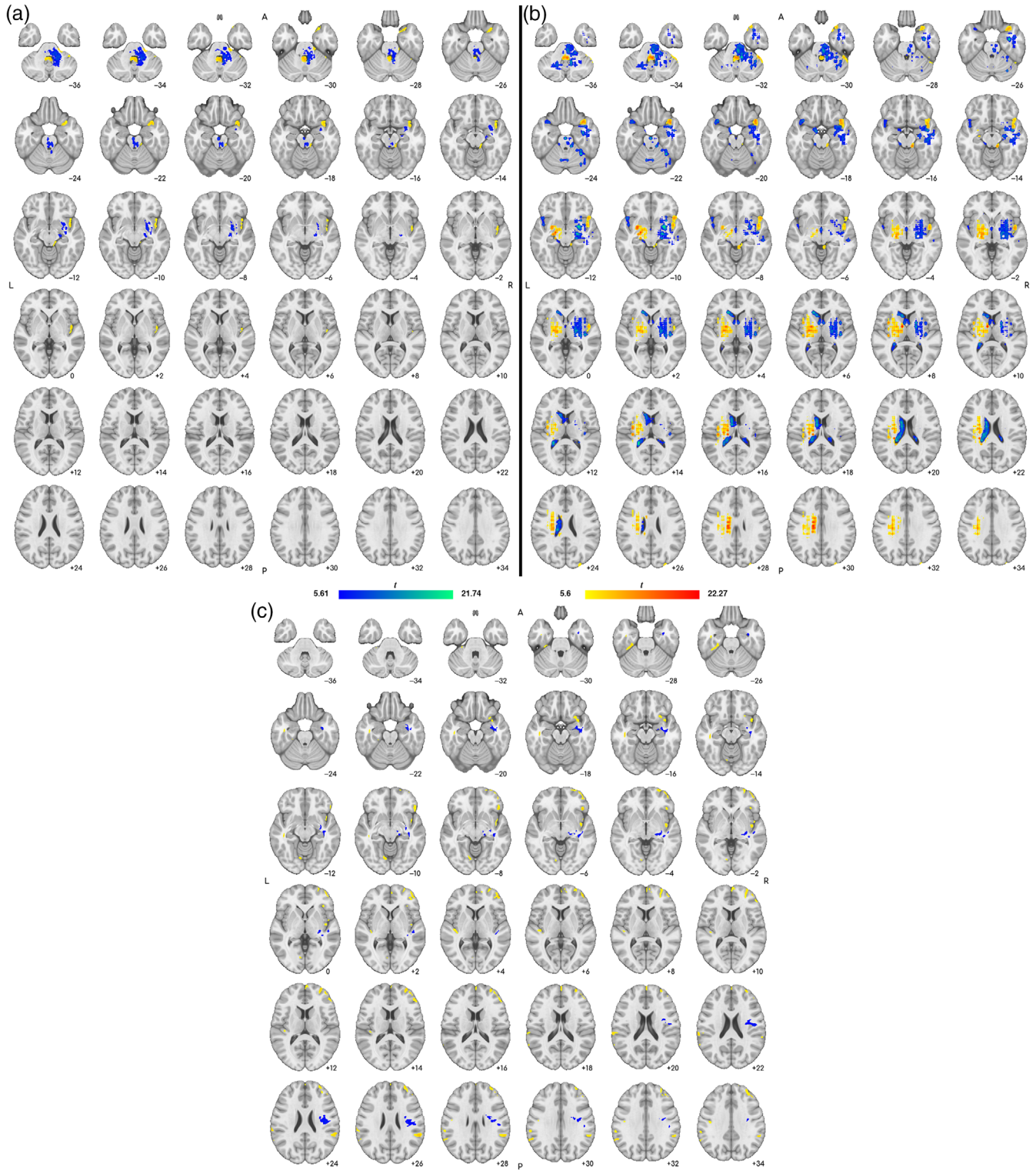

FIGURE 4 Voxel-based diktiometry (post-registration (a) and pre-registration (b)) sensitivity results and traditional voxel-based diffusion analysis results (c) from $z = -36$ to $z = +34$ in MNI space for the mean diffusivity (MD). Left-dominant Parkinson's disease (PD) subjects have been laterally flipped. (Note that patient-right is shown on image-right, i.e. neurological convention.)

4.4 | Comparison to traditional VBA approaches

In order to confirm that the registration and statistical analysis aspects of our method are robust, we used the same pipelines to perform traditional VBA. This has displayed results highly consistent with those of Xiao et al. (2021), despite differences in registration and statistical processing, confirming the robustness of the VBA method.

The goal of this article is to show an alternative to traditional VBA approaches that are extensively used throughout the literature

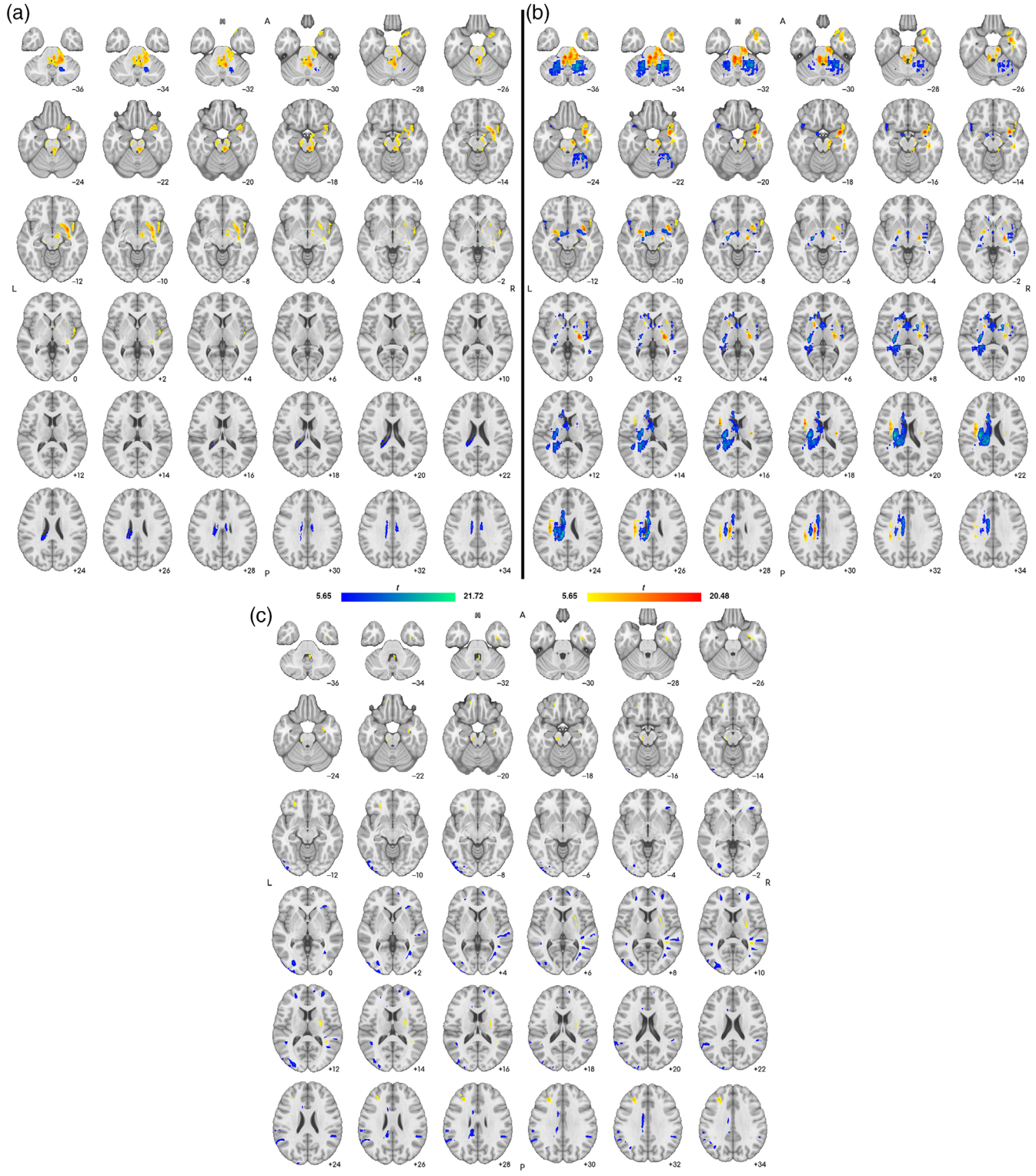

FIGURE 5 Voxel-based diktiometry (post-registration (a) and pre-registration (b)) sensitivity results and traditional voxel-based diffusion analysis results (c) from $z = -36$ to $z = +34$ in MNI space for the anisotropy (A). Left-dominant Parkinson's disease (PD) subjects have been laterally flipped. (Note that patient-right is shown on image-right, i.e. neurological convention.)

for imaging biomarker discovery and for better understanding the pathophysiology of neurological disorders (Atkinson-Clement et al., 2017). Due to the necessarily uncertain nature of biomarker discovery (and therefore no ground truth biomarkers), there is no directly quantitative way to compare methods. Nevertheless, some interesting qualitative and theoretical comparisons can be made.

One interesting theoretical difference between the proposed diktiometric method and the traditional voxel-based approach is that the

FIGURE 6 Voxel-based diktiometry (post-registration (a) and pre-registration (b)) sensitivity results and traditional voxel-based diffusion analysis results (c) (third column) from $z = -36$ to $z = +34$ in MNI space for the fractional anisotropy (FA). Left-dominant Parkinson's disease (PD) subjects have been laterally flipped. (Note that patient-right is shown on image-right, i.e. neurological convention.)

former is sensitive to *nonlocal patterns* in the underlying voxel values rather than the values themselves. This is crucial in that it allows for the identification of discriminative features conditioned on the entire image even if the marginal distribution of said feature at said voxel is not highly discriminative in itself.

4.4.1 | Interpreting diktiographic sensitivity maps

One negative aspect of our approach compared to traditional VBA is the complexity in which the visual results must be interpreted. Due to the complex nonlinear and nonlocal nature of the CNN and sensitivity

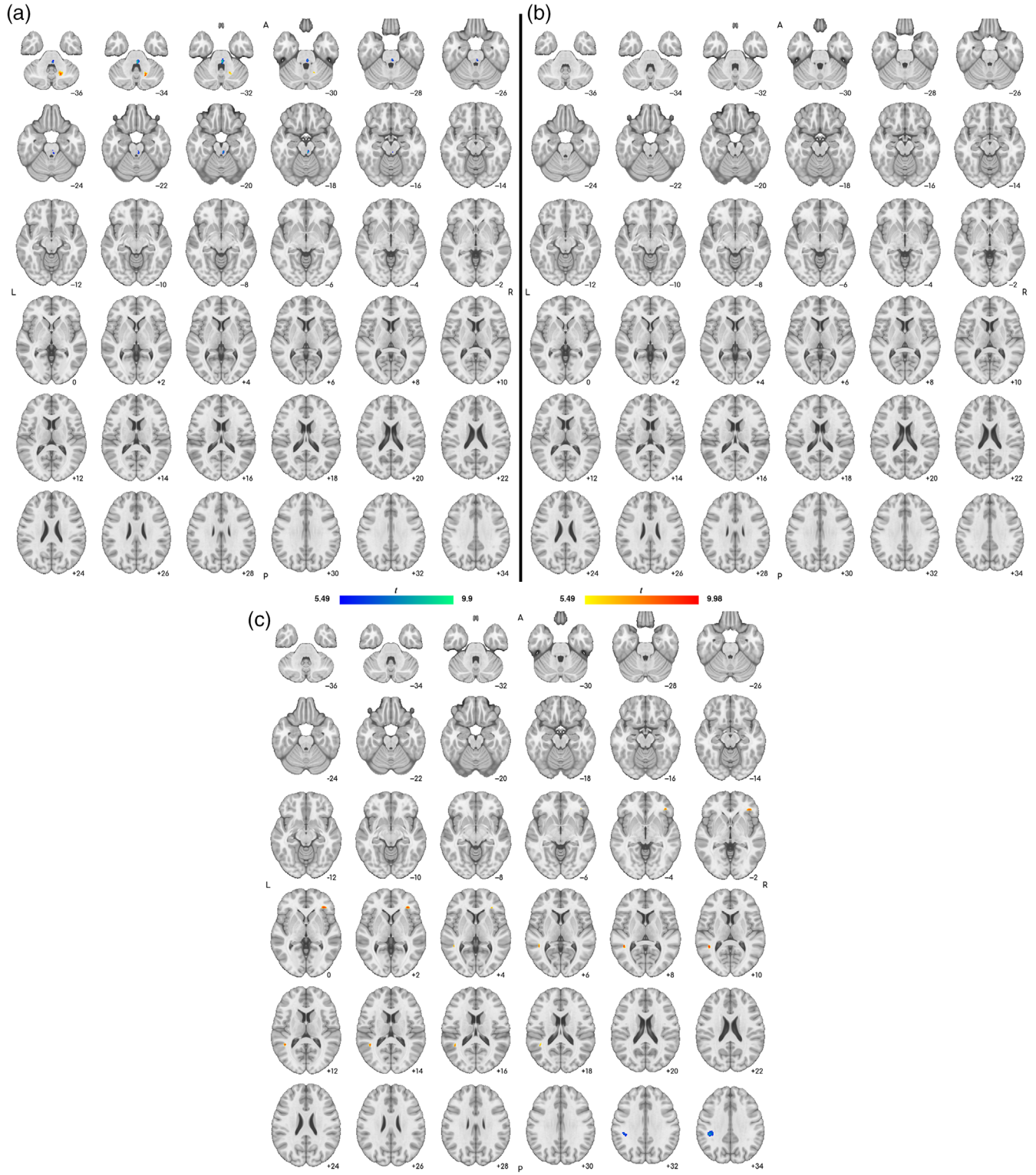

FIGURE 7 Voxel-based diktiometry (post-registration (a) and pre-registration (b)) sensitivity results and traditional voxel-based diffusion analysis results (c) (third column) from $z = -36$ to $z = +34$ in MNI space for the pseudo-planarity (PsPl). Left-dominant Parkinson's disease (PD) subjects have been laterally flipped. (Note that patient-right is shown on image-right, i.e. neurological convention.)

analysis used, the analysis of any particular region of the image may not be done in isolation and regions should be considered as potentially related to each other.

For example, comparing across the pre- and post-registration approaches provides information about the importance of morphological information and anatomically constant co-ordinate systems. As

deformable registration minimises morphological differences by definition, any signal arising from the post-registration approach that does not appear in the pre-registration approach is likely due to morphological changes rather than changes in diffusion characteristics. Signals arising from the pre-registration approach that are not in the post-registration approach are likely due to the network requiring a consistent spatial co-ordinate space to form longer-distance connections, such as between equivalent anatomy on opposite hemispheres or between distant anatomical regions in the same hemisphere.

4.4.2 | Coupling and suppression of distinct biomarkers

Fundamentally, our analysis couples together nonlocal biomarkers that may actually be independent. To see that, consider that there are two ways that a collection of biomarkers may be dependent on each other: they may be correlated in the input data distribution and/or they may be nonlinearly coupled together by the classifier. To give an example of the latter, if the sensitivity maps highlight two regions, there is a possibility that they are coupled together, forming two complementary facets of a singular, more nonlocal biomarker (e.g. an asymmetry biomarker). For traditional VBA, this is not possible as the classifiers themselves are computed independently of each other and thus ensure that the identified features are also treated independently; any coupling between features must be a product of the input data distribution, not the classifier. This "biomarker classification coupling" problem can become even more problematic as we consider multiple nonlocal biomarkers that might interact with each other. We have some evidence that this may be the case, for example in Figure 4a,b, we see a lateralised biomarker in the basal ganglia (slides −10 to −4) in the post-registration approach, which is bilateral in the pre-registration approach, indicating that the post-registration method is looking solely at the disease onset side whereas the pre-registration approach is looking at the asymmetry between the two sides. This could indicate two separate biomarkers or two different methods for extracting a singular biomarker. In addition, many areas that were highlighted using a single metric (e.g. MD) were also highlighted in others (e.g. A and FA) which again could lead to the interpretation of distinct biomarkers or different facets of a singular biomarker.

We took a heuristic approach, assuming that the distant highlighted regions should be interpreted as independent unless they correspond to the same anatomy (e.g. when the same region is highlighted on both sides of the brain with opposite sensitivity signs, we interpreted it as a single asymmetry biomarker). We also made the assumption that the diffusion metrics were independent of each other.

In addition, our methodology may also suppress some nonlocal biomarkers even if they are accessible to the current CNN architecture. This is because the results visualised are from an ensemble of different CNNs rather than any singular one. There is a distinct possibility that these networks detect and are sensitive to a different set of biomarkers. Thus, the sensitivity maps only indicate features that are commonly and robustly selected by the majority of trained CNNs, with other biomarkers remaining invisible.

One area of future work that would address both these issues is to investigate methods for decoupling the sensitivity maps into independent components.

This is not an issue in traditional VBA approaches due to their simplicity and the more limited scope of (strictly local) biomarkers they can detect. This is both an advantage and a disadvantage as nonrobust local biomarkers can still be detected with the traditional VBA approach even if they have only a marginal correlation to the underlying disease.

4.4.3 | CNN sensitivity at specific regions

Cognisant of the previous two sections, one can now begin to interpret and compare biomarkers across our proposed approach and traditional VBA.

Using the traditional method, we have largely reproduced the results generated by the VBA performed by Xiao et al. (2021), specifically in terms of the white matter bundle extending from the cortex to the brain stem on the collateral to symptom onset. This validates that the statistical mapping method used for all three approaches is coherent with the literature. (Note that Xiao et al. (2021) used a white-matter mask in their approach whereas we did not, leading to a number of grey-matter and sulcal regions also to be identified.) This also validates that simple CNNs trained on a PD diagnosis task do look at relevant regions of the brain.

Notably, our approach appears to generate a number of additional results in comparison to the traditional method. This may confirm Prasuhn et al.'s (2020) observation that linear classification methods are insufficient. The fact that the traditional method and the pre-registration method rarely overlap also confirms this, suggesting that the regions identified by traditional VBA are not salient enough for extraction as meaningful features. One clear example of this is the reliance of the neural networks on a mix of cerebellar and cerebral structures, unlike the traditional approach which never found any local biomarkers in the cerebellum, which is in agreement with other recent VBD investigations (Atkinson-Clement et al., 2017). This is of particular interest as the community has been recently calling for more investigation of the cerebellum in the aetiology of PD (Mirdamadi, 2016).

Our population-wise results of the post-registration networks show a distinct focus of the neural networks on the basal ganglia and cerebellum, indicating that the network can both identify these regions easily as well as use them to distinguish between PD patients and healthy controls. For the pre-registration network, these locations are also highlighted, indicating that the diffusion characteristics of these regions, rather than their morphology which is removed in the pre-registration workflow, are predictive of PD. What is of particular interest is that these locations are not highlighted in the traditional approach, indicating that the predictive characteristics are truly nonlocal, referring to a collection of correlated diffusion changes rather than a cluster of pixels that happen to co-vary. The traditional approach appeared to highlight many areas that are not associated with the symptomatology of PD, likely due to statistical artefacts,

whereas the proposed methods highlight areas that are either already known in PD physiopathology (e.g. the basal ganglia) or are under investigation such as the cerebellum.

Our studies agree with the more general observation made by Xiao et al. (2021) in that differences between PD patients and controls seem to be lateralised. However, there appears to be a difference in how this lateralisation can be interpreted, specifically between the pre- and post-registration approaches. One possible explanation for this is that the pre-registration approach is comparing the two hemispheres, looking for asymmetries which would not be as easily performed by the post-registration network that cannot rely on particular pixels location always representing the same anatomy.

4.4.4 | Role of different diffusion characteristics

Unsurprisingly, MD played the strongest role in the analysis, generating the highest number of significant voxels for both the proposed diktiometric and the traditional voxel-based diffusion analysis approaches. Given that other voxel-based diffusion studies uniformly use MD in their analysis, this is coherent with the literature and confirms the role of MD in understanding the diffusion characteristics of neurological disorders (Atkinson-Clement et al., 2017).

Interestingly, FA appears to have played only a secondary role in the analysis of the CNNs, largely due to its strong connection to MD in its normalisation term. This means that the regions identified by FA sensitivity maps were almost always a subset of those identified by an MD sensitivity, only with a sign inversion as increasing the MD naturally lowered the FA. Although this theoretically could have been overcome by the contribution of the anisotropy (σ) term, this was not observed. Thus, in this sensitivity analysis, it was more meaningful to look to the unnormalised anisotropy term rather than FA to see information traditionally associated with organised microstructure. In addition, we observed the same qualitative results in the traditional voxel-based diffusion analysis. This calls into question the role of FA in diffusion analysis studies more generally as it may be possible that correlated MD increases with FA decreases (and vice versa) are possibly solely due to MD and the contribution of the denominator term, although much more investigation would need to be done to determine what diffusion parameters are sufficiently decoupled to avoid this issue.

The pseudo-planarity appeared to play a very minor role in the analysis, which indicates that the majority of the sensitivity was assigned to the orthogonal MD and A terms. This is unsurprisingly as in the literature, measurements of MD and FA tend to be more significant than the other metrics that measure the shape of the tensor (Abe et al., 2010).

4.5 | Limitations and future work

4.5.1 | Technical limitations of simple CNNs

One of the major technical limitations of this framework involves the *translation invariance* property of traditional, simple CNNs. Although translation invariance is often seen as a strength of CNNs, in the context of medicine, it means that the network has to be able to first localise particular structures before being able to use their diffusion characteristics. This is in contrast to simpler methods such as registration + SVM (Chen et al., 2020) in which the metrics can be extracted from consistent anatomical locations directly as input. This is particularly evident in the post-registration network lacking symmetric features that are opposite in magnitude, indicating an assessment of asymmetry.

However, it should be noted that this limits the degree to which one can use geometric data augmentation techniques, which we hypothesise is one of the reasons why applying CNNs on unregistered images (i.e. the post-registration approach) had a slighter higher performance than when applied on registered ones (i.e. the pre-registration approach). This experiment also relied on a traditional CNN architecture which also leaves open the possibility that a more complex architecture could incorporate more localisation information.

4.5.2 | Diagnostic performance and use

Lastly, the classification performance of these networks needs to be improved before they can be used as a diagnostic tool on their own. Although the motivation behind this study was to develop a tool similar to VBA, the end goal of these systems should be to improve the diagnosis of patients more accurately and at earlier time-points in the disease's progression. Although the method's 70% accuracy is much lower than we have come to expect from deep learning methods in general, it must be noted that it is still, to the best of the author's knowledge, one of the best in the literature that uses only diffusion tensor information. Not even the T1 structural images, which are often acquired at the same time, are used in the current framework in order for the networks to focus on specifically diffusion-related biomarkers. Thus, there are four ways we envision this technique to be more useful as a diagnostic tool:

1. the inclusion of more imaging modalities,
2. architectural improvements,
3. prediction of symptom severity, rather than diagnosis,
4. increasing the size of the training data set or
5. investigation of a more heterogeneous cohort including early-stage PD.

With the exception of including T1 imaging, this would require extensive research and data collection.

By focusing on symptomatology rather than diagnosis, it may be possible to use these methods in a broader array of clinical contexts related to PD. For example, dementia and other severe cognitive disorders are common counter-indications for deep brain stimulation due to its potentially deleterious impact on cognition, so if cognition-specific biomarkers can be identified, these patients could be more readily screened using imaging (Rodriguez-Oroz et al., 2005). Similarly, more patient-specific maps could be used to help guide deep brain stimulation procedures by targeting areas that are more affected by

that patient's specific symptomatology. In addition, separating the maps based on symptomatology may also help to disentangle the effects specific to PD from those caused by other age-related disorders common in PD cohorts.

4.5.3 | Other areas of future work

Aside from technical development and network optimisation, there is a large potential for the use of more complex diffusion models than the standard diffusion tensor which could still be used in the framework of CNN sensitivity analysis. Specific examples of these models which are already known to be useful in the context of PD include higher order tensors such as those extracted in diffusion kurtosis imaging (J.-J. Wang et al., 2011) and fibre orientation and density distributions (Xiao et al., 2021). This would also have the benefit of nuancing the results from a scientific perspective, giving us more insight into the diffusion-related effects of PD.

Even with the tensor model, there are still avenues to explore. By acting directly on the diffusion tensors themselves, sensitivity information may be extractable for other information, such as tensor orientation, that is not usually investigated. However, more research would need to be done to ensure that the sensitivity maps for tensor orientation can be meaningfully aggregated (i.e. meaningfully combined in a population-wise manner) and interpreted. The approach may also be improved through the use of more advanced diffusion protocols that have been developed since the PPMI data set was launched. However, any improvement in the protocol may be offset by the more limited supply of training data at least until new, large databases are constructed.

5 | CONCLUSIONS

This article presents VBD, a technique for combining traditional VBA with CNNs in order to identify, visualise, and analyse regions of the brain associated with a particular disorder. The strength of this method comes from the inversion of the traditional VBA paradigm. In VBA, regions of the image that show statistically robust differences are identified which could then be used to help guide diagnosis. In VBD, a strong classifier is used first, and the classifier sensitivity is evaluated for statistical significance, displaying what information is empirically more useful for said classifier. This inversion allows for the ROI to display a more nonlocal character unlike traditional VBA in which (setting aside smoothing and spatial correction) the emphasis is on singular voxels at particular registered locations.

This method has shown evidence for diffusion biomarkers for PD that are specifically nonlocal in nature. For example, there is evidence to suggest that asymmetry in the diffusivity of the white matter may be a usable biomarker for lateralisation. These biomarkers are inherently nonlocal in that they describe how diffusion characteristics in different, separate regions of the brain co-vary and correlate with each other and not only the disease status.

Interestingly, for traditionally CNN-based classification, it appears that the white matter areas in the cerebellum and brain stem are particularly indicative of PD, more so than the cerebrum which provides additional evidence of the cerebellar role in PD (Mirdamadi, 2016).

Overall, VBD provides an interesting new tool for the investigation of nonlocal imaging biomarkers of neurological disorders.

ACKNOWLEDGMENTS

Alfonso Estudillo Romero is supported through the Fondation Recherche Médicale (FRM) DIC20161236441, the SAD Région Bretagne programme, and the Institut des Neurosciences Cliniques de Rennes (INCR). John S. H. Baxter is supported by the Institut national de la santé et de la recherche médicale (INSERM).

CONFLICT OF INTEREST

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

All data used in this study were taken from the Parkinson's Progression Markers Initiative (PPMI) created by the Micheal J. Fox Foundation. These data are openly available to the public and no closed data were used in this article.

ORCID

Alfonso Estudillo-Romero https://orcid.org/0000-0002-1921-9913
Claire Haegelen https://orcid.org/0000-0002-8341-4887
Pierre Jannin https://orcid.org/0000-0002-7415-071X
John S. H. Baxter https://orcid.org/0000-0003-3548-4343

REFERENCES

Abe, O., Takao, H., Gonoi, W., Sasaki, H., Murakami, M., Kabasawa, H., Kawaguchi, H., Goto, M., Yamada, H., Yamasue, H., Kasai, K., Aoki, S., & Ohtomo, K. (2010). Voxel-based analysis of the diffusion tensor. *Neuroradiology*, *52*(8), 699–710.

Andersson, J. L. R., & Sotiropoulos, S. N. (2016). An integrated approach to correction for off-resonance effects and subject movement in diffusion MR imaging. *NeuroImage*, *125*, 1063–1078.

Atkinson-Clement, C., Pinto, S., Eusebio, A., & Coulon, O. (2017). Diffusion tensor imaging in Parkinson's disease: Review and meta-analysis. *NeuroImage: Clinical*, *16*, 98–110.

Chen, S., Zhang, J., Ruan, X., Deng, K., Zhang, J., Zou, D., He, X., Li, F., Bin, G., Zeng, H., & Huang, B. (2020). Voxel-based morphometry analysis and machine learning based classification in pediatric mesial temporal lobe epilepsy with hippocampal sclerosis. *Brain Imaging and Behavior*, *14*, 1945–1954.

Cousineau, M., Jodoin, P. M., Garyfallidis, E., Côté, M. A., Morency, F. C., Rozanski, V., Grand'Maison, M., Bedell, B. J., & Descoteaux, M. (2017). A test-retest study on Parkinson's PPMI dataset yields statistically significant white matter fascicles. *NeuroImage: Clinical*, *16*, 222–233.

Descoteaux, M., Wiest-Daesslé, N., Prima, S., Barillot, C., & Deriche, R. (2008). Impact of Rician adapted non-local means filtering on HARDI. In D. Metaxas, L. Axel, G. Fichtinger, & G. Székely (Eds.), *Medical image computing and computer-assisted intervention – MICCAI 2008* (pp. 122–130). Springer Berlin Heidelberg.

Fonov, V., Evans, A., McKinstry, R., Almli, C., & Collins, D. (2009). Unbiased nonlinear average age-appropriate brain templates from birth to adulthood. *NeuroImage*, *47*(Suppl. 1), S102.

Fonov, V., Evans, A. C., Botteron, K., Almli, C. R., McKinstry, R. C., & Collins, D. L. (2011). Unbiased average age-appropriate atlases for pediatric studies. *NeuroImage*, *54*(1), 313–327.

Frackowiak, R., Friston, K., Frith, C., Dolan, R., & Mazziotta, J. (Eds.). (1997). *Human brain function*. Academic Press.

Haegelen, C., Coupé, P., Fonov, V., Guizard, N., Jannin, P., Morandi, X., & Collins, D. L. (2013). Automated segmentation of basal ganglia and deep brain structures in MRI of Parkinson's disease. *International Journal of Computer Assisted Radiology and Surgery*, *8*(1), 99–110.

Iglesias, J. E., Liu, C., Thompson, P. M., & Tu, Z. (2011). Robust brain extraction across datasets and comparison with publicly available methods. *IEEE Transactions on Medical Imaging*, *30*(9), 1617–1634.

Johnson, H., Harris, G., & Williams, K. (2007). BRAINSFit: Mutual information registrations of whole-brain 3D images, using the insight toolkit. *The Insight Journal*, *180*, 1–10.

Kikinis, R., Pieper, S. D., & Vosburgh, K. G. (2014). 3D slicer: A platform for subject-specific image analysis, visualization, and clinical support. In F. Jolesz (Ed.), *Intraoperative imaging and image-guided therapy* (pp. 277–289). Springer.

Li, X., Morgan, P. S., Ashburner, J., Smith, J., & Rorden, C. (2016). The first step for neuroimaging data analysis: DICOM to NIfTI conversion. *Journal of Neuroscience Methods*, *264*, 47–56.

Li, Y., Guo, T., Guan, X., Gao, T., Sheng, W., Zhou, C., Wu, J., Xuan, M., Gu, Q., Zhang, M., Yang, Y., & Huang, P. (2020). Fixel-based analysis reveals fiber-specific alterations during the progression of Parkinson's disease. *NeuroImage: Clinical*, *27*, 102355.

Marek, K., Chowdhury, S., Siderowf, A., Lasch, S., Coffey, C. S., Caspell-Garcia, C., Simuni, T., Jennings, D., Tanner, C. M., Trojanowski, J. Q., Shaw, L. M., Seibyl, J., Schuff, N., Singleton, A., Kieburtz, K., Toga, A. W., Mollenhauer, B., Galasko, D., Chahine, L. M., … Taucher, J. (2018). The Parkinson's progression markers initiative (PPMI) – Establishing a PD biomarker cohort. *Annals of Clinical and Translational Neurology*, *5*(12), 1460–1477.

Mirdamadi, J. L. (2016). Cerebellar role in Parkinson's disease. *Journal of Neurophysiology*, *116*(13), 917–919.

Mordvintsev, A., Olah, C., & Tyka, M. (2015). Inceptionism: Going deeper into neural networks. https://research.googleblog.com/2015/06/inceptionism-going-deeper-into-neural.html

Norton, I., Essayed, W. I., Zhang, F., Pujol, S., Yarmarkovich, A., Golby, A. J., Kindlmann, G., Wassermann, D., Estepar, R. S. J., Rathi, Y., Pieper, S., Kikinis, R., Johnson, H. J., Westin, C. F., & O'Donnell, L. J. (2017). SlicerDMRI: Open source diffusion MRI software for brain cancer research. *Cancer Research*, *77*(21), e101–e103.

Prasuhn, J., Heldmann, M., Münte, T. F., & Brüggemann, N. (2020). A machine learning-based classification approach on Parkinson's disease diffusion tensor imaging datasets. *Neurological Research and Practice*, *2*(1), 1–5.

Rodriguez-Oroz, M. C., Obeso, J., Lang, A., Houeto, J. L., Pollak, P., Rehncrona, S., Kulisevsky, J., Albanese, A., Volkmann, J., Hariz, M., Quinn, N. P., Speelman, J. D., Guridi, J., Zamarbide, I., Gironell, A., Molet, J., Pascual-Sedano, B., Pidoux, B., Bonnet, A. M., … Van Blercom, N. (2005). Bilateral deep brain stimulation in Parkinson's disease: A multicentre study with 4 years follow-up. *Brain*, *128*(10), 2240–2249.

Schuff, N., Wu, I. W., Buckley, S., Foster, E. D., Coffey, C. S., Gitelman, D. R., Mendick, S., Seibyl, J., Simuni, T., Zhang, Y., Jankovic, J., Hunter, C., Tanner, C. M., Rees, L., Factor, S., Berg, D., Wurster, I., Gauss, K., Sprenger, F., … Marek, K. (2015). Diffusion imaging of nigral alterations in early Parkinson's disease with dopaminergic deficits. *Movement Disorders*, *30*(14), 1885–1892.

Simonyan, K., Vedaldi, A., & Zisserman, A. (2013). Deep inside convolutional networks: Visualising image classification models and saliency maps. arXiv preprint arXiv:1312.6034.

Talai, A. S., Sedlacik, J., Boelmans, K., & Forkert, N. D. (2021). Utility of multimodal MRI for differentiating of Parkinson's disease and progressive supranuclear palsy using machine learning. *Frontiers in Neurology*, *12*, 648548.

Tournier, J. D., Smith, R., Raffelt, D., Tabbara, R., Dhollander, T., Pietsch, M., Christiaens, D., Jeurissen, B., Yeh, C. H., & Connelly, A. (2019). MRtrix3: A fast, flexible and open software framework for medical image processing and visualisation. *NeuroImage*, *202*, 116137.

Tustison, N. J., Avants, B. B., Cook, P. A., Zheng, Y., Egan, A., Yushkevich, P. A., & Gee, J. C. (2010). N4itk: Improved N3 bias correction. *IEEE Transactions on Medical Imaging*, *29*(6), 1310–1320.

Wang, F., Kaushal, R., & Khullar, D. (2020). Should health care demand interpretable artificial intelligence or accept "black box" medicine? *Annals of Internal Medicine*, *172*(1), 59–60.

Wang, J. J., Lin, W. Y., Lu, C. S., Weng, Y. H., Ng, S. H., Wang, C. H., Liu, H. L., Hsieh, R. H., Wan, Y. L., & Wai, Y. Y. (2011). Parkinson disease: Diagnostic utility of diffusion kurtosis imaging. *Radiology*, *261*(1), 210–217.

Wasserthal, J., Neher, P. F., Hirjak, D., & Maier-Hein, K. H. (2019). Combined tract segmentation and orientation mapping for bundle-specific tractography. *Medical Image Analysis*, *58*, 101559.

Xiao, Y., Peters, T. M., & Khan, A. R. (2021). Characterizing white matter alterations subject to clinical laterality in drug-naïve de novo Parkinson's disease. *Human Brain Mapping*, *42*(14), 4465–4477.

Zhang, Y., & Burock, M. A. (2020). Diffusion tensor imaging in Parkinson's disease and Parkinsonian syndrome: A systematic review. *Frontiers in Neurology*, *11*, 531993.

How to cite this article: Estudillo-Romero, A., Haegelen, C., Jannin, P., & Baxter, J. S. H. (2022). Voxel-based diktiometry: Combining convolutional neural networks with voxel-based analysis and its application in diffusion tensor imaging for Parkinson's disease. *Human Brain Mapping*, *43*(16), 4835–4851. https://doi.org/10.1002/hbm.26009

APPENDIX A

GRADIENT PROPAGATION TO DIFFUSION METRICS

The result of gradient propagation through the network to the input yields the derivatives $\frac{\delta N}{\delta D_{(x,y,z),(i,j)}}$ where $D_{(x,y,z),(i,j)}$ is the value of the ith row, jth column element of the diffusion tensor at location x, y, z. As the following is equivalent for all voxels, we will remove x, y, z and drop the indexing brackets for the purposes of notation. (Also for notation, both the different and gradient notations are used depending on which is more succinct. For total clarity, $\nabla_x y$ is a vector whose entries are all $\frac{\delta y}{\delta z}$ where z is some element of x, which is itself a list or a vector of variables.)

In order to retrieve the values of μ and σ, it is necessary to perform an eigendecomposition of the diffusion matrix:

$$D = R\Lambda R^T, \tag{A1}$$

where R is an orthonormal matrix and Λ is a nonnegative diagonal matrix containing the eigenvalues of the diffusion matrix. Note that this decomposition is not fully unique: it is possible to re-order the elements of Λ and still have a valid matrix. The benefit of μ and σ is that they are invariant to this ordering whereas other diffusion metrics largely depend on having a specific eigenvalue ordering (e.g. comparing the relative sizes of the largest and second-largest eigenvalues).

Given that E is a diagnonal matrix, we can easily derive a simple formula for the elements of D in terms of the rotation matrix and eigenvalues, λ_k, for some ordering:

$$\begin{aligned} d_{i,j} &= R_i \Lambda \left(R_j\right)^T \\ d_{i,j} &= \sum_k r_{i,k}\lambda_k r_{j,k}, \end{aligned} \tag{A2}$$

which yields the derivative:

$$\frac{\delta d_{i,j}}{\delta \lambda_k} = r_{i,k} r_{j,k}. \tag{A3}$$

We can then analytically propagate gradients back through this operation under the assumption that R is constant with respect to λ using the chain rule:

$$\begin{aligned} \frac{\delta N}{\delta \lambda_k} &= \sum_{i,j} \frac{\delta N}{\delta d_{i,j}} \frac{\delta d_{i,j}}{\delta \lambda_k} \\ \frac{\delta N}{\delta \lambda_k} &= \sum_{i,j} \frac{\delta N}{\delta d_{i,j}} r_{i,k} r_{j,k} \end{aligned}. \tag{A4}$$

In order to retrieve the gradients with respect to μ and σ, we propose the following invertible co-ordinate transformation:

$$\begin{aligned} \mu &= \tfrac{1}{3}(\lambda_1 + \lambda_2 + \lambda_3) \\ \sigma &= \sqrt{(\lambda_1 - \mu)^2 + (\lambda_2 - \mu)^2 + (\lambda_3 - \mu)^2} \\ \theta &= \arctan\left(\lambda_1 - \mu, \sqrt{\tfrac{1}{3}}(\lambda_2 - \lambda_3)\right) \\ \lambda_1 &= \mu + \sigma\left(\sqrt{\tfrac{2}{3}}\cos\theta\right) \\ \lambda_2 &= \mu + \sigma\left(-\sqrt{\tfrac{1}{6}}\cos\theta + \sqrt{\tfrac{1}{2}}\sin\theta\right) \\ \lambda_3 &= \mu + \sigma\left(-\sqrt{\tfrac{1}{6}}\cos\theta - \sqrt{\tfrac{1}{2}}\sin\theta\right) \end{aligned}, \tag{A5}$$

which, using the chain rule, yields the derivatives:

$$\begin{aligned} \frac{\delta N}{\delta \mu} &= \frac{\delta N}{\delta \lambda_1} + \frac{\delta N}{\delta \lambda_2} + \frac{\delta N}{\delta \lambda_3} \\ \frac{\delta N}{\delta \sigma} &= \sqrt{\tfrac{2}{3}}\cos\theta \frac{\delta N}{\delta \lambda_1} + \left(\sqrt{\tfrac{1}{2}}\sin\theta - \sqrt{\tfrac{1}{6}}\cos\theta\right)\frac{\delta N}{\delta \lambda_2} \\ &\quad - \left(\sqrt{\tfrac{1}{2}}\sin\theta + \sqrt{\tfrac{1}{6}}\cos\theta\right)\frac{\delta N}{\delta \lambda_3}. \\ \frac{\delta N}{\delta \theta} &= -\sqrt{\tfrac{2}{3}}\sigma\sin\theta \frac{\delta N}{\delta \lambda_1} + \sigma\left(\sqrt{\tfrac{1}{2}}\cos\theta + \sqrt{\tfrac{1}{6}}\sin\theta\right)\frac{\delta N}{\delta \lambda_2} \\ &\quad - \sigma\left(\sqrt{\tfrac{1}{2}}\cos\theta - \sqrt{\tfrac{1}{6}}\sin\theta\right)\frac{\delta N}{\delta \lambda_3} \end{aligned} \tag{A6}$$

An alternative way of computing the gradients is to consider the above co-ordinate transform as a local basis transformation. The benefit of this transformation is that it is a locally orthogonal basis given $\sigma \neq 0$, and thus have orthogonal sensitivities. That is:

$$\nabla_\lambda \mu \cdot \nabla_\lambda \sigma = \nabla_\lambda \mu \cdot \nabla_\lambda \theta = \nabla_\lambda \sigma \cdot \nabla_\lambda \theta = 0. \tag{A7}$$

In order to get an orthonormal basis, we have to ensure that the vectors are all unit length, that is $|\nabla_\lambda \widehat{\mu}| = |\nabla_\lambda \widehat{\sigma}| = \left|\nabla_\lambda \widehat{\theta}\right| = 1$. Note that this is already the case for σ, that is $\widehat{\sigma} = \sigma$ and can be achieved for μ using simple scaling, that is $\widehat{\mu} = \sqrt{\tfrac{1}{3}}\mu$, so the last vector $\nabla_\lambda \widehat{\theta}$ can be found using the cross product: $\nabla_\lambda \widehat{\theta} = \nabla_\lambda \widehat{\sigma} \times \nabla_\lambda \widehat{\mu}$. This gives us a magnitude-preserving expression for the sensitivity of the network:

$$\begin{aligned} \frac{\delta N}{\delta \widehat{\mu}} &= \left[\frac{\delta N}{\delta \lambda_1}\frac{\delta N}{\delta \lambda_2}\frac{\delta N}{\delta \lambda_3}\right]^\top \cdot \left[\frac{1}{\sqrt{3}}\frac{1}{\sqrt{3}}\frac{1}{\sqrt{3}}\right]^\top \quad \left(= \frac{1}{\sqrt{3}}\frac{\delta N}{\delta \mu}\right) \\ \frac{\delta N}{\delta \widehat{\sigma}} &= \left[\frac{\delta N}{\delta \lambda_1}\frac{\delta N}{\delta \lambda_2}\frac{\delta N}{\delta \lambda_3}\right]^\top \cdot \left[\frac{\lambda_1 - \mu}{\sigma}\frac{\lambda_2 - \mu}{\sigma}\frac{\lambda_3 - \mu}{\sigma}\right]^\top \quad \left(= \frac{\delta N}{\delta \sigma}\right). \\ \frac{\delta N}{\delta \widehat{\theta}} &= \left[\frac{\delta N}{\delta \lambda_1}\frac{\delta N}{\delta \lambda_2}\frac{\delta N}{\delta \lambda_3}\right]^\top \cdot \left[\frac{\lambda_3 - \lambda_2}{\sqrt{3}\sigma}\frac{\lambda_1 - \lambda_3}{\sqrt{3}\sigma}\frac{\lambda_2 - \lambda_1}{\sqrt{3}\sigma}\right]^\top \quad \left(= \sigma^{-1}\frac{\delta N}{\delta \theta}\right) \end{aligned} \tag{A8}$$

Unlike μ and σ which are invariant to the ordering of the eigenvalues, θ has an ordering dependency, meaning that a consistent ordering must be used for θ to be aggregated across a population and θ must also be provably continuous at boundary cases where this ordering may be effected by random noise. In the case where the eigenvalues are ascending (i.e. $\lambda_1 \geq \lambda_2 \geq \lambda_3$), θ can be thought of as the *pseudo-planarity*, that is an approximate measure of planarity, which

describes how much of the remaining degree of freedom is used to make two of the three eigenvalues similar. $\theta = 0$ implies that the tensor is as close to being linear (i.e. $\lambda_2 \approx \lambda_3$) as possible given its mean diffusivity and anisotropy. As $\theta \to \pi/3$, the tensor is as planar as possible with $\lambda_1 \approx \lambda_2$ again given the mean diffusivity and anisotropy. As suggested by the trigonometric functions, θ also has meaningful units, radians.

Interestingly, the sensitivity with respect to θ which is neither cyclic nor has an ordering dependence. This means that our method can measure in a population-wise manner, the sensitivity of the classification network with respect to the shape of the tensor not reflected by its overall size (i.e. mean diffusivity) or overall anisotropy, but by a third, completely orthogonal measure.