Received: 27 September 2022 | Revised: 11 May 2023 | Accepted: 2 June 2023

DOI: 10.1002/hbm.26399

TECHNICAL REPORT

WILEY

An automatic interpretable deep learning pipeline for accurate Parkinson's disease diagnosis using quantitative susceptibility mapping and T1-weighted images

Yida Wang[1] | Naying He[2] | Chunyan Zhang[3] | Youmin Zhang[2] | Chenglong Wang[1] | Pei Huang[4] | Zhijia Jin[2] | Yan Li[2] | Zenghui Cheng[2] | Yu Liu[2] | Xinhui Wang[2] | Chen Chen[3] | Jingliang Cheng[3] | Fangtao Liu[2] | Ewart Mark Haacke[2,5] | Shengdi Chen[4] | Guang Yang[1,6,7] | Fuhua Yan[2]

[1]Shanghai Key Laboratory of Magnetic Resonance, East China Normal University, Shanghai, China

[2]Department of Radiology, Ruijin Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China

[3]Department of MRI, The First Affiliated Hospital of Zhengzhou University, Zhengzhou, Henan, China

[4]Department of Neurology, Ruijin Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China

[5]Department of Biomedical Engineering, Wayne State University, Detroit, Michigan, USA

[6]Institute of Brain and Education Innovation, East China Normal University, Shanghai, China

[7]Shanghai Center for Brain Science and Brain-Inspired Technology, Shanghai, China

Correspondence
Fuhua Yan, Department of Radiology, Ruijin Hospital, Shanghai Jiao Tong University School of Medicine, No. 197 Ruijin Er Road, Shanghai 200025, China.
Email: yfh11655@rjh.com.cn

Guang Yang, Shanghai Key Laboratory of Magnetic Resonance, East China Normal University, 3663 N. Zhongshan Road, Shanghai 200062, China.
Email: gyang@phy.ecnu.edu.cn

Jingliang Cheng, Department of MRI, The First Affiliated Hospital of Zhengzhou University, No. 1 Jianshe Dong Road, Er Qi District, Zhengzhou, Henan 450052, China.
Email: fccchengjl@zzu.edu.cn

Abstract

Parkinson's disease (PD) diagnosis based on magnetic resonance imaging (MRI) is still challenging clinically. Quantitative susceptibility maps (QSM) can potentially provide underlying pathophysiological information by detecting the iron distribution in deep gray matter (DGM) nuclei. We hypothesized that deep learning (DL) could be used to automatically segment all DGM nuclei and use relevant features for a better differentiation between PD and healthy controls (HC). In this study, we proposed a DL-based pipeline for automatic PD diagnosis based on QSM and T1-weighted (T1W) images. This consists of (1) a convolutional neural network model integrated with multiple attention mechanisms which simultaneously segments caudate nucleus, globus pallidus, putamen, red nucleus, and substantia nigra from QSM and T1W images, and (2) an SE-ResNeXt50 model with an anatomical attention mechanism, which uses QSM data and the segmented nuclei to distinguish PD from HC. The mean dice values for segmentation of the five DGM nuclei are all >0.83 in the internal testing cohort, suggesting that the model could segment brain nuclei accurately. The

Abbreviations: AG, anatomical gate; ASSD, average symmetric surface distance; AUC, area under curve; CI, confidence interval; CN, caudate nucleus; CNN, convolutional neural network; DL, deep learning; DSC, Dice coefficient per case; FA, flip angle; GP, globus pallidus; Grad-CAM, gradient-weighted class activation mapping; GRE, gradient echo; HC, healthy control; MRI, magnetic resonance imaging; PD, Parkinson's disease; PUT, putamen; QSM, quantitative susceptibility mapping; RN, red nucleus; ROC, receiver operating characteristic; ROI, region of interest; SN, substantia nigra; TE, echo time; TR, repetition time.

Yida Wang and Naying He made equal contributions.

Funding information
National Natural Science Foundation of China, Grant/Award Numbers: 81971576, 82271954, 61731009; National Key R&D Program of China, Grant/Award Number: 2022YFC2009905/2022YFC2009900; Innovative Research Team of High-level Local Universities in Shanghai; Xing-Fu-Zhi-Hua Foundation of ECNU; Microscale Magnetic Resonance Platform of ECNU; Open Project of Shanghai Key Laboratory of Magnetic Resonance

proposed PD diagnosis model achieved area under the the receiver operating characteristic curve (AUCs) of 0.901 and 0.845 on independent internal and external testing cohorts, respectively. Gradient-weighted class activation mapping (Grad-CAM) heatmaps were used to identify contributing nuclei for PD diagnosis on patient level. In conclusion, the proposed approach can potentially be used as an automatic, explainable pipeline for PD diagnosis in a clinical setting.

KEYWORDS
deep learning, Parkinson's disease, quantitative susceptibility mapping, segmentation

1 | INTRODUCTION

Parkinson's disease (PD) is a neurodegenerative movement disorder and has been proven to be associated with abnormal brain iron accumulation (Berg & Youdim, 2006; Kalia & Lang, 2015). The major pathologic changes caused by PD is the loss of dopaminergic neurons from the substantia nigra (SN), which would lead to a decrease in neuromelanin and an increase in iron deposition. The cardinal signs of PD patients are tremor, rigidity, and bradykinesia, all of which can seriously affect the quality of life (Tysnes & Storstein, 2017). Rating scales are used for the evaluation of motor impairment and disability in PD patients, but most of these scales are subjective and may be difficult to use in some hospitals (Jankovic, 2008). Although PD patients with typical symptoms can be diagnosed easily, it is still difficult to distinguish PD with atypical parkinsonism in the early stage due to the similar symptoms. It is very critical to differentiate early-stage PD from other movement disorders for developing effective treatment strategies to improve the life quality of the patient.

Quantitative susceptibility mapping (QSM) is a novel postprocessing technique used to calculate the magnetic susceptibility distribution of tissues in vivo from gradient echo (GRE) magnetic resonance phase images (Haacke et al., 2015; Li & Leigh, 2004). QSM has already been used widely in quantifying in vivo iron content, especially in deep grey matter structures (Haacke et al., 2015). The iron distribution in deep brain nuclei, as shown in QSM images could provide pathophysiological-related information underlying PD (He et al., 2021; Lotfipour et al., 2012). QSM has been shown to be able to detect increases in iron in the deep gray matter (DGM) of the brain in early-stage PD, suggesting that it could potentially serve as an early diagnostic tool for clinical practice (Eskreis-Winkler et al., 2017).

Previous QSM studies in PD diagnosis can be roughly divided into two categories: whole-brain analysis and region of interest (ROI)-based analysis. Whole-brain analysis methods extract high-level contextual features from the 3D imaging data. For example, whole brain T2-weighted magnetic resonance (MR) images have been used as input into the convolutional neural network (CNN) to distinguish PD from healthy controls (HC) and the models achieved good performance (Kaur et al., 2020; Sivaranjini & Sujatha, 2019). However, these methods did not take full advantage of the shape and position priors of brain nuclei structures, possibly leading to a suboptimal performance. The ROI-based methods focus on extracting quantitative features from specific brain nuclei regions such as the SN, red nucleus (RN), globus pallidus (GP), caudate nucleus (CN), and putamen (PUT), according to a priori knowledge of their role in PD diagnosis. For example, a joint feature-sample selection framework (Adeli et al., 2016) was proposed to extract useful features from SN, RN, and other clinical important regions based on T1-weighted (T1W) images for PD diagnosis. Texture analysis in the SN from QSM and R2* maps was utilized to classify PD and HC (Li, Zhai, et al., 2019). A hybrid feature extraction method was introduced for PD diagnosis based on the SN as seen in QSM images (Xiao et al., 2019). Another group used a histogram analysis of the QSM data for the SN and PUT which yielded the highest diagnostic performance (Zhang et al., 2022). While the pathological manifestations of PD in the SN are relatively clear, there is still inconsistency about the role of iron measurements throughout the DGM regions of the brain.

These ROI-based studies (Li, Zhai, et al., 2019; Xiao et al., 2019) relied on manually delineated ROIs for downstream analysis, which would make their application time-consuming and expensive in the real-world clinical environment. To address this challenge, it is highly desirable to develop a fully automatic model to segment brain nuclei regions. Though QSM images can provide superior image contrast between iron-rich subcortical nuclei and their surroundings (Cobzas et al., 2015), the CN and PUT boundaries are not always clearly visible on QSM images. However, T1W images usually do a good job to show the CN and PUT boundaries, implying that multimodal magnetic resonance imaging (MRI) images (T1W and QSM) can be used for better automatic segmentation of multiple brain nuclei (Jin et al., 2022).

Recently, deep learning (DL) methods have shown superior performance in medical imaging applications (Suganyadevi et al., 2022). CNN can automatically learn discriminative and representative features from large annotated datasets to address complex computer vision problems. Well-known CNN networks such as VGG-Net (Simonyan & Zisserman, 2014), ResNet (He et al., 2016), ResNeXt (Xie et al., 2017), and U-Net (Ronneberger et al., 2015), together with their variants, have been used widely in medical imaging for lesion detection, segmentation, and diagnosis (Suganyadevi et al., 2022). It has been demonstrated that CNNs could enhance the efficacy of clinicians in image-based diagnosis. However, due to the nonlinear structure of CNN, most existing models are considered as black boxes whose internal inference processes are hard to interpret, which, in turn, makes it hard for clinicians to confirm the models' predictions

and thus limits their acceptance in clinical practice (Riccardo et al., 2018).

To address the problem of lack of interpretability, gradient-weighted class activation mapping (Grad-CAM) (Selvaraju et al., 2020) has been proposed to generate visual explanations for the model output by visualizing discriminative regions on the input images. An attention mechanism approach could also increase the performance and explainability of the network. Different types of attention mechanisms have been integrated into CNN and achieved some success in medical imaging applications (Niu et al., 2021). Squeeze and excitation (SE) block (Hu et al., 2017) provides channel attention that could generate an attention mask across channels and highlight important channels. Spatial attention could help CNN models focus on the most relevant regions in the feature maps and ignore irrelevant background and can be seen as an adaptive spatial region selection mechanism (Du et al., 2018).

Therefore, we hypothesized that it might be useful to automatically segment all candidate nuclei instead of one or two specified nuclei and use features from all these nuclei with attention mechanism to improve the diagnosis of PD, and in the process, reveal those deep gray nuclei relevant to PD at patient level with the help of Grad-CAM.

In this study, we present a cascade DL pipeline for PD diagnosis with MR images. Five brain nuclei regions were analyzed by our segmentation model before being used as a priori information by an anatomical attention-guided CNN model for PD diagnosis. Grad-CAM heatmaps were calculated to highlight the regions relevant to the prediction. The experimental results showed that the anatomical attention mechanism provided by the automatic segmentation not only improved the performance of the classification model, but also increased its interpretability by generating more meaningful heatmaps. Quantitative and qualitative comparisons with other DL methods demonstrated that our approach achieved highly accurate and reliable PD diagnosis performance, and with minimal human interaction, it could achieve near-perfect performance.

2 | MATERIALS AND METHODS

2.1 | Subject recruitment, MRI scanning, and processing

Two different datasets, Dataset 1 ($n1 = 379$) and Dataset 2 ($n2 = 155$), were used in our study. In Dataset 1, 92 PD patients and 287 HC were prospectively recruited from Ruijin Hospital. In Dataset 2, 83 PD patients and 72 HC were collected from the First Affiliated Hospital of Zhengzhou University. This study was approved by the local ethics committees and informed consent was signed by each subject at both institutions. All the PD subjects were on medication. The inclusion and exclusion criteria were the same as given in a previously published study (He et al., 2021). In brief, the diagnosis and severity of motor symptoms were assessed by two neurologists based on the Movement Disorders Society (MDS) PD criteria (Berg et al., 2018), the Hoehn and Yahr (H&Y) stage and the MDS Unified Parkinson's Disease Rating Scale Part-III (UPDRS-III) (Goetz et al., 2008). The exclusion criteria for PD patients were as follows: (1) H&Y scale more than 2.5; (2) Mini-Mental State Examination (MMSE) score <24; (3) a history of cerebrovascular disease, brain tumor, head trauma, or any other type of psychiatric disorders; (4) a history of medication known to cause parkinsonism or affect clinical assessment; (5) contraindications to an MRI examination, and (6) younger than 40 years of age. The inclusion criteria for the HCs were as follows: (1) no family history of movement disorders; (2) no neurological or psychiatric disorders; (3) an MMSE score of at least 24 and (4) older than 40 years of age. PD subjects were evaluated 1 year later to confirm the diagnosis with the same MDS-PD diagnostic criteria used at the baseline assessment.

Dataset 1 was collected on a 3T scanner (Ingenia, Philips Healthcare, Netherlands) with a 15-channel phased array head coil using a 3D GRE imaging sequence. All scanning was done parallel to the anterior commissure-posterior commissure line. T1W and QSM were derived from the STrategically Acquired Gradient Echo (STAGE) imaging data (Chen et al., 2018). STAGE could provide similar results for any manufacturer's system thereby providing standardization of brain imaging (Haacke et al., 2020). It is more stable and consistent across different scanners. Therefore, we used STAGE image data in this study. The imaging parameters were as follows: echo times (TEs) = 7.5 and 17.5 ms, repetition time = 25 ms, flip angles (FA) = 6° and 24°, in-plane resolution = 0.67 mm × 1.34 mm (interpolated to 0.67 mm × 0.67 mm), pixel bandwidth = 220 Hz/pixel, imaging matrix size = 384 × 384, slice thickness = 2 mm, 64 slices and a scan time = 1 min 53 s for each FA.

Dataset 2 was collected on a 3T scanner (Prisma, Siemens Healthcare, Erlangen, Germany) with a 64-channel phased array head coil using a 3D GRE imaging sequence. T1W and QSM were also derived from the STAGE imaging data. The STAGE protocol included two fully flow-compensated multiecho susceptibility-weighted imaging scans. One scan used a 6° FA with two TEs of 7.5/17.5 ms, and the other used a 24° FA with two TEs of 8.75/18.75 ms. Other imaging parameters were: in-plane resolution = 0.67 mm × 1.34 mm, imaging matrix size = 384 × 288, slice thickness = 2 mm, 64 slices, and an iPAT factor = 2. The total acquisition time was 5 min 8 s.

The QSM data were created using the second echo (TE = 17.5 ms) of the STAGE data since the longer TE has better signal-to-noise in mapping the iron content. The QSM processing was done as follows: (1) Brain extraction (BET) was performed using the BET tool on the first echo magnitude images with a threshold of 0.2, an erosion factor of 2, and an island factor of 2000 (Smith, 2002). (2) 3DSRNCP (3D sorting following a noncontinuous path) was used to unwrap the phase (Abdul-Rahman et al., 2007); (3) SHARP (sophisticated harmonic artifact reduction processing) was used to remove background phase (Schweser et al., 2011); and (4) iterative QSM was used with four iterations and a threshold of 0.1 to reconstruct the susceptibility maps (Haacke et al., 2010; Tang et al., 2013).

In Dataset 1, one neuroradiologist segmented the CN, GP, PUT, RN, and SN in bilateral hemispheres manually. Another radiologist

with 5 years of experience in neuroimaging double checked all the structural boundaries. Finally, those ROIs approved by these two radiologists were used as the ground truth for the segmentation model.

2.2 | Data preparation

Dataset 1 was randomly split into a training cohort (74 PD/230 HC) and an independent testing cohort (18 PD/57 HC). Dataset 2 (83 PD/72 HC) was used as an external testing cohort to assess the performance and generalizability of the diagnostic model. The details of the basic characteristics of the subjects in the three cohorts were summarized in Table 1. Each subject was normalized to [0, 1] before contrast-limited adaptive histogram equalization (CLAHE) (Reza, 2004) was used to enhance and harmonize local image contrast. Since all images for a subject were derived from a single STAGE acquisition, there was no need for registration of QSM and T1W images.

2.3 | Model structure

For accurate PD diagnosis, we proposed a two-stage DL framework, which consisted of (1) a segmentation model for brain nuclei segmentation, and (2) a classification model for PD diagnosis. In the first stage, we trained a segmentation model using QSM and T1W images as input to automatically segment five deep gray nuclei. In the second stage, QSM images and the corresponding segmented brain nuclei regions were simultaneously fed into a classification model for PD diagnosis. The proposed classification network contained two branches, the first one was used to extract discriminative feature representation from QSM images, and the second one was used to capture the anatomical priori information from the segmented brain nuclei. The attention gate fused feature maps from two branches for better PD diagnosis. The details about the two DL models and attention gate were presented in the following section.

A modified 3D implementation of CA-Net (Gu et al., 2021) was utilized to segment brain nuclei on QSM and T1W images. Brain nuclei exhibit clearer edges on QSM and T1W images, so QSM and T1W images were concatenated in the channel dimension and fed into the CA-Net to obtain all the five nuclei. In the CA-Net, spatial attention, channel attention, and scale attention modules were integrated into U-Net at the same time for more accurate and explainable medical image segmentation. Spatial attention forced the network to focus on the region of interest. Channel attention adjusted weights of different channels of feature maps so that the most relevant feature channels will be highlighted. Scale attention enhanced the most relevant features among different scales so that the network can segment objects of different sizes.

The structure of classification model adopted to distinguish PD from HC in this work was shown in Figure 1. The proposed network consisted of a 3D SE-ResNeXt50 branch and an anatomical attention branch, and was thus called AG (attention gated)-SE-ResNeXt50. In the anatomical attention module, the segmented brain nuclei were used as priori information to make the network extract features in the useful brain structures for better PD diagnosis. The cropped 3D QSM images were fed into the 3D SE-ResNeXt50 branch which was composed of two convolution layers, a 3D max pooling layer, SE-ResNeXt bottleneck blocks, a global max pooling layer, and a fully connected layer. SE-ResNeXt bottleneck block integrated SE module into ResNeXt bottleneck. Compared with the standard ResNet bottleneck, ResNeXt used a grouped convolution layer to replace the traditional convolution operation, which improved the accuracy and controlled the model complexity and number of parameters.

In the anatomical attention branch, two convolution layers and a 3D max pooling layer were used to extract features from the regions of the five brain nuclei segmented by CA-Net. The subsequent convolution layers with the kernel size of $1 \times 1 \times 1$ and the stride of 2 were used to keep the feature size the same as the image feature size in the SE-ResNeXt50 branch. The anatomical gate (AG) (Sun et al., 2020) fused feature maps on different scales from two branches in the channel dimension. Then, the fused feature maps went through two parallel $1 \times 1 \times 1$ convolution layers followed by a sigmoid layer to get the two weighted maps and multiplied with original feature maps. Finally, the two weighted feature maps were added together to get the anatomical-guided feature maps.

2.4 | Model training

Five-fold cross-validation was used for model training and selection in the training cohort. For each fold, we selected the best model according to the model's performance on the cross-validation dataset. We

TABLE 1 Basic characteristics of the subjects in the training, internal testing, and external testing cohorts.

	Training cohort		Internal testing cohort		External testing cohort	
	PD ($n = 74$)	HC ($n = 230$)	PD ($n = 18$)	HC ($n = 57$)	PD ($n = 83$)	HC ($n = 72$)
Age (years)	63.27 ± 7.72	62.83 ± 6.54	63.44 ± 7.06	62.02 ± 7.35	59.38 ± 9.14	57.86 ± 9.46
Gender (M/F)	36/38	80/150	10/8	23/34	35/44	37/35
Disease duration (years)	3.42 ± 2.39		3.95 ± 3.30		3.48 ± 3.07	
UPDRS-III	26.06 ± 17.89		30.10 ± 15.71		23.71 ± 13.83	

Note: *Values* are represented in the form mean ± standard deviation. The UPDRS-III was administered while subjects were in the ON-medicated state.
Abbreviations: HC, healthy controls; PD, Parkinson's disease; UPDRS-III, Unified Parkinson's Disease Rating Scale Part-III.

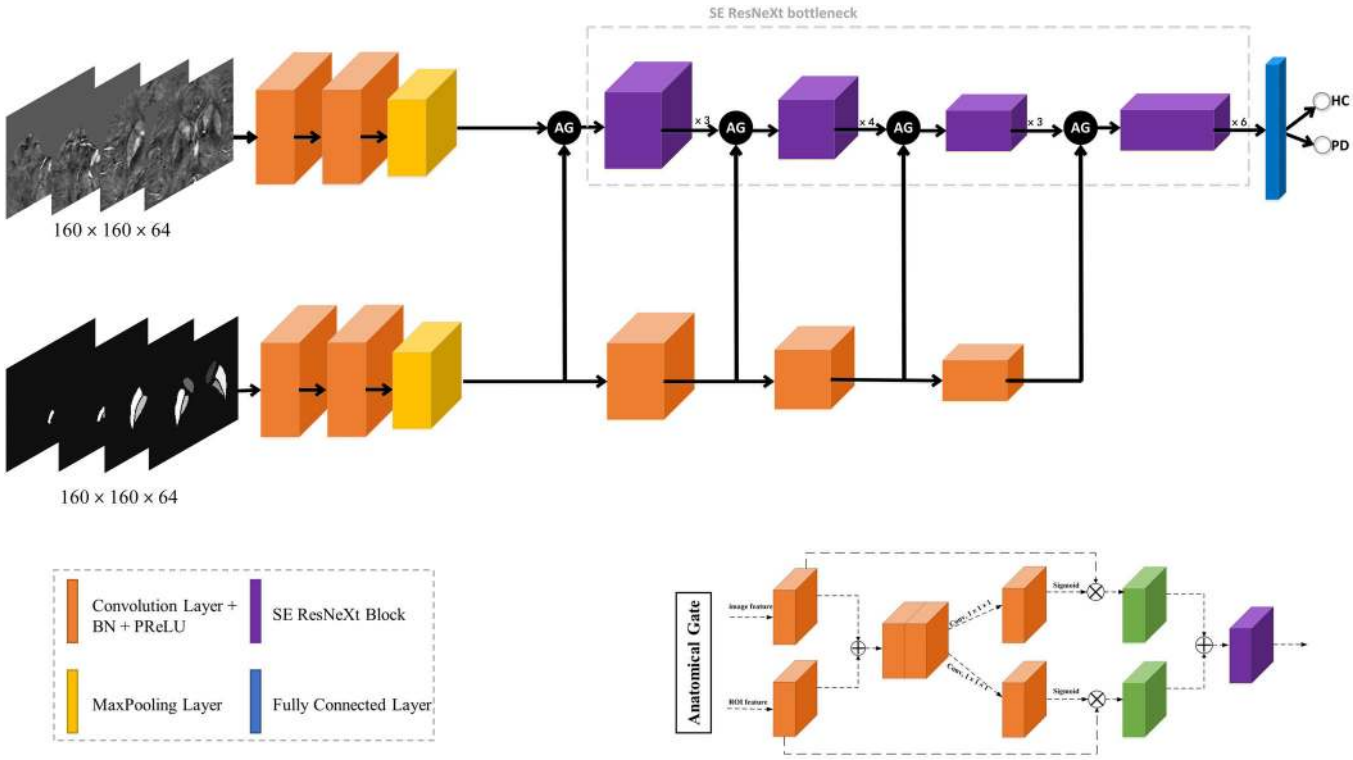

FIGURE 1 Architecture of the classification model. Anatomical gate (bottom right) was used to fuse feature maps generated by two branches to obtain anatomical-guided feature maps. AG, anatomical gate; HC, healthy controls; PD, Parkinson's disease; SE, squeeze and excitation.

integrated the results by averaging the predicted probability of the five trained models to obtain the final segmentation or classification results. During the training process, we repeated random PD cases to remove the imbalance of samples in the training data. To increase the robustness of the model, we used an online data augmentation strategy including random shifting within ±10 pixels, rotating within ±5°, and shearing from 0.8 to 1.2, vertical and horizontal flipping, elastic transform from 0 to 0.1 and gamma transform from 0.9 to 1.1. We center cropped the original QSM and T1W images to $160 \times 160 \times 64$ image patches which covered all five brain nuclei, and used them as the network input.

For training of both segmentation and classification models, we used focal loss (Lin et al., 2017) as the loss function, which can be described as:

$$\text{Focal Loss}(p,p_t) = -(\alpha(1-p_t)^{\gamma} p \log(p_t) + (1-\alpha)p_t{}^{\gamma}(1-p)\log(1-p_t))$$

where p and p_t represents label and predicted probability. Hyperparameters α and γ were set to 0.25 and 2 in the experiments. We used the Adam optimizer with momentum of 0.9, a weight decay of 0.0001, and an initial learning rate of 10^{-4} to minimize the loss. The batch size was set to 16. The training process was stopped if the loss on the validation dataset did not decrease over 30 epochs. The DL models were implemented in Python 3.8.0, using PyTorch 1.9.0. Model training was conducted on a workstation equipped with three NVIDIA A100 graphic processing units (GPU) with 40 GB of graphic memory each.

To find the best model for PD diagnosis, and demonstrate the influence of different input on PD diagnosis, we trained a series models using different combinations of input, including (1) 3D versions of VGG16 and ResNet50 models using QSM images as input; (2) SE-ResNeXt50 network using QSM images only, QSM and T1W images, and QSM images masked with the corresponding segmented brain nuclei regions, respectively. These models were named using a "network-input" format, such as VGG16-QSM. The results of the above models were compared with the proposed model. Considering the article length, we included the details of these experiments in the Supplementary material S1. To evaluate the influence of errors in the automatic segmentation on the diagnosis, we conducted two experiments: (1) manual outlined ROIs were used when evaluating the proposed diagnostic model and (2) manual-outlined ROIs were used to train and evaluate a new instance of proposed diagnostic model. All the models were trained with the same DL settings and the same training and testing cohorts in Dataset 1 as mentioned above.

2.5 | Assessment of model predictions

The Dice coefficient per case (DSC) and average symmetric surface distance (ASSD) were used to measure the agreement between the

manual delineation (ground truth) and the predicted segmentation. The DSC and ASSD are defined as:

$$\mathrm{DSC}(P,R)=\frac{2\,|P\cap R|}{|P|+|R|}$$

$$\mathrm{ASDD}=\frac{1}{|P|+|R|}\left(\sum_{x\in S_P} d(x,S_R)+\sum_{y\in S_R} d(y,S_P)\right)$$

where P, R are the predicted results and ground truth, respectively, and S_P, S_R denote their corresponding surfaces. Here, (x, y) denotes a voxel on the surface of P and R, respectively and d denotes the distance from a voxel to the surface. Higher DSC values and smaller ASSD values correspond to better segmentation results.

The performance of PD diagnosis was assessed by the receiver operating characteristic (ROC) analysis and confusion matrix. Evaluation metrics including accuracy, area under the ROC curve (AUC), sensitivity, specificity, positive predictive value, negative predictive value, precision, recall, and F1 score were calculated at a cutoff value that maximized the Youden index in the training cohort. The difference of the predicted probability between PD and HC was assessed with Wilcoxon test and p-value < .001 was regarded as statistically significant. Statistical analyses were performed with R software (version 4.0.4, http://www.R-project.org).

3 | RESULTS

3.1 | Segmentation

Overall, the segmentation from the CA-Net model performed well in terms of the DSC values and ASSD values (Table 2). The mean DSC values (Figure 2) for the five brain nuclei were all greater than 0.83 with a small standard deviation, indicating that the model achieved stable segmentation results. Figure 3 demonstrated for two cases that the CA-Net was able to segment the five regions well compared with the manual drawings. These boundaries could then be used to assess volume or average susceptibility for example. However, for a small number of cases, some segmented nuclei may exhibit visually discernable differences from the ground truth, as illustrated in the CN Case 2 in Figure 3 where the CA-Net placed a boundary in a location not drawn by the raters. To calculate the inter–rater variability, a third radiologist manually drew the ROIs in 30 cases randomly selected from the testing cohort. The mean values of Dice between two raters were 0.849, 0.860, 0.860, 0.870, and 0.833 for CN, GP, PUT, RN, and SN, respectively.

3.2 | Classification

To evaluate the classification performance of the proposed AG-SE-ResNeXt50 model, we plotted the ROC curves, the distribution of predicted probabilities, and the confusion matrix over both the training and internal testing cohort (Figure 4). The proposed model demonstrated promising performance with AUC values of 0.970 and 0.901 on the training and internal testing cohorts, respectively. Significant differences were found between the predicted probabilities of PD and HC cohorts (Wilcoxon, p < .001). When using manually segmented ROIs in the classification, AG-SE-ResNeXt50 achieved AUC values of 0.994 and 1.0 over the training and internal testing cohort, respectively.

The proposed model achieved an AUC of 0.901, significantly higher than those of VGG16-QSM (0.422), ResNet50-QSM (0.765), SE-QSM (0.784), SE-QSM/T1 (0.709), and SE-maskedQSM (0.856) (Figure 5) in the internal testing cohort. As illustrated in Table 3, while VGG16-QSM failed for the classification task, ResNet50-QSM, SE-QSM, SE-QSM/T1, and SE-maskedQSM achieved acceptable performance with accuracies of 66.7%, 70.7%, 58.7%, and 88.0%,

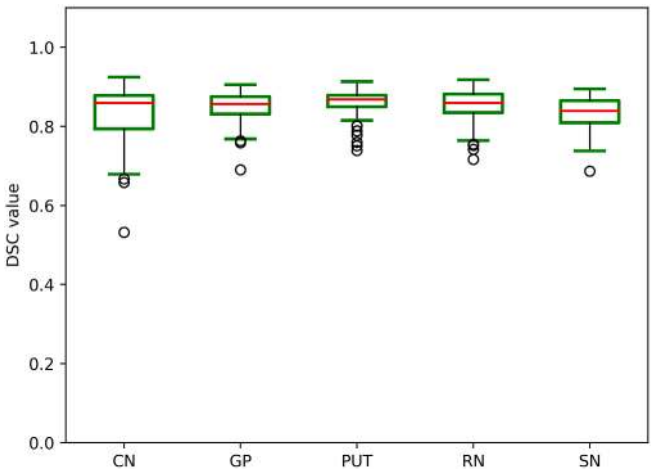

FIGURE 2 Box-plot showing the distribution of dice coefficient per case (DSC) scores in the caudate nucleus (CN), globus pallidus (GP), putamen (PUT), red nucleus (RN), and substantia nigra (SN) over the internal testing cohort.

TABLE 2 Metrics of segmentation of all nuclei in the internal testing cohort.

	CN	GP	PUT	RN	SN
DSC	0.831 ± 0.071	0.849 ± 0.037	0.858 ± 0.034	0.852 ± 0.041	0.831 ± 0.041
ASSD (mm)	0.663 ± 0.430	0.509 ± 0.110	0.531 ± 0.104	0.281 ± 0.068	0.251 ± 0.074

Note: All values are represented in the form of mean ± standard deviation.
Abbreviations: ASSD, average symmetric surface distance; CN, caudate nucleus; DSC, dice similarity coefficient; GP, globus pallidus; PUT, putamen; RN, red nucleus; SN, substantia nigra.

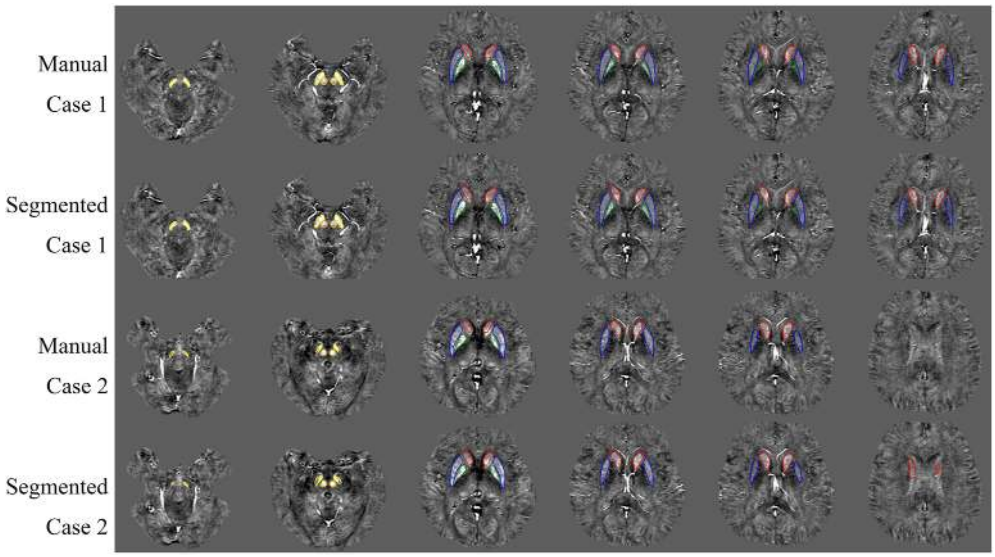

FIGURE 3 Visual comparison of segmented nuclei of two PD patients with the ground truth. The top two rows represent the manual drawing and segmented ROIs of Case 1, with dice coefficient per case (DSC) values for caudate nucleus (CN), globus pallidus (GP), putamen (PUT), red nucleus (RN), and substantia nigra (SN) are 0.838, 0.905, 0.874, 0.893, and 0.861, respectively, while the bottom two rows illustrated Case 2, with DSC values for CN, GP, PUT, RN, and SN are 0.679, 0.890, 0.883, 0.809, and 0.863, respectively. The red, green, blue, orange, and yellow contours outline CN, GP, PUT, RN, and SN regions, respectively.

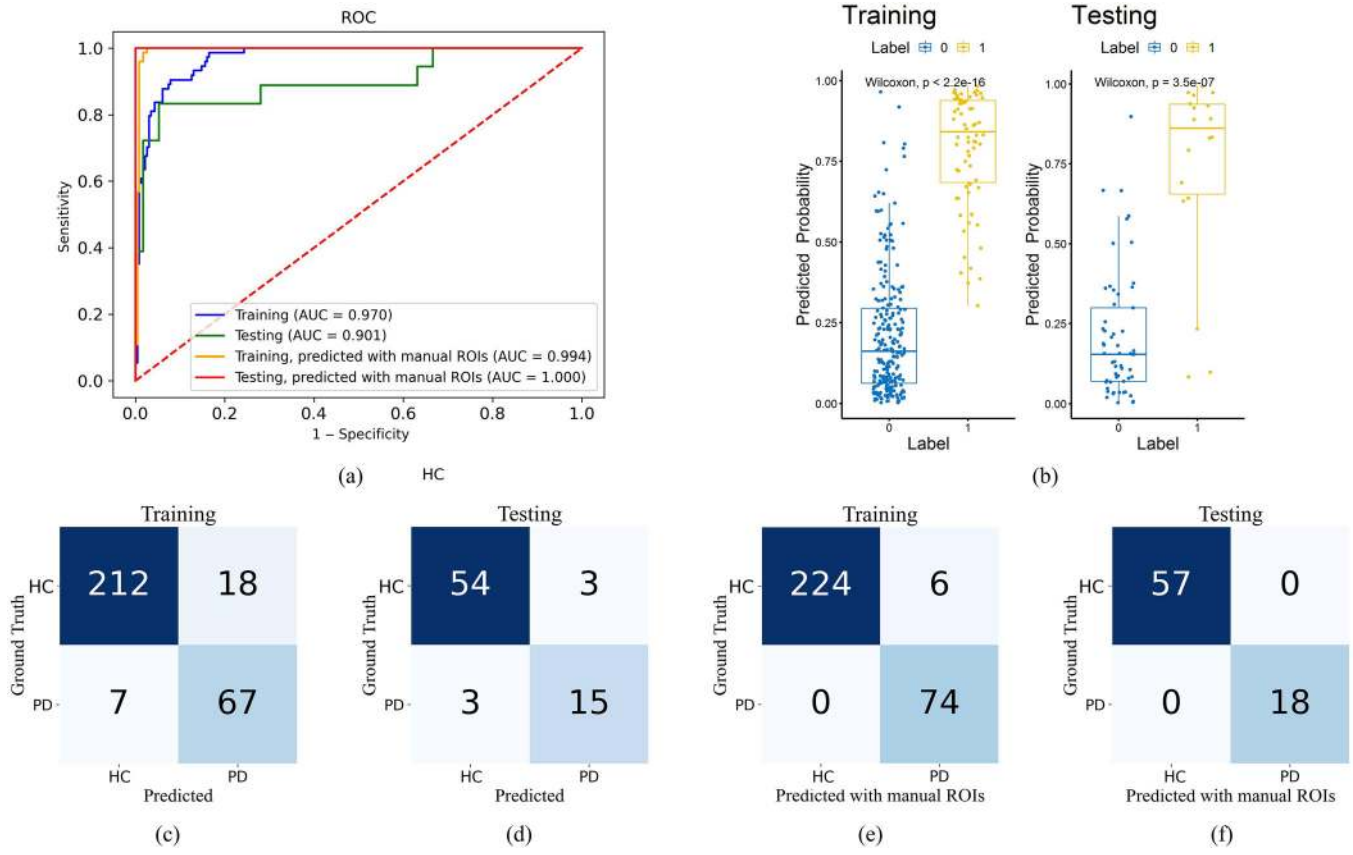

FIGURE 4 Parkinson's disease (PD) diagnosis results (a) the receiver operating characteristic (ROC) curves predicted with automatic segmented and manual region of interests (ROIs) in the training and internal testing cohorts; (b) The probability distribution of the predicted results using automatic segmented ROIs in the training and internal testing cohorts; (c–f) The confusion matrix of the classification model predicted with automatic segmented and manual ROIs on the training and internal testing cohort. PD patients/healthy controls (HC) were treated as positive/negative, respectively.

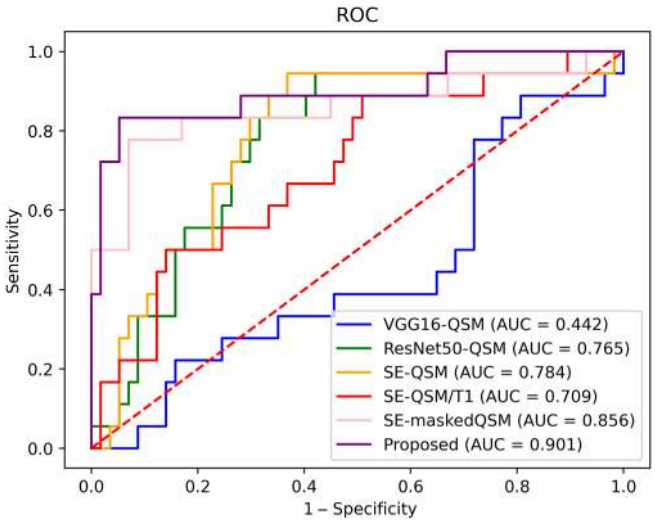

FIGURE 5 Receiver operating characteristic (ROC) curves of different models for Parkinson's disease (PD) diagnosis over the internal testing cohort. AUC, area under the ROC curve; QSM, quantitative susceptibility mapping; SE, squeeze and excitation.

respectively. The proposed AG-SE-ResNeXt50 model achieved a much higher accuracy of 92.0%, and the proposed diagnostic model tested with manually segmented ROIs achieved the highest accuracy of 98.7%. In the independent external testing cohort, the proposed model achieved an AUC of 0.845, an accuracy of 0.787, a sensitivity of 0.771, and a specificity of 0.806 (Table 3) and the ROC curve and confusion matrix were shown in Figure 6.

To reveal the region the AG-SE-ResNeXt50 model focused on during PD diagnosis, the Grad-CAM technique was used to visualize the activation maps for two PD patients in the testing cohort, and the resulting heatmaps were overlaid onto the QSM images as shown in Figure 7. The top row shows a true positive case (Case 1 in Figure 3), whereas the second row and the third row illustrate a positive case (Case 2 in Figure 3) predicted incorrectly with the automatically segmented ROI (second) but correctly with the ground truth ROI (third).

4 | DISCUSSION

In this study, we proposed an automatic DL model which used anatomical priori information to distinguish PD from HC subjects using QSM and T1W images. The segmentation model based on CA-Net integrated comprehensive attention modules, including spatial attention, channel attention, and scale attention into U-Net, so that it can segment five brain nuclei regions with various shapes and sizes simultaneously and accurately. The classification model, namely AG-SE-ResNeXt50, which used QSM images and the segmented brain nuclei regions to diagnose PD, achieved an AUC of 0.901 and 0.845 on the independent internal and external testing cohorts and provided good interpretability.

Quantitative and visual comparison showed that the trained CA-Net model achieved accurate segmentation results. The proposed CA-Net model contained multiple mechanism attentions and had advantages in dealing with segmentation targets with a variation of scales and shape. The accuracy of segmentation ensured the resulting ROIs could be used to provide anatomical quantitative information for DGM volumes and iron content for potential PD classification. Before DL achieved robust results in medical image segmentation, atlas-based segmentation had been widely used in DGM nuclei segmentation, often combined with other classic image segmentation algorithms, such as LevelSet, for better results. However, segmentation based on DL can easily produce better results, often in much less time. For example, our segmentation model achieved a DSC value of 0.83 for SN, higher than those reported in previous atlas-based studies (Garzon et al., 2018; Guo et al., 2018; Li, Chen, et al., 2019) with range of 0.77–0.81. Further, these classic atlas-based methods (Li, Chen, et al., 2019; Lim et al., 2013) required much longer computational time (~1 h) to segment each instance than DL-based model (~3 min). Generally, building DL model is straightforward and once built, the models are often faster and can achieve better results than classic segmentation algorithms.

There were also many studies using DL for nuclei segmentation and PD diagnosis on neuromelanin-sensitive (NM-MRI) images. For example, several studies (Gaurav et al., 2022; Krupicka et al., 2019; Le Berre et al., 2019) used U-Net for SN segmentation on neuromelanin-sensitive and achieved DSC values near 0.80. NM-MRI can be used to delineate the boundaries of structures such as SNpc, ventral tegmental area, and locus coeruleus, and to visualize the NM contents in these structures. However, it is not ideal for segmentation of other DGM nuclei. PD diagnosis can also be performed on NM-MRI, from perspective of NM distribution. For example, Shinde et al. (2019) used a bounding box around the brain-stem created manually on the axial slices of the NM-MRI as input to the 2D ResNet50 for PD diagnosis and achieved an accuracy of 80%.

Different from NM-MRI, QSM can visualize clearly the contour of ferrous DGM nuclei, such as CN, PUT, GP, SN, and RN. It can also be used to quantify the iron deposition in these structures. For PD patients, abnormal iron deposition in DGM nuclei, especially SN, is known to be a sign with relatively high sensitivity and specificity, and observable with medical imaging. Compared with the above studies using NM-MRI, our model could simultaneously segment five brain nuclei using QSM and T1W images and achieved a higher mean DSC value (0.83). Chai et al. (2022) have also simultaneously segmented five nuclei with a contrast attention U-Net on T1W and QSM images, and achieved comparable DSC values for the segmentation of CN, GP, PUT, and RN, but much lower DSC than our study (0.71 vs. 0.83) for SN segmentation. Different to this work, our segmentation model employed a scale attention mechanism that could automatically learn adaptive weights to calibrate the feature maps at different scales and combine features of different scales for accurate segmentation of all nuclei of different sizes. Generally, NM-MRI and QSM have their own respective advantages in PD diagnosis and we will utilize both of them to improve the diagnostic performance of PD in the future.

In the proposed AG-SE-ResNeXt50 model, grouped convolution could greatly reduce the number of parameters while ensuring

TABLE 3 Comparison of the performance of different models for PD diagnosis.

				Training	Testing								
	Input	Network	AG	AUC (95% CI)	AUC (95% CI)	ACC	SEN	SPE	PPV	NPV	PRE	REC	F1
Internal													
VGG16-QSM	QSM	VGG16	w/o	0.594 (0.571–0.661)	0.422 (0.295–0.611)	0.360	0.889	0.193	0.258	0.846	0.258	0.889	0.400
ResNet50-QSM	QSM	ResNet50	w/o	0.851 (0.803–0.897)	0.765 (0.631–0.878)	0.667	0.944	0.579	0.415	0.971	0.415	0.944	0.576
SE-QSM	QSM	SE-ResNeXt50	w/o	0.907 (0.871–0.939)	0.784 (0.660–0.910)	0.707	0.944	0.632	0.447	0.973	0.447	0.944	0.607
SE-QSM/T1	QSM, T1	SE-ResNeXt50	w/o	0.873 (0.832–0.912)	0.709 (0.571–0.842)	0.587	0.889	0.491	0.356	0.933	0.356	0.889	0.508
SE-maskedQSM	masked QSM	SE-ResNeXt50	w/o	0.915 (0.883–0.946)	0.856 (0.694–0.968)	0.880	0.778	0.912	0.737	0.929	0.737	0.778	0.757
Proposed	QSM, ROI	AG-SE-ResNeXt50	w	0.970 (0.952–0.986)	0.901 (0.789–0.982)	0.920	0.833	0.947	0.833	0.947	0.833	0.833	0.833
Tested with GT-ROI				**1.000** (1.000–1.000)	0.984 (0.943–1.000)	0.987	**1.000**	0.983	0.952	**1.000**	0.952	**1.000**	0.976
Proposed/GT-ROI	QSM, GT-ROI	AG-SE-ResNeXt50	w	0.994 (0.984–1.000)	**1.000** (1.000–1.000)	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**
External													
Proposed	QSM, ROI	AG-SE-ResNeXt50	w	NA	0.845 (0.783–0.903)	0.787	0.771	0.806	0.821	0.753	0.821	0.771	0.795

Abbreviations: ACC, accuracy; AG, anatomical gate; AUC, area under ROC curve; CI, confidence interval; F1, F1 score; GT-ROI, ground-truth region of interest; NPV, negative predictive value; PD, Parkinson's disease; PPV, positive predictive value; PRE, precision; REC, recall; ROC, receiver operating characteristic; ROI, segmented region of interest; SE, squeeze and excitation; SEN, sensitivity; SPE, specificity; w, with; w/o, without. The bold values indicate the highest value in the column.

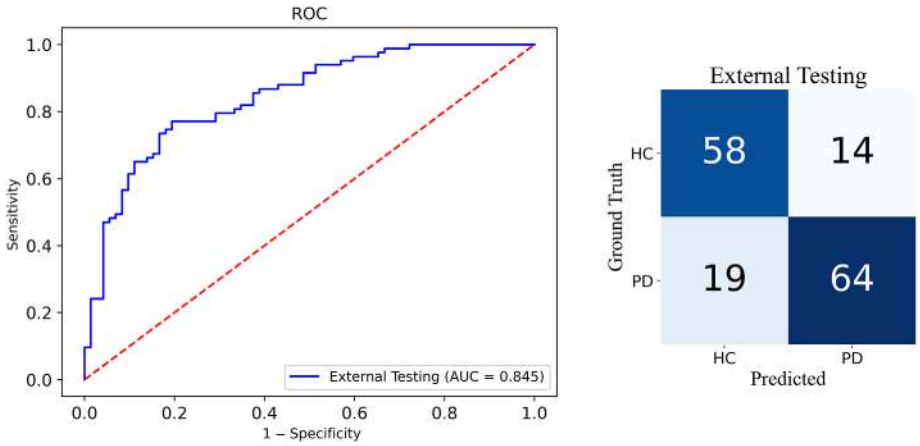

FIGURE 6 Receiver operating characteristic (ROC) curve (left) and confusion matrix (right) for Parkinson's disease (PD) diagnosis over the external testing cohort. AUC, area under the curve; HC, healthy controls.

effectiveness of the model (Xie et al., 2017) and the SE module could learn to use global information to selectively discriminate features and suppress less useful ones among channels (Hu et al., 2017). The center-cropped QSM and T1W images covered all five nuclei regions and were used as network input, which could remove irrelevant background, reducing the memory usage while making full use of the 3D structure all the nuclei. The results demonstrated the effectiveness of our proposed DL-based model integrated with anatomical attention mechanism for structural definition and PD diagnosis. It can be seen clearly from the Grad-CAM heatmaps in Figure 7 that the SN, RN, and CN regions are highlighted, indicating that model used discriminative features from these relevant nuclei for PD diagnosis. However, for Case 2 which was incorrectly predicted by the proposed model using segmented ROIs, SN and RN are not highlighted, implying that (1) the other nuclei, such as CN might contribute relatively more in this case, compared with SN and RN, and (2) inaccurate segmented CN influenced the diagnosis of this case, where CN contribute more to the diagnosis of PD. Thus, with the help of Grad-CAM activation maps, we could potentially know the relative contribution of different deep gray nuclei to the PD diagnosis on an individual patient level.

It is worth noting that when the proposed model was used to predict PD status with ground truth ROIs, it achieved AUC values of 1.000 and 0.984 over the training and internal testing cohorts, respectively, which is even better than the training AUC. A close inspection on all the six test cases that were incorrectly classified by the proposed model using model-segmented ROIs revealed that each case had at least one nucleus with a DSC value <0.70. This result has several important implications: (1) using inaccurate ROIs for anatomical attention will negatively influence the classification; (2) even with inaccurate ROIs, the proposed classification model can be trained to learn distinct features generalizable to predict when using ground truth ROIs correctly; and (3) with minimal human interaction to revise the poorly segmented nuclei, a near-perfect diagnosis of PD can be achieved.

We also compared the performance of the proposed classification model with different models and input data: VGG16-QSM, ResNet50-QSM, SE-QSM, SE-QSM/T1, and SE-maskedQSM. Among these models, the proposed model achieved the best performance and the results demonstrated that using anatomical attention provided by brain nuclei segmentation can help to improve PD diagnosis. The DGM nuclei are known to be important to PD diagnosis (Eskreis-Winkler et al., 2017), and features extracted from these ROIs are known to be relevant to the diagnosis (Adeli et al., 2016; Rana et al., 2015; Xiao et al., 2019). Therefore, we see that using the information derived from the DGM nuclei to provide anatomical attention could greatly improve the efficiency of learning useful features and promote the model performance.

Previous studies about PD diagnosis based on classic machine learning (Adeli et al., 2016; Rana et al., 2015; Xiao et al., 2019) employed handcrafted features to represent information about the ROIs. In our case, these ROIs were obtained from manual delineation by several radiologists, which was burdensome and costly in human effort and time. In general, these results will be dependent on the experience of the radiologists. The proposed method automatically segmented brain nuclei regions to provide anatomical information for PD diagnosis, so it eliminated the manual workload and achieved an accuracy of 92.0%, higher than those of previous studies in the range of 80% to 90%. It worth noting that the proposed diagnosis model using the manual-outlined ROIs outperformed the one using automatic segmentation (98.7% vs. 92.0%), implying that (1) a better segmentation model may be used to further improve the performance of proposed pipeline, and (2) when anatomical attention is combined with accurate segmentation, near-perfect performance can be achieved in PD diagnosis.

While the proposed method has achieved promising results, there are also some limitations in this work. First, although brain nuclei regions could provide anatomical priori knowledge, inaccurate segmentation may potentially influence the classification performance. In future work, we will design a multi-task architecture (Vandenhende et al., 2021) to perform joint brain nuclei segmentation and PD diagnosis. Second, the datasets were collected from two centers, but the number of the testing cohort was small. Therefore, it is desirable to collect more multi-center and multi-vendor MRI data for further evaluation of the performance of the proposed method.

In summary, in this work, we proposed a fully automated approach for PD diagnosis based on QSM and T1W images, which achieved superior performance relative to previous studies and can potentially be used to help clinicians distinguish PD from HC with

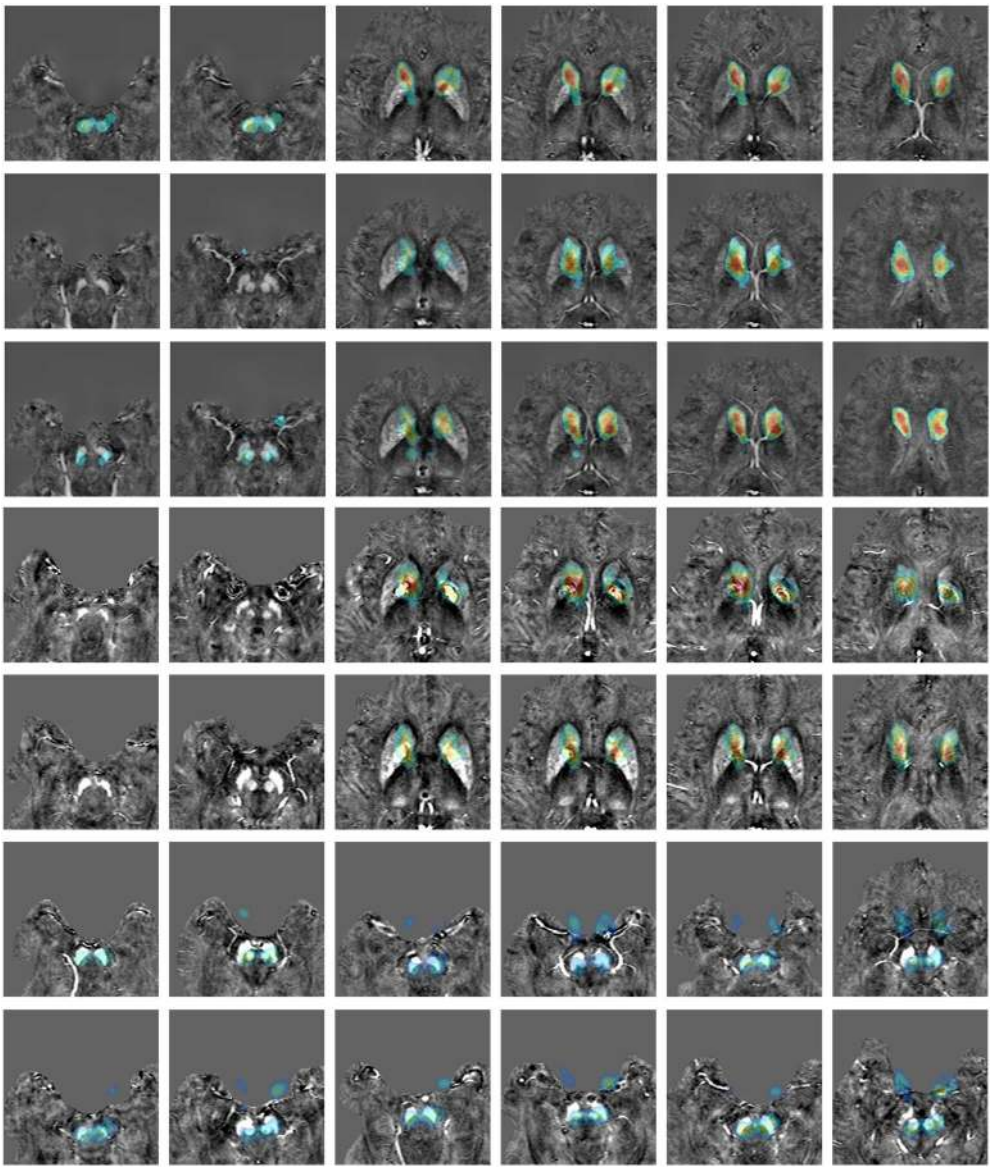

FIGURE 7 The gradient-weighted class activation mapping (Grad–CAM) visualization of AG–SE-ResNeXt50 on quantitative susceptibility mapping images of Parkinson's disease (PD) patients in the testing cohort. Top row: PD patient (Case 1 in Figure 3) predicted correctly with segmented regions of interest (ROIs). Second row: PD patient (Case 2 in Figure 3) predicted incorrectly as negative by the proposed model using segmented ROIs. Third row: Case 2 predicted correctly by the proposed model using the manually outlined ROI. Fourth and fifth rows: another two PD patients predicted incorrectly with segmented ROIs. Sixth and seventh rows: another six PD patients predicted correctly with segmented ROIs. AG: anatomical gate; SE, squeeze and excitation.

improved accuracy. The Grad-CAM heatmaps can be used to provide information about relevant nuclei for an individual patient to help radiologists confirm the model's prediction.

ACKNOWLEDGMENTS

We would like to thank all our participants for their contributions.

FUNDING INFORMATION

This work was supported, in part, by the National Natural Science Foundation of China (grant number: 81971576 and 82271954 for Fuhua Yan, MD and Naying He, MD, grant number: 61731009 for Guang Yang); by grants from National Key R&D Program of China (2022YFC2009905/2022YFC2009900) and the Innovative Research Team of High-level Local Universities in Shanghai. This work was supported by Xing-Fu-Zhi-Hua Foundation of ECNU, Microscale Magnetic Resonance Platform of ECNU, and the Open Project of Shanghai Key Laboratory of Magnetic Resonance.

CONFLICT OF INTEREST STATEMENT

The authors have no potential conflicts of interest to declare.

DATA AVAILABILITY STATEMENT

The source and means of obtaining the data used in this article has been described in the Section 2 of this article. The data are available via a reasonable request with a formal data sharing agreement to the author Fuhua Yan, Naying He, and Jingliang Cheng. Source code of the models used in this study is publicly available at: https://github.com/wangyidada/PD_diagnosis.git.

ORCID

Yida Wang https://orcid.org/0000-0001-7501-5037
Jingliang Cheng https://orcid.org/0000-0002-6996-329X
Ewart Mark Haacke https://orcid.org/0000-0003-4965-7954
Guang Yang https://orcid.org/0000-0001-8942-427X
Fuhua Yan https://orcid.org/0000-0002-5910-1506

REFERENCES

Abdul-Rahman, H. S., Gdeisat, M. A., Burton, D. R., Lalor, M. J., Lilley, F., & Moore, C. J. (2007). Fast and robust three-dimensional best path phase unwrapping algorithm. *Applied Optics*, *46*, 6623–6635. https://doi.org/10.1364/ao.46.006623

Adeli, E., Shi, F., An, L., Wee, C. Y., Wu, G., Wang, T., & Shen, D. (2016). Joint feature-sample selection and robust diagnosis of Parkinson's disease from MRI data. *NeuroImage*, *141*, 206–219. https://doi.org/10.1016/j.neuroimage.2016.05.054

Berg, D., Adler, C. H., Bloem, B. R., Chan, P., Gasser, T., Goetz, C. G., Halliday, G., Lang, A. E., Lewis, S., Li, Y., Liepelt-Scarfone, I., Litvan, I., Marek, K., Maetzler, C., Mi, T., Obeso, J., Oertel, W., Olanow, C. W., Poewe, W., … Postuma, R. B. (2018). Movement disorder society criteria for clinically established early Parkinson's disease. *Movement Disorders*, *33*, 1643–1646. https://doi.org/10.1002/mds.27431

Berg, D., & Youdim, M. B. (2006). Role of iron in neurodegenerative disorders. *Topics in Magnetic Resonance Imaging*, *17*, 5–17. https://doi.org/10.1097/01.rmr.0000245461.90406.ad

Chai, C., Wu, M., Wang, H., Cheng, Y., Zhang, S., Zhang, K., Shen, W., Liu, Z., & Xia, S. (2022). CAU-Net: A deep learning method for deep gray matter nuclei segmentation. *Frontiers in Neuroscience*, *16*, 918623. https://doi.org/10.3389/fnins.2022.918623

Chen, Y. S., Liu, S. F., Wang, Y., Kang, Y., & Haacke, E. M. (2018). STrategically acquired gradient Echo (STAGE) imaging, part I: Creating enhanced T1 contrast and standardized susceptibility weighted imaging and quantitative susceptibility mapping. *Magnetic Resonance Imaging*, *46*, 130–139. https://doi.org/10.1016/j.mri.2017.10.005

Cobzas, D., Sun, H., Walsh, A. J., Lebel, R. M., Blevins, G., & Wilman, A. H. (2015). Subcortical gray matter segmentation and voxel-based analysis using transverse relaxation and quantitative susceptibility mapping with application to multiple sclerosis. *Journal of Magnetic Resonance Imaging*, *42*, 1601–1610. https://doi.org/10.1002/jmri.24951

Du, W., Wang, Y., & Qiao, Y. (2018). Recurrent spatial-temporal attention network for action recognition in videos. *IEEE Transactions on Image Processing*, *27*, 1347–1360. https://doi.org/10.1109/TIP.2017.2778563

Eskreis-Winkler, S., Zhang, Y., Zhang, J. W., Liu, Z., Dimov, A., Gupta, A., & Wang, Y. (2017). The clinical utility of QSM: Disease diagnosis, medical management, and surgical planning. *NMR in Biomedicine*, *30*, e3668. https://doi.org/10.1002/nbm.3668

Garzon, B., Sitnikov, R., Backman, L., & Kalpouzos, G. (2018). Automated segmentation of midbrain structures with high iron content. *NeuroImage*, *170*, 199–209. https://doi.org/10.1016/j.neuroimage.2017.06.016

Gaurav, R., Valabregue, R., Yahia-Cherif, L., Mangone, G., Narayanan, S., Arnulf, I., Vidailhet, M., Corvol, J. C., & Lehericy, S. (2022). NigraNet: An automatic framework to assess nigral neuromelanin content in early Parkinson's disease using convolutional neural network. *NeuroImage: Clinical*, *36*, 103250. https://doi.org/10.1016/j.nicl.2022.103250

Goetz, C. G., Tilley, B. C., Shaftman, S. R., Stebbins, G. T., Fahn, S., Martinez-Martin, P., Poewe, W., Sampaio, C., Stern, M. B., Dodel, R., Dubois, B., Holloway, R., Jankovic, J., Kulisevsky, J., Lang, A. E., Lees, A., Leurgans, S., LeWitt, P. A., Nyenhuis, D., … LaPelle, N. (2008). Movement Disorder Society-sponsored revision of the unified Parkinson's Disease Rating Scale (MDS-UPDRS): Scale presentation and clinimetric testing results. *Movement Disorders*, *23*, 2129–2170. https://doi.org/10.1002/mds.22340

Gu, R., Wang, G. T., Song, T., Huang, R., Aertsen, M., Deprest, J., Ourselin, S., Vercauteren, T., & Zhang, S. T. (2021). CA-net: Comprehensive attention convolutional neural networks for explainable medical image segmentation. *IEEE Transactions on Medical Imaging*, *40*, 699–711. https://doi.org/10.1109/Tmi.2020.3035253

Guo, T., Song, Y., Li, J., Fan, M., Yan, X., He, A., Huang, D., Shen, C., Zhang, G., & Yang, G. (2018). Seed point discontinuity-based segmentation method for the substantia nigra and the red nucleus in quantitative susceptibility maps. *Journal of Magnetic Resonance Imaging*, *48*, 1112–1119. https://doi.org/10.1002/jmri.26023

Haacke, E. M., Chen, Y., Utriainen, D., Wu, B., Wang, Y., Xia, S., He, N., Zhang, C., Wang, X., Lagana, M. M., Luo, Y., Fatemi, A., Liu, S., Gharabaghi, S., Wu, D., Sethi, S. K., Huang, F., Sun, T., Qu, F., … Yan, F. (2020). STrategically Acquired Gradient Echo (STAGE) imaging, part III: Technical advances and clinical applications of a rapid multi-contrast multi-parametric brain imaging method. *Magnetic Resonance Imaging*, *65*, 15–26. https://doi.org/10.1016/j.mri.2019.09.006

Haacke, E. M., Liu, S. F., Buch, S., Zheng, W. L., Wu, D. M., & Ye, Y. Q. (2015). Quantitative susceptibility mapping: Current status and future directions. *Magnetic Resonance Imaging*, *33*, 1–25. https://doi.org/10.1016/j.mri.2014.09.004

Haacke, E. M., Tang, J., Neelavalli, J., & Cheng, Y. C. (2010). Susceptibility mapping as a means to visualize veins and quantify oxygen saturation. *Journal of Magnetic Resonance Imaging*, *32*, 663–676. https://doi.org/10.1002/jmri.22276

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, USA, pp. 770–778. https://doi.org/10.1109/CVPR.2016.90

He, N. Y., Ghassaban, K., Huang, P., Jokar, M., Wang, Y., Cheng, Z. H., Jin, Z. J., Li, Y., Sethi, S. K., He, Y. X., Chen, Y. S., Gharabaghi, S., Chen, S. D., Yan, F. H., & Haacke, E. M. (2021). Imaging iron and neuromelanin simultaneously using a single 3D gradient echo magnetization transfer sequence: Combining neuromelanin, iron and the nigrosome-1 sign as complementary imaging biomarkers in early stage Parkinson's disease. *NeuroImage*, *230*, 117810. https://doi.org/10.1016/j.neuroimage.2021.117810

Hu, J., Shen, L., & Sun, G. (2017). Squeeze-and-excitation networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 99. https://doi.org/10.1109/Cvpr.2018.00745

Jankovic, J. (2008). Parkinson's disease: Clinical features and diagnosis. *Journal of Neurology Neurosurgery and Psychiatry*, *79*, 368–376. https://doi.org/10.1136/jnnp.2007.131045

Jin, Z. J., Wang, Y., Jokar, M., Li, Y., Cheng, Z. H., Liu, Y., Tang, R. B., Shi, X. F., Zhang, Y. M., Min, J. H., Liu, F. T., He, N. Y., Yan, F. H., & Haacke, E. M. (2022). Automatic detection of neuromelanin and iron in the midbrain nuclei using a magnetic resonance imaging-based brain template. *Human Brain Mapping*, *43*, 2011–2025. https://doi.org/10.1002/hbm.25770

Kalia, L. V., & Lang, A. E. (2015). Parkinson's disease. *Lancet*, *386*, 896–912. https://doi.org/10.1016/S0140-6736(14)61393-3

Kaur, S., Aggarwal, H., & Rani, R. (2020). Diagnosis of Parkinson's disease using deep CNN with transfer learning and data augmentation. *Multimedia Tools and Applications*, *80*, 10113–10139. https://doi.org/10.1007/s11042-020-10114-1

Krupicka, R., Marecek, S., Mala, C., Lang, M., Klempir, O., Duspivova, T., Siroka, R., Jarosikova, T., Keller, J., Sonka, K., Ruzicka, E., & Dusek, P. (2019). Automatic substantia nigra segmentation in neuromelanin-sensitive MRI by deep neural network in patients with prodromal and manifest synucleinopathy. *Physiological Research*, *68*, S453–S458. https://doi.org/10.33549/physiolres.934380

Le Berre, A., Kamagata, K., Otsuka, Y., Andica, C., Hatano, T., Saccenti, L., Ogawa, T., Takeshige-Amano, H., Wada, A., Suzuki, M., Hagiwara, A., Irie, R., Hori, M., Oyama, G., Shimo, Y., Umemura, A., Hattori, N., & Aoki, S. (2019). Convolutional neural network-based segmentation can help in assessing the substantia nigra in neuromelanin MRI. *Neuroradiology*, *61*, 1387–1395. https://doi.org/10.1007/s00234-019-02279-w

Li, G., Zhai, G., Zhao, X., An, H., Spincemaille, P., Gillen, K. M., Ku, Y., Wang, Y., Huang, D., & Li, J. (2019). 3D texture analyses within the substantia nigra of Parkinson's disease patients on quantitative susceptibility maps and R2(*) maps. *NeuroImage*, *188*, 465–472. https://doi.org/10.1016/j.neuroimage.2018.12.041

Li, L., & Leigh, J. S. (2004). Quantifying arbitrary magnetic susceptibility distributions with MR. *Magnetic Resonance in Medicine*, *51*, 1077–1082. https://doi.org/10.1002/mrm.20054

Li, X., Chen, L., Kutten, K., Ceritoglu, C., Li, Y., Kang, N., Hsu, J. T., Qiao, Y., Wei, H., Liu, C., Miller, M. I., Mori, S., Yousem, D. M., van Zijl, P. C. M., & Faria, A. V. (2019). Multi-atlas tool for automated segmentation of brain gray matter nuclei and quantification of their magnetic susceptibility. *NeuroImage*, *191*, 337–349. https://doi.org/10.1016/j.neuroimage.2019.02.016

Lim, I. A. L., Faria, A. V., Li, X., Hsu, J. T. C., Airan, R. D., Mori, S., & van Ziji, P. C. M. (2013). Human brain atlas for automated region of interest selection in quantitative susceptibility mapping: Application to determine iron content in deep gray matter structures. *NeuroImage*, *82*, 449–469. https://doi.org/10.1016/j.neuroimage.2013.05.127

Lin, T. Y., Goyal, P., Girshick, R., He, K., & Dollár, P. (2017). Focal loss for dense object detection. IEEE International Conference on Computer Vision (ICCV), pp. 2999–3007. https://doi.org/10.1109/ICCV.2017.324

Lotfipour, A. K., Wharton, S., Schwarz, S. T., Gontu, V., Schafer, A., Peters, A. M., Bowtell, R. W., Auer, D. P., Gowland, P. A., & Bajaj, N. P. S. (2012). High resolution magnetic susceptibility mapping of the substantia nigra in Parkinson's disease. *Journal of Magnetic Resonance Imaging*, *35*, 48–55. https://doi.org/10.1002/jmri.22752

Niu, Z. Y., Zhong, G. Q., & Yu, H. (2021). A review on the attention mechanism of deep learning. *Neurocomputing*, *452*, 48–62. https://doi.org/10.1016/j.neucom.2021.03.091

Rana, B., Juneja, A., Saxena, M., Gudwani, S., Senthil Kumaran, S., Agrawal, R. K., & Behari, M. (2015). Regions-of-interest based automated diagnosis of Parkinson's disease using T1-weighted MRI. *Expert Systems with Applications*, *42*, 4506–4516. https://doi.org/10.1016/j.eswa.2015.01.062

Reza, A. M. (2004). Realization of the contrast limited adaptive histogram equalization (CLAHE) for real-time image enhancement. *Journal of VLSI Signal Processing Systems for Signal Image and Video Technology.*, *38*, 35–44. https://doi.org/10.1023/B:VLSI.0000028532.53893.82

Riccardo, G., Anna, M., Salvatore, R., Franco, T., Fosca, G., & Dino, P. (2018). A survey of methods for explaining black box models. *ACM Computing Surveys*, *51*, 1–42. https://doi.org/10.1145/3236009

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation. *Medical Image Computing and Computer-Assisted Intervention, Pt III*, *9351*, 234–241. https://doi.org/10.1007/978-3-319-24574-4_28

Schweser, F., Deistung, A., Lehr, B. W., & Reichenbach, J. R. (2011). Quantitative imaging of intrinsic magnetic tissue properties using MRI signal phase: An approach to in vivo brain iron metabolism? *NeuroImage*, *54*, 2789–2807. https://doi.org/10.1016/j.neuroimage.2010.10.070

Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2020). Grad-CAM: Visual explanations from deep networks via gradient-based localization. *International Journal of Computer Vision*, *128*, 336–359. https://doi.org/10.1007/s11263-019-01228-7

Shinde, S., Prasad, S., Saboo, Y., Kaushick, R., Saini, J., Pal, P. K., & Ingalhalikar, M. (2019). Predictive markers for Parkinson's disease using deep neural nets on neuromelanin sensitive MRI. *NeuroImage: Clinical*, *22*, 101748. https://doi.org/10.1016/j.nicl.2019.101748

Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. Computer Science. https://doi.org/10.48550/arXiv.1409.1556

Sivaranjini, S., & Sujatha, C. M. (2019). Deep learning based diagnosis of Parkinson's disease using convolutional neural network. *Multimedia Tools and Applications*, *79*, 15467–15479. https://doi.org/10.1007/s11042-019-7469-8

Smith, S. M. (2002). Fast robust automated brain extraction. *Human Brain Mapping*, *17*, 143–155. https://doi.org/10.1002/hbm.10062

Suganyadevi, S., Seethalakshmi, V., & Balasamy, K. (2022). A review on deep learning in medical image analysis. *International Journal of Multimedia Information Retrieval*, *11*, 19–38. https://doi.org/10.1007/s13735-021-00218-1

Sun, L., Shao, W., Zhang, D., & Liu, M. (2020). Anatomical attention guided deep networks for ROI segmentation of brain MR images. *IEEE Transactions on Medical Imaging*, *39*, 2000–2012. https://doi.org/10.1109/TMI.2019.2962792

Tang, J., Liu, S., Neelavalli, J., Cheng, Y. C., Buch, S., & Haacke, E. M. (2013). Improving susceptibility mapping using a threshold-based K-space/image domain iterative reconstruction approach. *Magnetic Resonance in Medicine*, *69*, 1396–1407. https://doi.org/10.1002/mrm.24384

Tysnes, O. B., & Storstein, A. (2017). Epidemiology of Parkinson's disease. *Journal of Neural Transmission*, *124*, 901–905. https://doi.org/10.1007/s00702-017-1686-y

Vandenhende, S., Georgoulis, S., Gansbeke, W. V., Proesmans, M., & van Gool, L. (2021). Multi-task learning for dense prediction tasks: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *44*, 3614–3633. https://doi.org/10.1109/TPAMI.2021.3054719

Xiao, B., He, N., Wang, Q., Cheng, Z., Jiao, Y., Haacke, E. M., Yan, F., & Shi, F. (2019). Quantitative susceptibility mapping based hybrid feature extraction for diagnosis of Parkinson's disease. *NeuroImage: Clinical*, *24*, 102070. https://doi.org/10.1016/j.nicl.2019.102070

Xie, S. N., Girshick, R., Dollar, P., Tu, Z. W., & He, K. M. (2017). Aggregated residual transformations for deep neural networks. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 5987–5995. https://doi.org/10.1109/Cvpr.2017.634

Zhang, Y., Yang, M., Wang, F., Chen, Y., Liu, R., Zhang, Z., & Jiang, Z. (2022). Histogram analysis of quantitative susceptibility mapping for the diagnosis of Parkinson's disease. *Academic Radiology*, *29*, 71–79. https://doi.org/10.1016/j.acra.2020.10.027

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Wang, Y., He, N., Zhang, C., Zhang, Y., Wang, C., Huang, P., Jin, Z., Li, Y., Cheng, Z., Liu, Y., Wang, X., Chen, C., Cheng, J., Liu, F., Haacke, E. M., Chen, S., Yang, G., & Yan, F. (2023). An automatic interpretable deep learning pipeline for accurate Parkinson's disease diagnosis using quantitative susceptibility mapping and T1-weighted images. *Human Brain Mapping*, *44*(12), 4426–4438. https://doi.org/10.1002/hbm.26399