



# Automatic substantia nigra segmentation with Swin-Unet in susceptibility- and T2-weighted imaging: application to Parkinson disease diagnosis

Tongxing Wang<sup>1#</sup>, Yajing Wang<sup>1#</sup>, Haichen Zhu<sup>2</sup>, Zhen Liu<sup>3</sup>, Yu-Chen Chen<sup>1</sup>, Liwei Wang<sup>1</sup>, Shaofeng Duan<sup>4</sup>, Xindao Yin<sup>1</sup>, Liang Jiang<sup>1</sup>

<sup>1</sup>Department of Radiology, Nanjing First Hospital, Nanjing Medical University, Nanjing, China; <sup>2</sup>Lab of Image Science and Technology, Key Laboratory of Computer Network and Information Integration (Ministry of Education), School of Computer Science and Engineering, Southeast University, Nanjing, China; <sup>3</sup>Department of Radiology, The Affiliated ChuZhou Hospital of AnHui Medical University, Chuzhou, China; <sup>4</sup>GE HealthCare, Precision Health Institution, Shanghai, China

*Contributions:* (I) Conception and design: T Wang, Y Wang; (II) Administrative support: L Jiang, X Yin; (III) Provision of study materials or patients: H Zhu, Z Liu, YC Chen, L Wang, L Jiang; (IV) Collection and assembly of data: T Wang, Y Wang; (V) Data analysis and interpretation: H Zhu, Z Liu, YC Chen, L Wang, S Duan; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

<sup>#</sup>These authors have contributed equally to this work.

*Correspondence to:* Xindao Yin, MD, PhD; Liang Jiang, MD. Department of Radiology, Nanjing First Hospital, Nanjing Medical University, Changle Road No. 68, Nanjing 210006, China. Email: y.163yy@163.com; jiangliang0402@163.com.

**Background:** Accurately distinguishing between Parkinson disease (PD) and healthy controls (HCs) through reliable imaging method is crucial for appropriate therapeutic intervention. However, PD diagnosis is hindered by the subjective nature of the evaluation. We aimed to develop an automatic deep-learning method that can segment the substantia nigra areas on susceptibility-weighted imaging (SWI) and T2-weighted imaging (T2WI) and further differentiate patients with PD from HCs using a machine learning algorithm.

**Methods:** Magnetic resonance imaging (MRI) data from 83 patients with PD and 83 age- and sex-matched HCs were obtained on the same 3.0-T MRI scanner. A deep learning method with Swin-Unet was developed to segment volumes of interest (VOIs) on SWI and then map the VOIs on SWI to the corresponding T2WI; features were then extracted from the VOIs on SWI and T2WI. Three machine learning models were developed and compared to differentiate those with PD from HCs.

**Results:** Swin-Unet achieved a better Dice coefficient than did U-Net in SWI segmentation (0.832 *vs.* 0.712). Machine learning models outperformed visual analysis ( $P > 0.05$ ), and logistic regression (LR) achieved the best performance [area under the curve (AUC)  $\geq 0.819$ ] and the most stable (relative standard deviations in AUC  $\leq 0.05$ ). The test results showed that the AUC of the LR model based on SWI segmentation was 0.894 while that of the LR model based on T2WI segmentation was 0.876. There was no significant difference in VOIs based on manual labeling or automatic segmentation across T2WI, SWI, or a combination of the two ( $P > 0.05$ ). The AUCs of the LR model based on automatic segmentation were close to those of the model based on manual labeling ( $P > 0.05$ ).

**Conclusions:** Our approach could provide a powerful and useful method for automatically and rapidly diagnosing PD in the clinic with only T2WI.

**Keywords:** Parkinson's disease; magnetic resonance imaging (MRI); deep learning; machine learning; substantia nigra (SN)

Submitted Jan 06, 2024. Accepted for publication Jul 15, 2024. Published online Aug 21, 2024.

doi: 10.21037/qims-24-27

View this article at: <https://dx.doi.org/10.21037/qims-24-27>

## Introduction

Parkinson disease (PD), the second most common neurodegenerative disease, is characterized by neuronal loss in the substantia nigra (SN) (1). The prevalence of PD increases with age, leading to reduced quality of life and increased mortality (2). Therefore, the early and accurate diagnosis of PD is essential. However, the clinical presentations of PD are heterogeneous, making diagnosis challenging. Neuroimaging, particularly susceptibility-weighted imaging (SWI), has facilitated the estimation of SN neuronal loss in PD and thus is considered valuable for the diagnosis of this disease (3-5). The “swallow tail” sign, the appearance of a healthy nigrosome-1 (N1) on axial T2-weighted imaging (T2WI) and SWI, has been demonstrated to be a useful radiological sign for differentiating patients with PD from healthy controls (HCs) (6,7). SWI has been shown to be more sensitive than conventional magnetic resonance imaging (MRI) sequences in detecting loss of the swallow tail sign (8). However, the occurrence of the swallow tail sign in HCs is inconsistent, and the disappearance of the swallow tail sign can also be found in some cognitive disorders (9). The limitations of low specificity and low accuracy, especially in routine MRI, have hindered the clinical application of the swallow tail sign.

Radiomics is a method for extracting high-throughput data to provide a detailed characterization of radiographic images. It can capture lesion characteristics such as heterogeneity and shape and may—alone or in combination with demographic, histologic, genomic, or proteomic data—be used for clinical decision-making (10-12). Radiomic features can be roughly subdivided into statistical including histogram-based, texture-based, model-based, transform-based, and shape-based. Radiomics is also leveraged in the diagnosis of PD (13-15). A previous radiomics analysis indicated that radiomics findings based on dopamine transporter single-photon emission computed tomography (SPECT) can serve as a biomarker to track the progression of PD (16). However, its complicated sequences and time-consuming procedures limit its clinical application. A previous study also showed that some SN radiomics features based on SWI signal intensity could distinguish patients with PD from HCs (14). Although radiomics has good diagnostic accuracy for PD, the manual outlining of the SN remains a

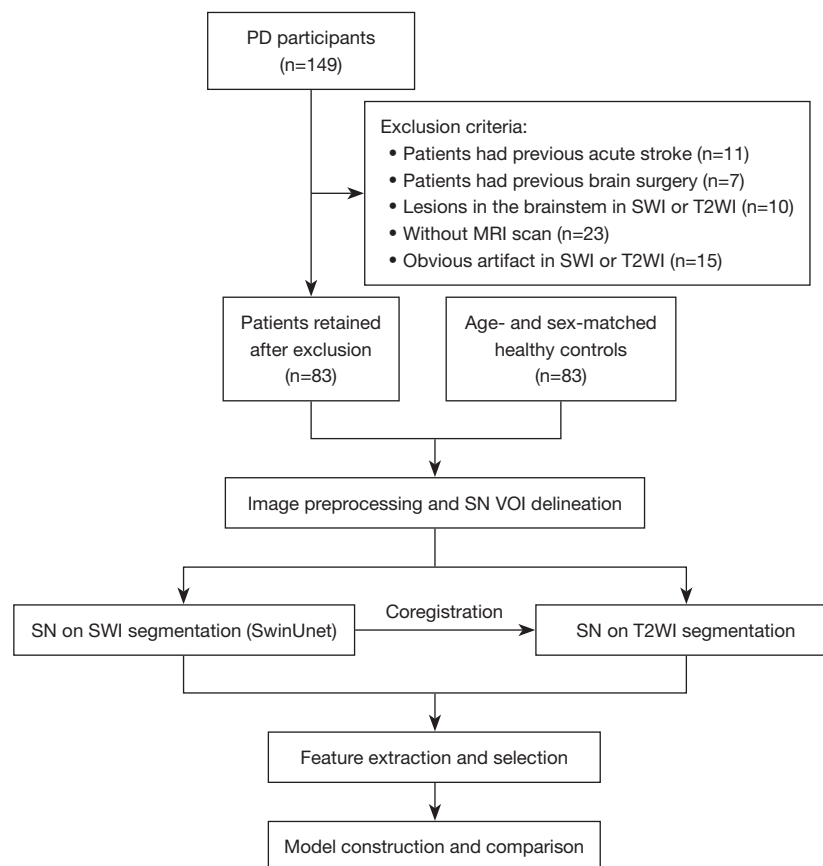
labor-intensive process. Furthermore, the results of manual segmentation can vary widely due to interrater differences. Deep segmentation networks have shown fast speed and high accuracy in automatic image segmentation (17,18). Automated detection, segmentation, and classification can free up clinical doctors' time for higher value tasks and reduce errors due to fatigue and subjectivity (19). The encoder-decoder-based network U-Net is typically employed as a baseline model in different medical image segmentation benchmarks by virtue of its simple structure and advantages in segmenting subtle tissues (20). Notably, a recent study reported that neostriatum radiomics signatures based on T2WI achieved good diagnostic performance for PD and could potentially serve as a basis for the clinical diagnosis of PD (21). However, whether SN radiomics features based on T2WI can help to distinguish PD from HCs has not yet been established. In addition, the efficacy of differentiation models developed using machine learning based on SN radiomics remains uncertain.

In this study, we used Swin-Unet (22) to segment the SN on SWI and T2WI, thus reducing or eliminating the time needed for radiologists to label volumes of interest (VOIs). In Swin-Unet, some conventional U-Net encoders are replaced with a transformer block to extract global image information and achieve a better segmentation performance (23). Subsequently, we used the radiomics features extracted from the VOI segmentation to diagnose PD and compared these with those of manual labeling methods. To this end, we constructed three classifiers [i.e., support vector machine (SVM), logistic regression (LR), and random forest (RF)], developed a stable machine learning model to distinguish patients with PD from HCs, and then estimated the generalizability of the model in a test group. We present this article in accordance with the TRIPOD reporting checklist (available at <https://qims.amegroups.com/article/view/10.21037/qims-24-27/rc>).

## Methods

### Study sample

This retrospective study was approved by the Ethics Committee of Nanjing Medical University (No. 2019-664) and was conducted in accordance with the Declaration of Helsinki (as revised in 2013). The requirement for written



**Figure 1** Flowchart of the included patients. PD, Parkinson disease; SWI, susceptibility-weighted imaging; T2WI, T2-weighted imaging; MRI, magnetic resonance imaging; SN, substantia nigra; VOI, volume of interest.

informed consent was waived by the ethics committee due to the nature of the retrospective design. We recruited patients with PD from Nanjing First Hospital between January 2017 and January 2021 who underwent 3-T brain imaging, including an SWI sequence. All patients met the UK Parkinson Disease Society Brain Bank clinical diagnostic criteria for PD (24) and were further evaluated with the Hoehn and Yahr (H&Y) scale (25). Patients with PD that had a history of other neurological and psychiatric diseases and secondary parkinsonism due to head trauma or medication use or that had features of atypical parkinsonism syndromes that could have interfered with the results were excluded (26). Age- and sex-matched healthy participants were used as the control group. All participants underwent a thorough interview concerning their medical history and a clinical examination following the International Parkinson and Movement Disorder Society (MDS) Unified Parkinson Disease Rating Scale protocol. HCs were required to have no history of any neurological or psychiatric disorders. All

participants underwent 3-T brain imaging, including an SWI sequence. SWI and T2WI images with visible motion artifacts on the SN area that could affect feature extraction were excluded. Ultimately, a total of 83 patients with PD and 83 HCs were reviewed. A flowchart of this study is shown in *Figure 1*. The patients with PD and HCs were randomly allocated to training ( $n=116$ ) and test groups ( $n=50$ ) at a ratio of 7:3.

### *MRI acquisition and analysis*

All participants underwent MRI on a 3.0-T MRI scanner (MAGNETOM Prisma, Siemens Healthineers, Erlangen, Germany). The MRI protocol included T2WI [repetition time (TR), 3,000 ms; echo time (TE), 103 ms; field of view (FOV), 220 mm × 220 mm; flip angle (FA), 150°; slice thickness, 6 mm] and SWI (TR, 27 ms; TE, 20 ms; FOV, 220 mm × 220 mm; FA, 15°; slice thickness, 2 mm).

The swallow-tail sign normally appears as a high-signal

structure that has a linear, comma-like or wedge-like shape and is bordered on both sides by low-signal structures (the SN pars compacta and the medial lemniscus). Two radiologists (Z.L., 8 years of experience; Y.C.C., 10 years of experience), blinded to each participant's status as a patient or control, evaluated the presence or absence of the swallow tail sign on axial sections from the SWI or T2WI sequence. Bilateral absence, unilateral absence, and faint presence of the swallow tail sign were considered negative, while bilateral presence was considered positive. In the event of a dispute, a resolution was reached through a collaborative decision-making process involving the two medical professionals mentioned above.

### Image preprocessing and VOI delineation

SWI and T2WI were performed for all participants in the same scan, so the world coordinates of the corresponding pixel dimension in the two sequences were consistent. The pixel dimensions of the SW images varied from  $0.429 \times 0.429 \times 0.5$  to  $0.859 \times 0.859 \times 2.5$  mm<sup>3</sup>, and those of the T2W images varied from  $0.286 \times 0.286 \times 7.2$  to  $0.687 \times 0.687 \times 6.5$  mm<sup>3</sup>. Considering the voxel size variance, for each participant, we reshape the SW images voxel size to  $1 \times 1 \times \beta$  mm<sup>3</sup> (where  $\beta$  is the original SWI z-dimension voxel size) and then cropped each SWI volume to the size of  $256 \times 256 \times N$  (where  $N$  represents the number of z-dimensions). The SN areas on SWI (SWI<sub>VOI</sub>) were then delineated manually on transverse slices using ITK-SNAP (www.itksnap.org) by one radiologist with 10 years of experience (L.W.) and then checked by another radiologist with 15 years of experience (X.Y.). If the initial VOIs were determined to be inaccurate, they were subsequently revised and redrawn for further analysis. The VOIs were then transformed from SWI into T2WI space using the transform matrix, which was calculated by matching the voxel coordinates in the SW image and T2W image (T2WI<sub>VOI</sub>). To confirm that the anatomical location of T2WI<sub>VOI</sub> was correct, the same two board-certified neuroradiologists checked the location visually using ITK-SNAP software by superimposing the transformed T2WI<sub>VOI</sub> on the T2W image.

### SN segmentation

For swallow-tail sign analysis, SN VOIs should be delineated first. However, accurate labeling requires considerable manpower, and on T2WI, the SN is relatively

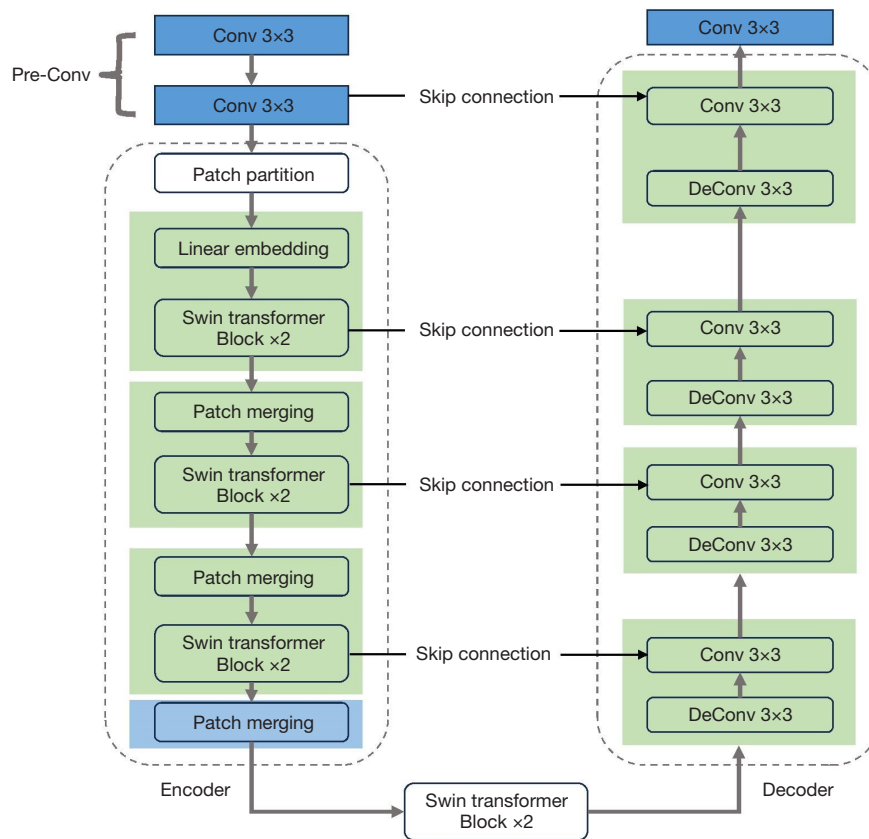
unclear, resulting in low diagnostic accuracy for patients with PD. U-Net has shown great potential in medical image segmentation; thus, we used Swin-UNet (22), a network based on U-Net with a Swin transformer structure, to segment VOIs from SWI scans. The structure of Swin-UNet is shown in Figure 2, with the input SWI size being  $256 \times 256$ . Given the advantages of convolution operations for the underlying visual feature extraction, the input images first underwent SWI feature extraction through two consecutive  $3 \times 3$  convolutional layers, and the output feature map was then sent into the Swin transformer blocks. In the Swin transformer block, the feature dependencies in the feature map were extracted using the self-attention mechanism with the following equation:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{C_k}}\right)V \quad [1]$$

where  $C_k$  is the number of feature channels used to normalize the data,  $Q$  is the query matrix,  $K$  is the key matrix, and  $V$  is the value matrix. To obtain features at different levels, patch partition or patch merging operations were introduced between the Swin transformer blocks, and then the feature map size was at last reduced four times to  $16 \times 16$ . At each map size level, we set the self-attention vector dimension to 96, 192, 384, and 768 successively. In the decoder part of the Swin-UNet, the reduced-resolution feature map was returned to the original size of the map layer by layer via a  $3 \times 3$  deconvolution layer, and the encoder features were incorporated into the decoder features through a skip connection for better segmentation results. Finally, the segmentation results were output from another  $3 \times 3$  deconvolution layer. Considering that VOIs only account for a very small portion of the overall image and most of the rest are background voxels in the SWI image, we trained Swin-UNet using weighted cross-entropy loss, as shown in the following equation:

$$\text{Loss} = -\sum_{i=1}^n w_c [y_{ci} \cdot \log x_{ci} + (1 - y_{ci}) \cdot \log(1 - x_{ci})] \quad [2]$$

where  $x_{ci}$  is the  $i$ -th pixel prediction probability of the  $c$ -th class,  $y_{ci}$  is the ground-truth label value of the  $i$ -th pixel, and  $w_c$  is the weight of the  $c$ -th class. The weight of VOI pixel category  $w_0$  is 1, and that of the background pixel category  $w_1$  is 0.1. The input data were first normalized to a mean of 0 and a variance of 1. We augmented the training dataset five times to prevent overfitting during the training period. The augmentation methods included flipping image,  $90^\circ$  and  $270^\circ$  rotation, and scaling at two random scales



**Figure 2** Schematic diagram of SWI segmentation. Conv, convolution; SWI, susceptibility-weighted imaging.

between 0.8 and 1.2. Thus, we were able to use almost 700-case SW images for training. The network was trained for 200 epochs to achieve loss convergence, and the training batch size was set to 28. The initial learning rate was set to  $1e-4$  and then decayed by multiplying by 0.98 in each epoch. The Swin-Unet network framework was developed by PyTorch on our personal computer (Intel Core i9 CPU, RTX3090 24 GB GPU, 32 GB RAM). Because the SW and T2W images were captured at the same time, the corresponding pixels in SWI and T2WI had the identical world coordinates. Based on this, we could map the VOIs in the SW images to the T2W images with the same world coordinates. The detailed steps are as follows: (I) the VOI voxel coordinates in the SW images were transformed using the affine transformation matrix  $M_s$  of the SW images to obtain the world coordinates  $y_{RAS}$  as follows:

$$y_{RAS} = M_s \cdot x_i \tag{3}$$

where  $x_i$  represents the voxel coordinates of the VOIs in the SW images. (II) The world coordinates of the VOIs in the

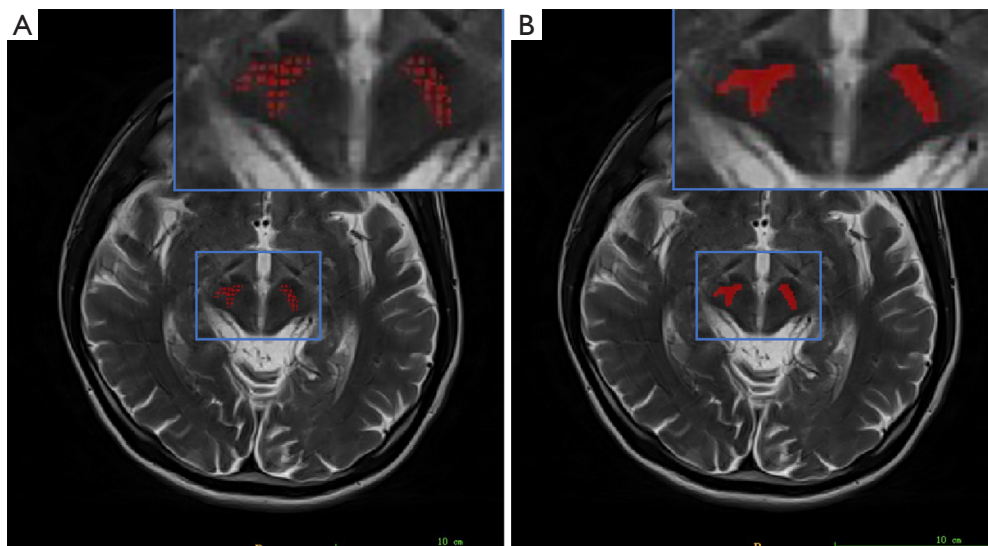
SW images were converted through the inverse operation of the affine transformation matrix  $M_t$  of the T2W images to the voxel coordinates  $z_i$  in the T2W images, where  $z_i = M_t^{-1} y_{RAS}$ , and were then marked in the T2W images. (III) Since the sizes of the T2W images were larger than those of the SW images, the VOI of the T2W images obtained by directly using the coordinate markers were discontinuous (Figure 3A); therefore, we used the close operation of the morphological function to eliminate VOI discontinuities, as shown in the following equation:

$$I \cdot S = (I \oplus S) \ominus S \tag{4}$$

where  $I$  is the VOI on the T2W images, and  $S$  is an operation core with a size of  $5 \times 5$ . The result of the operation is shown in Figure 3B.

**Feature extraction and selection**

The radiomics features of  $SWI_{VOI}$  (manual labeling on SWI),  $SWI_{seg}$  (segmentation on SWI),  $T2WI_{VOI}$  (coregistration



**Figure 3** Schematic diagram of T2WI segmentation. (A) The T2WI VOI obtained by directly using coordinate markers was discontinuous because the T2W images were larger than the SW images. (B) The T2WI VOI was continuous after application the close operation of morphological function. T2WI, T2-weighted imaging; VOI, volume of interest; T2W, T2-weighted; SW, susceptibility-weighted.

$SWI_{VOI}$  on T2WI), and  $T2WI_{seg}$  (coregistration  $SWI_{seg}$  on T2WI) were computed using PyRadiomics version 3.0.1 software (<https://pyradiomics.readthedocs.io/en/latest/>), which follows the principles of the Image Biomarker Standardization Initiative (IBSI). The radiomics features included six categories: shape-based [three-dimensional (3D)], first-order statistical, gray-level cooccurrence matrix (GLCM), gray-level run-length matrix (GLRLM), gray-level size-zone matrix (GLSZM), and gray-level dependence matrix (GLDM). Finally, a total of 1,132 features were extracted from each VOI. The mean and standard deviation of the features were normalized using the Z score method. To filter redundant features and reduce feature dimensions, the *t*-test was used to filter features based on their associated P values. Subsequently, the least absolute shrinkage and selection operator (LASSO) method, which is suitable for high-dimensional data regression, was used to select the most useful predictive features.

### Model construction

Three machine learning models were built for PD diagnosis: SVM (kernel: linear; cost: 0.1; number of support vectors: 53), LR (residual deviance: 86; null deviance: 128.6; residual deviance: 64.34; Akaike information criterion: 82.34), and RF (number of trees: 500; number of variables tried at each split: 2; out-of-bag estimate of error rate:

16.84%). Nested cross-validation was carried out to train the models with different machine learning methods. Leave one group out cross-validation was used for the outer loop, and 10-fold cross-validation was used for the inner loop. Each model was consequently constructed 100 times, and the corresponding 100 area under the curve (AUC) values and other metrics were calculated. The relative standard deviations (RSDs) were calculated using the following equation:

$$RSD = \frac{\sigma AUC}{\mu AUC} \times 100\% \quad [5]$$

where  $\sigma AUC$  is the standard deviation of the 100 AUC values, and  $\mu AUC$  is the mean of the 100 AUC values. The smaller the RSD%, the more stable the model is. After the aforementioned process, a final model with comparable performance and stability was chosen and validated with the test cohort.

### Statistical analysis

All statistical analyses were performed using R statistical version 4.0.3 software (The R Foundation for Statistical Computing). The Kolmogorov-Smirnov statistical test (alpha =0.05) was used to test the normality of continuous variables, which are presented as medians with interquartile ranges and were assessed with the Student's *t*-test and

**Table 1** The demographic characteristic of the PD and HC groups

Characteristic	Training group			Test group			P value (HC-training vs. HC-test)	P value (PD-training vs. PD-test)
	HC (n=58)	PD (n=58)	P value	HC (n=25)	PD (n=25)	P value		
Age (years), median (IQR)	63 (45, 75)	65 (47, 75)	0.143 <sup>a</sup>	63 (43, 76)	64 (49, 72)	0.357 <sup>a</sup>	0.382 <sup>a</sup>	0.523 <sup>a</sup>
Gender, male/female	25/33	28/30	0.576 <sup>b</sup>	10/15	12/13	0.569 <sup>b</sup>	0.793 <sup>b</sup>	0.982 <sup>b</sup>
Age of onset (years), median (IQR)	NA	57 (38, 71)	NA	NA	54 (39, 70)	NA	NA	0.618 <sup>a</sup>

<sup>a</sup>, two-sample Student's *t*-test; <sup>b</sup>, Chi-square test. PD, Parkinson disease; HC, healthy control; HC-training, healthy controls in the training group; HC-test, healthy controls in the test group; PD-training, Parkinson disease patients in the training group; PD-test, Parkinson disease patients in the test group; IQR, interquartile range; NA, not applicable.

**Table 2** Evaluation results of SWI segmentation using different networks

Network	Sensitivity	Specificity	Precision	Dice
U-Net	0.790	0.831	0.742	0.712
Swin-Unet	0.869*	0.999*	0.838*	0.832*

\*, the best results. SWI, susceptibility-weighted imaging.

Mann-Whitney test. Categorical variables are presented as percentages and were assessed with the  $\chi^2$  test. Analysis of variance and the Tukey multiple-comparison test were used for multiple comparisons. The interobserver reproducibility of imaging analysis was assessed using the intraclass correlation coefficient (ICC), with an ICC >0.8 indicating good agreement. The intrarater consistency between the two VOI segmentations was calculated using the Dice coefficient, with a Dice coefficient >0.8 indicating the high reproducibility of segmentations. Receiver operating characteristic (ROC) curve analysis was performed with the “pROC” package in R, and the AUC, sensitivity, specificity, and accuracy were used to compare the efficacy of the models. The ROC curves of the machine learning models were compared using the DeLong test.

## Results

### Demographics and clinical characteristics

Of the 58 patients with PD in the training group, 21 were H&Y stage 1, 21 were H&Y stage 2, and 12 were H&Y stage 3. Of the 25 patients with PD in the test group, 10 were H&Y stage 1, 10 were H&Y stage 2, and 5 were H&Y stage 3. Of the 83 patients with PD, 40 (48.19%) were male, while 35 of the 83 (42.17%) HCs were male. The average age of the PD group was 65.12±15.35 years, whereas that of the

HC group was 64.38±14.17 years. There was no significant difference in age or sex between the PD group and the HC group ( $P>0.05$ ). The demographic and clinical characteristics of the participants are shown in *Table 1*. The interobserver ICC between the two researchers (Z.L. and Y.C.C.) in the imaging analysis was 0.85, and the Dice coefficient between the two researchers (L.J. and X.Y.) for VOI segmentations was 0.97. In the visual analysis, SWI<sub>SN</sub> was able to the PD with an accuracy of 0.750 and an ROC-AUC of 0.750, while T2WI<sub>SN</sub> diagnosed PD with an accuracy of 0.612 and an ROC-AUC of 0.612. The AUC of the SWI<sub>SN</sub> was significantly higher than that of the T2WI<sub>SN</sub> ( $P<0.01$ ).

### SWI segmentation

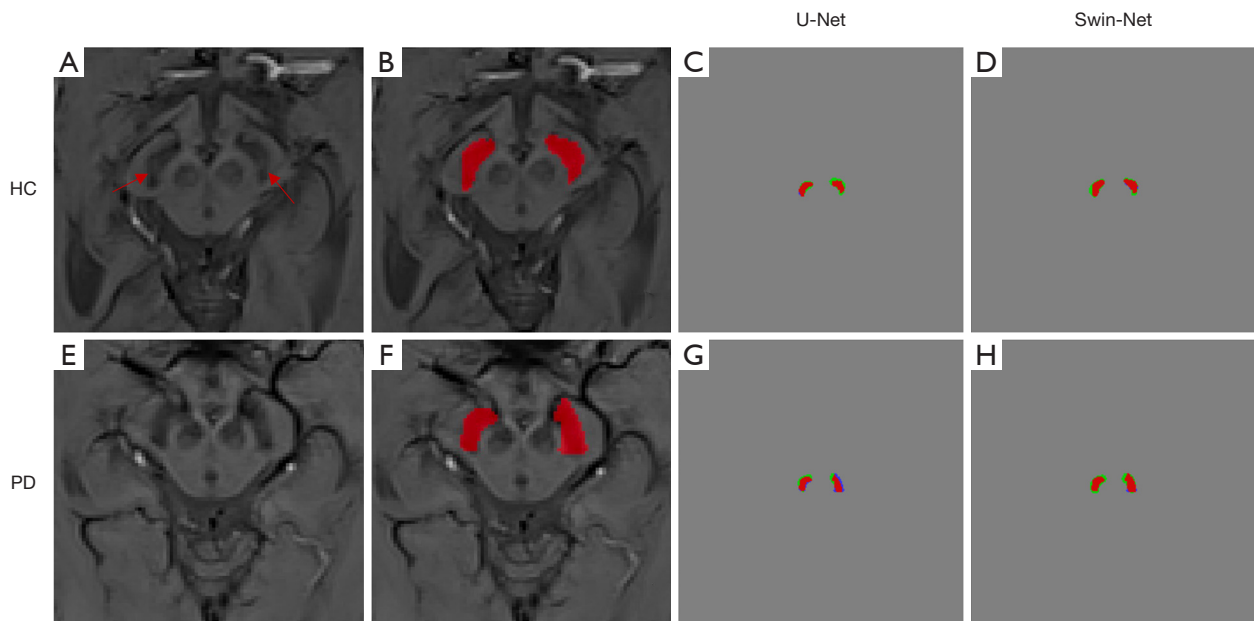
In segmenting the SWI scans, the Swin-Unet method, as compared with its basis, U-Net, achieved better sensitivity (0.869 vs. 0.790), specificity (0.999 vs. 0.831), precision (0.838 vs. 0.742), and Dice coefficient (0.832 vs. 0.712) (*Table 2*). *Figure 4* presents several SWI segmentation results obtained using Swin-Unet and U-Net.

### Feature selection

After LASSO screening, 6 features from SWI<sub>VOI</sub>, 11 features from SWI<sub>seg</sub>, 7 features from T2WI<sub>VOI</sub>, 12 features from T2WI<sub>seg</sub>, 7 features from SWI<sub>VOI</sub> + T2WI<sub>VOI</sub>, and 5 features from SWI<sub>seg</sub> + T2WI<sub>seg</sub> were selected. The tuning parameters and LASSO coefficients associated with PD are shown in *Figure S1*. The detailed features and weight coefficients are shown in *Figure 5*.

### Machine learning model

The performance metrics of the three models are shown



**Figure 4** Visualization of SWI VOI segmentation by U-Net and Swin-Net. (A-D) A 76-year-old male from the healthy control group. (A) Normal swallow-tail sign on both sides (red arrows). (B) The substantia nigra areas on SWI were delineated manually. Effectiveness of (C) U-Net and (D) Swin-Net in SWI VOI segmentation. (E-H) A 74-year-old male with Parkinson disease. Bilateral absence of the swallow-tail sign (E). The substantia nigra areas on SWI were delineated manually (F). Effectiveness of (G) U-Net and (H) Swin-Net in SWI VOI segmentation. True-positive pixels are denoted in red, false-positive pixels in green and false negative pixels in blue. HC, healthy control; PD, Parkinson disease; SWI, susceptibility-weighted imaging; VOI, volume of interest.

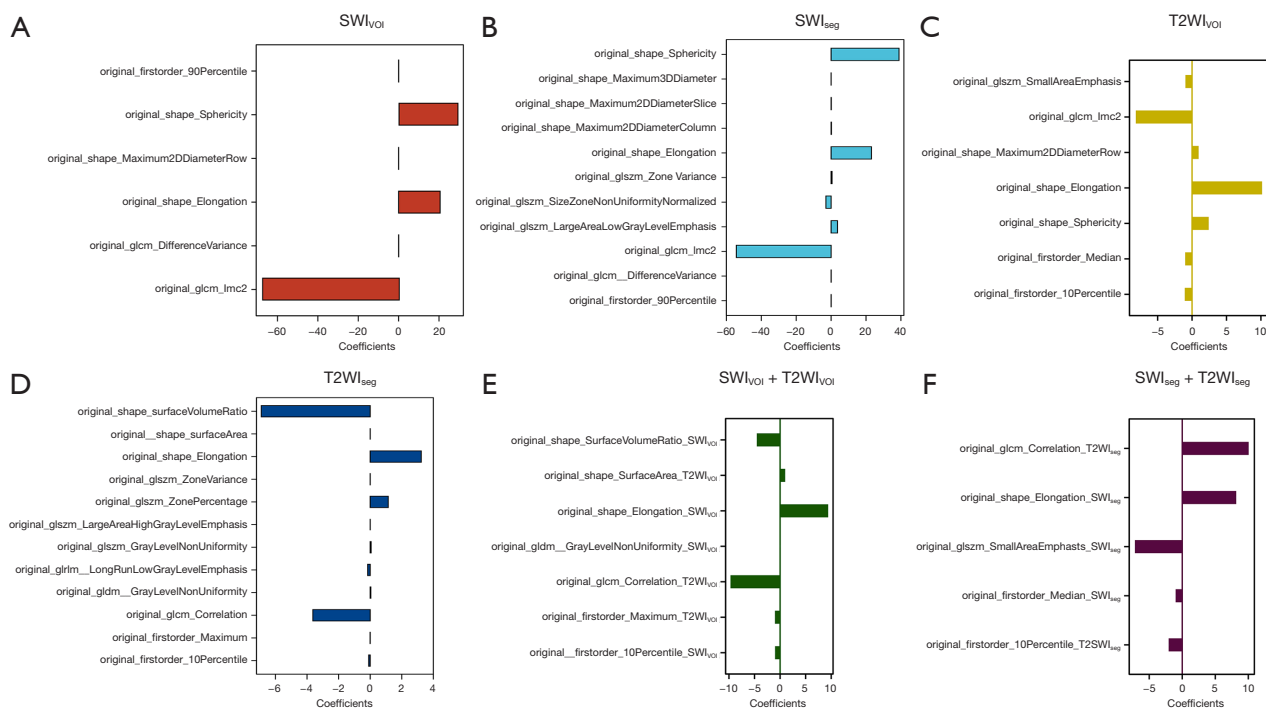
in *Table 3*. In the training group, the LR models based on  $SWI_{VOI}$  (AUC: 0.974; accuracy: 0.955),  $SWI_{seg}$  (AUC: 0.944; accuracy: 0.924),  $T2WI_{VOI}$  (AUC: 0.819; accuracy: 0.864),  $T2WI_{seg}$  (AUC: 0.852; accuracy: 0.864),  $SWI_{VOI} + T2WI_{VOI}$  (AUC: 0.927; accuracy: 0.935), and  $SWI_{seg} + T2WI_{seg}$  (AUC: 0.917; accuracy: 0.921) had the best diagnostic performance for patients with PD. The diagnostic performance of all machine learning models was significantly higher than that of visual analysis ( $P < 0.05$ ). In addition, assessment of model stability with different machine learning methods showed that the LR models based on  $SWI_{VOI}$  (RSD% of AUC: 0.04),  $SWI_{seg}$  (RSD% of AUC: 0.05),  $T2WI_{seg}$  (RSD% of AUC: 0.07),  $SWI_{VOI} + T2WI_{VOI}$  (RSD% of AUC: 0.05), and  $SWI_{seg} + T2WI_{seg}$  (RSD% of AUC: 0.04) were the most stable (*Table 3*). After the performance and stability were compared, LR was chosen as the final model for validation. There were no significant differences among T2WI, SWI, and or their combination in either the manually labeled VOIs or the segmentation VOI ( $P > 0.05$ ). The AUCs of the LR model based on segmentation VOI were close to those of the model based on manual labeling ( $P > 0.05$ ). The ROC curves of the models in the test group are shown in *Figure 6*.

The performance metrics of the model with the test group are shown in *Table 4*.

## Discussion

We developed and evaluated Swin-Net, a network that uses deep learning to segment SN areas on SWI scans from patients with PD and HCs and then maps the  $SWI_{VOIs}$  onto the corresponding T2WI to obtain T2WI SN; furthermore, we established an automatic PD diagnosis model using machine learning. The proposed segmentation method achieved good performance when using SWI scans of the SN. In addition, we built three machine learning models using radiomics features extracted from manually labeled and automatically segmented VOIs for the diagnosis of PD. The LR model attained the best performance and the greatest stability, both of which were significantly higher than those of visual analysis. On the test cohort, the AUCs of the LR models based on different sequences (T2WI, SWI, or T2WI + SWI) for the segmentation VOIs were all close to those for manually labeled VOIs, and there were no significant differences between the manually labeled





**Figure 5** Details of the features associated with PD and the weighting coefficients of different datasets after LASSO screening.  $SWI_{VOI}$ , features of the manual labeling on SWI;  $SWI_{seg}$ , features of the segmentation on SWI;  $T2WI_{VOI}$ , features of the manual labeling on T2WI;  $T2WI_{seg}$ , features of the segmentation on T2WI; SWI, susceptibility-weighted imaging; T2WI, T2-weighted imaging; PD, Parkinson disease; LASSO, least absolute shrinkage and selection operator.

and automatically segmented VOIs. Overall, our findings suggest that the Swin-Unet network can achieve good accuracy in the segmentation of the SN on SWI and T2WI and that our approach has high PD diagnostic value when only using T2WI, illustrating the potential for an automatic and fast PD diagnosis.

Uchida *et al.* (27,28) demonstrated that cognitive impairment in PD is associated with cerebral iron burden and striatal iron accumulations are correlated with neurophysiological signs in patients with PD. With the exception of the striatal iron accumulations, loss of dopaminergic neurons in the SN is known to occur in clinical parkinsonism (29,30). *NI* refers to an area with two hypointense tails with a hyperintense middle; its shape can be visualized as a “swallow tail” sign on SWI or T2WI (31,32). Therefore, previous studies have suggested that the absence of the swallow tail sign may have the potential to differentiate patients with PD from HCs (33). Radiomic features extracted from SWI have also been reported as biomarkers for diagnosing PD (14). However, manual segmentation on high-resolution images (SWI) is a highly laborious and

time-consuming process. Additionally, the SN is less clear on T2WI than on SWI, a major limitation for clinical trials. In this paper, we propose an efficient deep learning model (Swin-Unet) for SN segmentation. The Swin-Unet model includes an encoder, bottleneck, decoder, and skip connections (34). As the self-attention mechanism in the transformer block can extract more image information than can convolution, the Swin module in Swin-Unet represents an improvement in encoding ability over U-Net. Previous studies have demonstrated that Swin-Unet is superior to traditional U-Net on several medical image datasets (35,36). The results of this study showed that Swin-Unet achieved a better Dice coefficient than did U-Net in SWI segmentation (0.832 *vs.* 0.712), which is consistent with previous reports in the literature. In Swin-Unet, which is an Unet-like pure transformer for medical image segmentation, the tokenized image patches are fed into the transformer-based U-shaped encoder-decoder architecture with skip connections for local-global semantic feature learning. Therefore, the Swin-Unet ensures both high segmentation accuracy as well as robustness and generalizability.

**Table 3** The performance metrics of the classifiers using different modal features

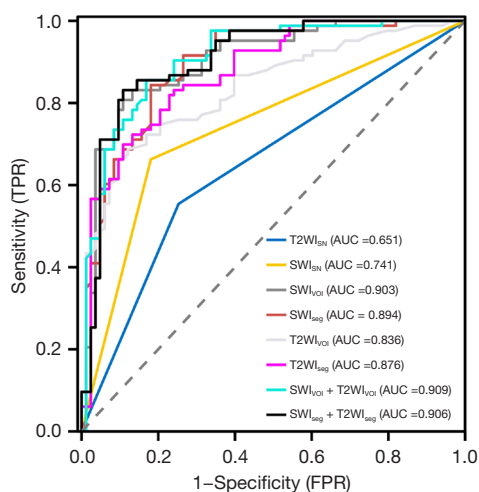
Model	Classifier	Accuracy	Sensitivity	Specificity	AUC	F1 score	RSD% in AUC
Visual analysis	SWI <sub>SN</sub>	0.750*	0.759*	0.741*	0.750*	0.748*	0.08*
	T2WI <sub>SN</sub>	0.612	0.586	0.638	0.612	0.610	0.09
SWI <sub>VOI</sub>	SVM	0.956	0.930	1.000	0.971	0.951	0.05
	LR	0.955*	0.943*	0.975*	0.974*	0.952*	0.04*
	RF	0.969	0.963	0.975	0.966	0.967	0.06
SWI <sub>seg</sub>	SVM	0.906	0.967	0.843	0.913	0.894	0.08
	LR	0.924*	0.933*	0.917*	0.944*	0.909*	0.05*
	RF	0.915	0.930	0.900	0.927	0.902	0.06
T2WI <sub>VOI</sub>	SVM	0.849	0.873	0.817	0.820	0.819	0.15
	LR	0.864*	0.891*	0.800*	0.819*	0.861*	0.02*
	RF	0.851	0.810	0.833	0.781	0.850	0.11
T2WI <sub>seg</sub>	SVM	0.855	0.867	0.843	0.844	0.852	0.09
	LR	0.864*	0.813*	0.867*	0.852*	0.859*	0.07*
	RF	0.865	0.860	0.867	0.881	0.861	0.10
SWI <sub>VOI</sub> + T2WI <sub>VOI</sub>	SVM	0.916	0.931	0.900	0.912	0.905	0.13
	LR	0.935*	0.933*	0.908*	0.927*	0.925*	0.05*
	RF	0.900	0.931	0.916	0.908	0.921	0.09
SWI <sub>seg</sub> + T2WI <sub>seg</sub>	SVM	0.909	0.913	0.908	0.910	0.901	0.11
	LR	0.921*	0.935*	0.910*	0.917*	0.921*	0.04*
	RF	0.918	0.912	0.903	0.902	0.901	0.17

\*, the best results. AUC, area under the curve; RSD, relative standard deviation; SWI<sub>SN</sub>, substantia nigra on SWI; T2WI<sub>SN</sub>, substantia nigra on T2WI; SWI<sub>VOI</sub>, features of the manual labeling on SWI; SVM, support vector machine; LR, logistic regression; RF, random forest; SWI<sub>seg</sub>, features of the segmentation on SWI; T2WI<sub>VOI</sub>, features of the manual labeling on T2WI; T2WI<sub>seg</sub>, features of the segmentation on T2WI; SWI, susceptibility-weighted imaging; T2WI, T2-weighted imaging.

Similar to these previous studies (14,15,37), we also used a general radiomics approach to extract features and build models for diagnosing PD. Ren *et al.* (14) demonstrated that predictive radiomics features extracted from the SN on SWI images could reflect the H&Y stage of PD to some extent. In our study, we extracted radiomics features from SN VOIs manually drawn on SWI and automatically segmented on SWI and T2WI. We found that three radiomics features (original\_shape\_Sphericity, original\_shape\_Elongation, and original\_GLCM\_Imc2) were closely correlated with a PD diagnosis after LASSO screening with high weighting coefficients. GLCM expresses the distribution of neighboring voxels and can reflect the signal mixing degree of the lesions by means of the relative relationship between the distribution and the site of the gray level,

which may be important markers of SN homogeneity. Although slice thickness, volume segmentation, and resolution differ between T2WI and SWI, these findings partly reflect the high accuracy of our segmentation method and the preservation of feature information from the SN. This may be attributable to the fact that radiomics can capture tissue properties such as shape and heterogeneity. In contrast to biopsy, which captures only a small portion of heterogeneity at only a single anatomic site, radiomics captures heterogeneity across the entire lesion volume (12).

Machine learning has been widely applied, for instance being used in a variety of biomedical studies (38,39), for automatically detecting road damage (40), for detecting plant diseases (41), and for helping autism-affected children (42). In our study, we tested three machine learning models for



**Figure 6** The receiver operating characteristic curves for diagnosing PD using the LR classifier with the test group. The AUC of the SWI<sub>SN</sub> was significantly higher than that of the T2WI<sub>SN</sub>. There were no significant differences among different sequences (T2WI, SWI, and T2WI + SWI) in the manually labeled VOIs or the automatically segmented VOIs. TPR, true positive rate; FPR, false positive rate; T2WI<sub>SN</sub>, substantia nigra on T2WI; AUC, area under curve; SWI<sub>SN</sub>, substantia nigra on SWI; SWI<sub>VOI</sub>, features of the manual labeling on SWI; SWI<sub>seg</sub>, features of the segmentation on SWI; T2WI<sub>VOI</sub>, features of the manual labeling on T2WI; T2WI<sub>seg</sub>, features of the segmentation on T2WI; SWI, susceptibility-weighted imaging; T2WI, T2-weighted imaging; VOI, volume of interest; PD, Parkinson disease; LR, logistic regression.

PD diagnosis: SVM, LR, and RF. Different classification models have variable performances in PD diagnosis, and in our study, the LR model outperformed the other ML methods. Binary LR is a traditional method for estimating the probability of a binary response based on one or more independent variables, providing not discrete outputs but probabilities associated with each observation (43). In addition, we found that the LR model was the most stable of all models analyzed. Therefore, LR was chosen as the final classifier. Notably, 10-fold cross-validation provided a more thorough control on classifier accuracy compared to commonly applied within-sample regression or leave-one-out cross-validation. The RSD obtained with the 10-fold cross-validation indicated that the classifications were rather robust. We validated the model with the test group and found that the AUC of the LR model based on SWI was slightly high than that of T2WI. One possible reason for this is that SWI can provide information on iron deposition in the SN (44). The AUCs of the LR model based on VOI segmentation on both SWI and T2WI were close to the AUC of manually labeled VOIs on SWI, and there were no significant differences among the three models. The results of our study confirmed that a machine learning model based on conventional MRI is capable of automatically diagnosing PD and has high generalizability. To our knowledge, no reported machine learning study has described the establishment of a PD diagnostic model derived from SWI and a subsequent comparison with models derived from

**Table 4** The performance metrics of the LR classifier using different modal features in the test group

Model	Accuracy	Sensitivity	Specificity	AUC	Recall	F1 score	P value
Visual analysis							<0.01
T2WI <sub>SN</sub>	0.651	0.554	0.747	0.651	0.582	0.615	
SWI <sub>SN</sub>	0.741	0.663	0.819	0.741	0.713	0.727	
Manual labeling							>0.05
SWI <sub>VOI</sub>	0.849	0.831	0.867	0.903	0.806	0.827	
T2WI <sub>VOI</sub>	0.783	0.687	0.880	0.894	0.734	0.758	
SWI <sub>VOI</sub> + T2WI <sub>VOI</sub>	0.843	0.855	0.831	0.909	0.818	0.830	
Segmentation VOI							>0.05
SWI <sub>seg</sub>	0.831	0.807	0.892	0.894	0.806	0.818	
T2WI <sub>seg</sub>	0.801	0.819	0.771	0.876	0.785	0.793	
SWI <sub>seg</sub> + T2WI <sub>seg</sub>	0.861	0.831	0.892	0.906	0.829	0.845	

LR, logistic regression; AUC, area under the curve; T2WI<sub>SN</sub>, substantia nigra on T2WI; SWI<sub>SN</sub>, substantia nigra on SWI; SWI<sub>VOI</sub>, features of the manual labeling on SWI; T2WI<sub>VOI</sub>, features of the manual labeling on T2WI; VOI, volume of interest; SWI<sub>seg</sub>, features of the segmentation on SWI; T2WI<sub>seg</sub>, features of the segmentation on T2WI; SWI, susceptibility-weighted imaging; T2WI, T2-weighted imaging.

T2WI. Moreover, this is the first machine learning study that has investigated the automatic diagnosis of PD by using and comparing SWI and T2WI. In addition, quantitative susceptibility mapping (QSM), as a noninvasive magnetic resonance technique, has been used to quantify local tissue susceptibility with high spatial resolution and particularly sensitive to the presence of iron (45). Previous studies have demonstrated that QSM value is an auxiliary biomarker for the early evaluation of cognitive decline in patients with PD (46). Machine learning based on QSM may have better differentiation performance, which should be further explored in the future studies.

Our study had several limitations. First, the patient population in the included studies was relatively small. However, it should be noted that the current patient number was determined based on those of previous studies on machine learning techniques for PD neuroimaging (21,47,48). In addition, a large sample size is usually necessary to avoid overfitting if a deep learning method is being used for diagnosis. Therefore, we used radiomics and a machine learning algorithm to diagnose PD. In a later phase of our research, the sample size will be further increased to diagnose PD using a deep learning method, the results of which will be compared with the those of this study. Second, we employed a single-center cohort design, and different external patient populations with different MRI scanners from multiple centers are needed to validate the diagnostic efficacy of our proposed model. Third, we did not manually draw the VOIs on T2WI because the SN was unclear, and the SN VOI on T2WI was obtained using only the segmentation method. Therefore, the segmentation efficacy for T2WI was not obtained. Structural MRI (T1WI and T2WI) is usually unremarkable in PD (49), which may be because T2WI is acquired using a standard 6-mm slice thickness. However, it should be noted that we individually checked the SN VOIs after segmentation and excluded images with significant errors for segmentation. Future work involving the scanning of a specific T2WI sequence with a 2-mm slice thickness is needed to further verify the diagnostic efficacy. Finally, the proposed algorithm in our study depends on several hyperparameters that need to be fine-tuned. Additional studies should be performed to determine the effect of hyperparameter selection on the overall performance.

## Conclusions

We developed an automated deep learning model with

the Swin-UNET network for segmenting SN areas on SWI, mapped the SWI segmentation voxels onto T2WI, and further diagnosed PD from HCs using a machine learning algorithm. We found that the LR model based on automatic VOI segmentation of SWI, T2WI and SWI + T2WI had a similar performance to that of a model based on manually labeled VOIs. The method proposed here may be feasible and useful in diagnosing PD using deep learning and machine learning techniques based on SWI or T2WI and potentially serving as a powerful and valuable tool for automatic and rapid PD diagnosis in the clinic. Its use may facilitate a more efficient clinical treatment trial design and could also guide clinical care via earlier intervention.

## Acknowledgments

*Funding:* This study was supported in part by the Natural Science Foundation of Jiangsu Province (No. BK20201118), the Jiangsu Provincial Special Program of Medical Science (No. BE2021604), the National Natural Science Foundation of China (No. 82202128), the Medical Research Project of Jiangsu Provincial Health Commission (No. M2022012), and the Xinghuo Talent Program of Nanjing First Hospital.

## Footnote

*Reporting Checklist:* The authors have completed the TRIPOD reporting checklist. Available at <https://qims.amegroups.com/article/view/10.21037/qims-24-27/rc>

*Conflicts of Interest:* All authors have completed the ICMJE uniform disclosure form (available at <https://qims.amegroups.com/article/view/10.21037/qims-24-27/coif>). S.D. is an employee of GE HealthCare (Shanghai, China). The other authors have no conflicts of interest to declare.

*Ethical Statement:* The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. This study was conducted in accordance with the Declaration of Helsinki (as revised in 2013) and was approved by the Ethics Committee of the Nanjing Medical University (No. 2019-664). The requirement for individual consent was waived by the ethics committee due to the retrospective nature of the analysis.

*Open Access Statement:* This is an Open Access article distributed in accordance with the Creative Commons

Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: <https://creativecommons.org/licenses/by-nc-nd/4.0/>.

## References

- Bryois J, Skene NG, Hansen TF, Kogelman LJA, Watson HJ, Liu Z; Brueggeman L, Breen G, Bulik CM, Arenas E, Hjerling-Leffler J, Sullivan PF. Genetic identification of cell types underlying brain complex traits yields insights into the etiology of Parkinson's disease. *Nat Genet* 2020;52:482-93.
- De Pablo-Fernández E, Lees AJ, Holton JL, Warner TT. Prognosis and Neuropathologic Correlation of Clinical Subtypes of Parkinson Disease. *JAMA Neurol* 2019;76:470-9.
- Khan AR, Hiebert NM, Vo A, Wang BT, Owen AM, Seergobin KN, MacDonald PA. Biomarkers of Parkinson's disease: Striatal sub-regional structural morphometry and diffusion MRI. *Neuroimage Clin* 2019;21:101597.
- Prasuhn J, Strautz R, Lemmer F, Dreischmeier S, Kasten M, Hanssen H, Heldmann M, Brüggemann N. Neuroimaging Correlates of Substantia Nigra Hyperechogenicity in Parkinson's Disease. *J Parkinsons Dis* 2022;12:1191-200.
- Vitali P, Pan MI, Palesi F, Germani G, Faggioli A, Anzalone N, Francaviglia P, Minafra B, Zangaglia R, Pacchetti C, Gandini Wheeler-Kingshott CAM. Substantia Nigra Volumetry with 3-T MRI in De Novo and Advanced Parkinson Disease. *Radiology* 2020;296:401-10.
- Prasuhn J, Neumann A, Strautz R, Dreischmeier S, Lemmer F, Hanssen H, Heldmann M, Schramm P, Brüggemann N. Clinical MR imaging in Parkinson's disease: How useful is the swallow tail sign? *Brain Behav* 2021;11:e02202.
- Schwarz ST, Afzal M, Morgan PS, Bajaj N, Gowland PA, Auer DP. The 'swallow tail' appearance of the healthy nigrosome - a new accurate test of Parkinson's disease: a case-control and retrospective cross-sectional MRI study at 3T. *PLoS One* 2014;9:e93814.
- Cao Q, Han X, Tang D, Qian H, Yan K, Shi X, Li Y, Zhang J. Diagnostic value of combined magnetic resonance imaging techniques in the evaluation of Parkinson disease. *Quant Imaging Med Surg* 2023;13:6503-16.
- Rizzo G, De Blasi R, Capozzo R, Tortelli R, Barulli MR, Liguori R, Grasso D, Logroscino G. Loss of Swallow Tail Sign on Susceptibility-Weighted Imaging in Dementia with Lewy Bodies. *J Alzheimers Dis* 2019;67:61-5.
- Dong D, Tang L, Li ZY, Fang MJ, Gao JB, Shan XH, Ying XJ, Sun YS, Fu J, Wang XX, Li LM, Li ZH, Zhang DF, Zhang Y, Li ZM, Shan F, Bu ZD, Tian J, Ji JF. Development and validation of an individualized nomogram to identify occult peritoneal metastasis in patients with advanced gastric cancer. *Ann Oncol* 2019;30:431-8.
- Yu J, Deng Y, Liu T, Zhou J, Jia X, Xiao T, Zhou S, Li J, Guo Y, Wang Y, Zhou J, Chang C. Lymph node metastasis prediction of papillary thyroid carcinoma based on transfer learning radiomics. *Nat Commun* 2020;11:4807.
- Mayerhoefer ME, Materka A, Langs G, Häggström I, Szczypiński P, Gibbs P, Cook G. Introduction to Radiomics. *J Nucl Med* 2020;61:488-95.
- Li XN, Hao DP, Qu MJ, Zhang M, Ma AB, Pan XD, Ma AJ. Development and Validation of a Plasma FAM19A5 and MRI-Based Radiomics Model for Prediction of Parkinson's Disease and Parkinson's Disease With Depression. *Front Neurosci* 2021;15:795539.
- Ren Q, Wang Y, Leng S, Nan X, Zhang B, Shuai X, Zhang J, Xia X, Li Y, Ge Y, Meng X, Zhao C. Substantia Nigra Radiomics Feature Extraction of Parkinson's Disease Based on Magnitude Images of Susceptibility-Weighted Imaging. *Front Neurosci* 2021;15:646617.
- Salmanpour MR, Shamsaei M, Saberi A, Hajianfar G, Soltanian-Zadeh H, Rahmim A. Robust identification of Parkinson's disease subtypes using radiomics and hybrid machine learning. *Comput Biol Med* 2021;129:104142.
- Rahmim A, Huang P, Shenkov N, Fotouhi S, Davoodi-Bojd E, Lu L, Mari Z, Soltanian-Zadeh H, Sossi V. Improved prediction of outcome in Parkinson's disease using radiomics analysis of longitudinal DAT SPECT images. *Neuroimage Clin* 2017;16:539-44.
- Li D, Xiao C, Liu Y, Chen Z, Hassan H, Su L, Liu J, Li H, Xie W, Zhong W, Huang B. Deep Segmentation Networks for Segmenting Kidneys and Detecting Kidney Stones in Unenhanced Abdominal CT Images. *Diagnostics (Basel)* 2022. doi: 10.3390/diagnostics12081788.
- Gu Z, Cheng J, Fu H, Zhou K, Hao H, Zhao Y, Zhang T, Gao S, Liu J. CE-Net: Context Encoder Network for 2D Medical Image Segmentation. *IEEE Trans Med Imaging* 2019;38:2281-92.
- Verma R, Kumar N, Patil A, Kurian NC, Rane S, Graham S, et al. MoNuSAC2020: A Multi-Organ Nuclei

- Segmentation and Classification Challenge. *IEEE Trans Med Imaging* 2021;40:3413-23.
20. Zunair H, Ben Hamza A. Sharp U-Net: Depthwise convolutional network for biomedical image segmentation. *Comput Biol Med* 2021;136:104699.
  21. Liu P, Wang H, Zheng S, Zhang F, Zhang X. Parkinson's Disease Diagnosis Using Neostriatum Radiomic Features Based on T2-Weighted Magnetic Resonance Imaging. *Front Neurol* 2020;11:248.
  22. Cao H, Wang Y, Chen J, Jiang D, Zhang X, Tian Q, Wang M. Swin-Unet: Unet-Like Pure Transformer for Medical Image Segmentation. In: Karlinsky L, Michaeli T, Nishino K. editors. *Computer Vision – ECCV 2022 Workshops. ECCV 2022. Lecture Notes in Computer Science*, Springer, 2022;13803:205-18.
  23. Dan Y, Zhu Z, Jin W, Li Z. S-Swin Transformer: simplified Swin Transformer model for offline handwritten Chinese character recognition. *PeerJ Comput Sci* 2022;8:e1093.
  24. Hughes AJ, Daniel SE, Kilford L, Lees AJ. Accuracy of clinical diagnosis of idiopathic Parkinson's disease: a clinico-pathological study of 100 cases. *J Neurol Neurosurg Psychiatry* 1992;55:181-4.
  25. Calne DB, Snow BJ, Lee C. Criteria for diagnosing Parkinson's disease. *Ann Neurol* 1992;32 Suppl:S125-7.
  26. SzeWCzyk-Krolikowski K, Tomlinson P, Nithi K, Wade-Martins R, Talbot K, Ben-Shlomo Y, Hu MT. The influence of age and gender on motor and non-motor features of early Parkinson's disease: initial findings from the Oxford Parkinson Disease Center (OPDC) discovery cohort. *Parkinsonism Relat Disord* 2014;20:99-105.
  27. Uchida Y, Kan H, Sakurai K, Arai N, Kato D, Kawashima S, Ueki Y, Matsukawa N. Voxel-based quantitative susceptibility mapping in Parkinson's disease with mild cognitive impairment. *Mov Disord* 2019;34:1164-73.
  28. Uchida Y, Kan H, Sakurai K, Inui S, Kobayashi S, Akagawa Y, Shibuya K, Ueki Y, Matsukawa N. Magnetic Susceptibility Associates With Dopaminergic Deficits and Cognition in Parkinson's Disease. *Mov Disord* 2020;35:1396-405.
  29. Norel X, Sugimoto Y, Ozen G, Abdelazeem H, Amgoud Y, Bouhadoun A, Bassiouni W, Goepp M, Mani S, Manikpurage HD, Senbel A, Longrois D, Heinemann A, Yao C, Clapp LH. International Union of Basic and Clinical Pharmacology. CIX. Differences and Similarities between Human and Rodent Prostaglandin E(2) Receptors (EP1-4) and Prostacyclin Receptor (IP): Specific Roles in Pathophysiologic Conditions. *Pharmacol Rev* 2020;72:910-68.
  30. Schaffner A, Li X, Gomez-Llorente Y, Leandrou E, Memou A, Clemente N, Yao C, Afsari F, Zhi L, Pan N, Morohashi K, Hua X, Zhou MM, Wang C, Zhang H, Chen SG, Elliott CJ, Rideout H, Ubarretxena-Belandia I, Yue Z. Vitamin B(12) modulates Parkinson's disease LRRK2 kinase activity through allosteric regulation and confers neuroprotection. *Cell Res* 2019;29:313-29.
  31. Cheng Z, Zhang J, He N, Li Y, Wen Y, Xu H, Tang R, Jin Z, Haacke EM, Yan F, Qian D. Radiomic Features of the Nigrosome-1 Region of the Substantia Nigra: Using Quantitative Susceptibility Mapping to Assist the Diagnosis of Idiopathic Parkinson's Disease. *Front Aging Neurosci* 2019;11:167.
  32. Schwarz ST, Mougou O, Xing Y, Blazejewska A, Bajaj N, Auer DP, Gowland P. Parkinson's disease related signal change in the nigrosomes 1-5 and the substantia nigra using T2\* weighted 7T MRI. *Neuroimage Clin* 2018;19:683-9.
  33. Pang H, Yu Z, Li R, Yang H, Fan G. MRI-Based Radiomics of Basal Nuclei in Differentiating Idiopathic Parkinson's Disease From Parkinsonian Variants of Multiple System Atrophy: A Susceptibility-Weighted Imaging Study. *Front Aging Neurosci* 2020;12:587250.
  34. Liu P, Song Y, Chai M, Han Z, Zhang Y. Swin-UNet++: A Nested Swin Transformer Architecture for Location Identification and Morphology Segmentation of Dimples on 2.25Cr1Mo0.25V Fractured Surface. *Materials (Basel)* 2021. doi: 10.3390/ma14247504.
  35. Islam MN, Hasan M, Hossain MK, Alam MGR, Uddin MZ, Soylu A. Vision transformer and explainable transfer learning models for auto detection of kidney cyst, stone and tumor from CT-radiography. *Sci Rep* 2022;12:11440.
  36. Liu Y, Zhao J, Luo Q, Shen C, Wang R, Ding X. Automated classification of cervical lymph-node-level from ultrasound using Depthwise Separable Convolutional Swin Transformer. *Comput Biol Med* 2022;148:105821.
  37. Shu ZY, Cui SJ, Wu X, Xu Y, Huang P, Pang PP, Zhang M. Predicting the progression of Parkinson's disease using conventional MRI and machine learning: An application of radiomic biomarkers in whole-brain white matter. *Magn Reson Med* 2021;85:1611-24.
  38. Nguyen HS, Ho DKN, Nguyen NN, Tran HM, Tam KW, Le NQK. Predicting EGFR Mutation Status in Non-Small Cell Lung Cancer Using Artificial Intelligence: A Systematic Review and Meta-Analysis. *Acad Radiol* 2024;31:660-83.
  39. Le NQK. Hematoma expansion prediction: still navigating the intersection of deep learning and radiomics. *Eur*

- Radiol 2024;34:2905-7.
40. Roy AM, Bhaduri J. DenseSPH-YOLOv5: An automated damage detection model based on DenseNet and Swin-Transformer prediction head-enabled YOLOv5 with attention mechanism. *Advanced Engineering Informatics* 2023;56:102007.
  41. Roy AM, Bose R, Bhaduri J. A fast accurate fine-grain object detection model based on YOLOv4 deep neural network. *Neural Comput & Applic* 2022;34:3895-921.
  42. Singh A, Raj K, Kumar T, Verma S, Roy AM. Deep Learning-Based Cost-Effective and Responsive Robot for Autism Treatment. *Drones* 2023;7:81.
  43. Lee H, Lee EJ, Ham S, Lee HB, Lee JS, Kwon SU, Kim JS, Kim N, Kang DW. Machine Learning Approach to Identify Stroke Within 4.5 Hours. *Stroke* 2020;51:860-6.
  44. Mazzucchi S, Frosini D, Costagli M, Del Prete E, Donatelli G, Cecchi P, Migaleddu G, Bonuccelli U, Ceravolo R, Cosottini M. Quantitative susceptibility mapping in atypical Parkinsonisms. *Neuroimage Clin* 2019;24:101999.
  45. Uchida Y, Kan H, Sakurai K, Oishi K, Matsukawa N. Quantitative susceptibility mapping as an imaging biomarker for Alzheimer's disease: The expectations and limitations. *Front Neurosci* 2022;16:938092.
  46. Yan Y, Wang Z, Wei W, Yang Z, Guo L, Wang Z, Wei X. Correlation of brain iron deposition and freezing of gait in Parkinson's disease: a cross-sectional study. *Quant Imaging Med Surg* 2023;13:7961-72.
  47. Park KW, Lee EJ, Lee JS, Jeong J, Choi N, Jo S, Jung M, Do JY, Kang DW, Lee JG, Chung SJ. Machine Learning-Based Automatic Rating for Cardinal Symptoms of Parkinson Disease. *Neurology* 2021;96:e1761-9.
  48. Srinivasan S, Ramadass P, Mathivanan SK, Panneer Selvam K, Shivahare BD, Shah MA. Detection of Parkinson disease using multiclass machine learning approach. *Sci Rep* 2024;14:13813.
  49. Tolosa E, Garrido A, Scholz SW, Poewe W. Challenges in the diagnosis of Parkinson's disease. *Lancet Neurol* 2021;20:385-97.

**Cite this article as:** Wang T, Wang Y, Zhu H, Liu Z, Chen YC, Wang L, Duan S, Yin X, Jiang L. Automatic substantia nigra segmentation with Swin-Unet in susceptibility- and T2-weighted imaging: application to Parkinson disease diagnosis. *Quant Imaging Med Surg* 2024;14(9):6337-6351. doi: 10.21037/qims-24-27