Radiomic Features of the Nigrosome-1 Region of the Substantia Nigra: Using Quantitative Susceptibility Mapping to Assist the Diagnosis of Idiopathic Parkinson's Disease

Zenghui Cheng [1†], Jiping Zhang [2†], Naying He [1], Yan Li [1], Yaofeng Wen [2], Hongmin Xu [1], Rongbiao Tang [1], Zhijia Jin [1], E. Mark Haacke [1,3], Fuhua Yan [1*] and Dahong Qian [4*]

[1] Department of Radiology, Ruijin Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China, [2] School of Biomedical Engineering, Shanghai Jiao Tong University, Shanghai, China, [3] Department of Radiology, Wayne State University, Detroit, MI, United States, [4] Institute of Medical Robotics, Shanghai Jiao Tong University, Shanghai, China

Introduction: The loss of nigrosome-1, which is also referred to as the swallow tail sign (STS) in T2*-weighted iron-sensitive magnetic resonance imaging (MRI), has recently emerged as a new biomarker for idiopathic Parkinson's disease (IPD). However, consistent recognition of the STS is difficult due to individual variations and different imaging parameters. Radiomics might have the potential to overcome these shortcomings. Therefore, we chose to explore whether radiomic features of nigrosome-1 of substantia nigra (SN) based on quantitative susceptibility mapping (QSM) could help to differentiate IPD patients from healthy controls (HCs).

Methods: Three-dimensional multi-echo gradient-recalled echo images ($0.86 \times 0.86 \times 1.00$ mm^3) were obtained at 3.0-T MRI for QSM in 87 IPD patients and 77 HCs. Regions of interest (ROIs) of the SN below the red nucleus were manually drawn on both sides, and subsequently, volumes of interest (VOIs) were segmented (these ROIs included four 1-mm slices). Then, 105 radiomic features (including 18 first-order features, 13 shape features, and 74 texture features) of bilateral VOIs in the two groups were extracted. Forty features were selected according to the ensemble feature selection method, which combined analysis of variance, random forest, and recursive feature elimination. The selected features were further utilized to distinguish IPD patients from HC using the SVM classifier with 10 rounds of 3-fold cross-validation. Finally, the representative features were analyzed using an unpaired t-test with Bonferroni correction and correlated with the UPDRS-III scores.

Results: The classification results from SVM were as follows: area under curve (AUC): 0.96 ± 0.02; accuracy: 0.88 ± 0.03; sensitivity: 0.89 ± 0.06; and specificity: 0.87 ± 0.07. Five representative features were selected to show their detailed difference between IPD patients and HCs: 10th percentile and median in IPD patients were higher than those

in HCs (all $p < 0.00125$), while Gray Level Run Length Matrix (GLRLM)-Long Run Low Gray Level Emphasis, Gray Level Size Zone Matrix (GLSZM)–Gray Level Non-Uniformity, and volume (all $p < 0.00125$) in IPD patients were lower than those in HCs. The 10th percentile was positively correlated with UPDRS-III score ($r = 0.35$, $p = 0.001$).

Conclusion: Radiomic features of the nigrosome-1 region of SN based on QSM could be useful in the diagnosis of IPD and could serve as a surrogate marker for the STS.

Keywords: nigrosome-1, substantia nigra, Parkinson's disease, quantitative susceptibility mapping, radiomics, support vector machine

INTRODUCTION

Parkinson's disease (PD) is a common neurodegenerative movement disorder, and mainly characterized by a loss of dopaminergic neurons and iron accumulation in the substantia nigra (SN), pathologically (Damier et al., 1999). The dopaminergic neuronal loss and iron deposition were reported to occur in nigrosomes of the SN pars compacta (SNpc) at the initial stage of PD, especially in the nigrosome-1, the largest among the five subdivisions (Damier et al., 1999; Lehericy et al., 2014). Therefore, imaging the nigrosome-1 using iron-sensitive magnetic resonance imaging (MRI), for example, T2*-weighted imaging, has recently been investigated and validated as a possible biomarker for idiopathic Parkinson's disease (IPD) (Noh et al., 2015; Reiter et al., 2015).

Recognizing the nigrosome-1 has been possible thanks to the presence of high iron signal surrounding it that produces what is referred to as the "swallow tail sign" (STS) (Schwarz et al., 2014). The loss of the STS is thought to be due to the increase in iron content subsequent to the depigmentation of the neuromelanin in the nigrosome-1 territory. The reported sensitivity, specificity, and accuracy of the loss of nigrosome-1 hyperintensity or the STS varied from 94 to 100%, 84.6 to 94.4%, and 94.6 to 96%, respectively (Noh et al., 2015; Reiter et al., 2015; Mahlknecht et al., 2017; Stezin et al., 2018). However, consistent recognition of the STS among reviewers has been difficult due to individual differences in the shape of the nigrosome-1 territory and to the choice of imaging parameters such as scanning plane, resolution, signal-to-noise ratio (SNR), and echo time, even on the ultra-high 7.0 T system MRI (Schmidt et al., 2017; Kim et al., 2018).

Conventional quantitative imaging (e.g., R2*) of SN was reported to have the potential to differentiate PD from healthy controls (HCs) by quantification of the local iron content (Martin et al., 2008; Du et al., 2011), which partially avoids some of the shortcomings referred above. Nevertheless, the quantification can be affected by several factors including intravoxel spin dephasing, non-locality of phase and tissue susceptibility, as well as field strength (Schweser et al., 2011; Li et al., 2012; He et al., 2015). Quantitative susceptibility mapping (QSM) has been proven to be a sensitive and reliable quantitative method reflecting the local tissue susceptibility with much better contrast compared to R2* or T2 methods (Langkammer et al., 2012; He et al., 2015; Murakami et al., 2015; Wang and Liu, 2015; Du et al., 2016). However, diagnostic accuracy did not surpass that of conventional imaging, which might be accredited to the

intergroup overlap of the susceptibility value when utilizing the mean value of a given structure (Kim et al., 2018).

Radiomics is a recently developed promising technique, which can be utilized to mine feature information (such as density, shape, size, and texture) of a certain region of interest underlying medical images and it can generate a great deal of quantitative features including some reflecting the inter-voxel spatiality (Feng et al., 2018). Radiomics has been mainly applied to various tumors, and its role in diagnosis, treatment evaluation, and prognosis has to some degree been verified (Cameron et al., 2016; Nie et al., 2016; Zhang et al., 2017).

To date, few investigations of radiomics on neurodegenerative disease have been carried out. We hypothesized that it might have the potential to overcome and/or complement the shortcomings of the current nigrosome-1 imaging referred to above. Therefore, our purpose was to explore whether radiomic features of the iron content in the nigrosome-1-containing part of the SN based on QSM data could help to differentiate IPD patients from HCs.

MATERIALS AND METHODS
Participants

This study was approved by our institutional review board and written informed consent was given and signed by all the participants. The diagnosis of IPD was based on the UK Parkinson's Disease Society Brain Bank Clinical Diagnostic Criteria (Hughes et al., 1992). The exclusion criteria were as follows: (1) secondary or atypical parkinsonism; (2) dementia: Mini-Mental State Examination (MMSE) score <24; (3) a history of cerebrovascular disease (e.g., infarction, hemorrhage), brain tumor, head trauma, or any other type of psychiatric disorders; (4) a history of medication known to cause parkinsonism or affect clinical assessment; or (5) contraindications to an MRI examination. HCs with gender and age matched from local communities were recruited according to the following inclusion criteria: (1) older than 40 years, without family history of movement disorders; (2) without any neurological or psychiatric disorders; and (3) an MMSE score of at least 24. There were 87 IPD patients (41 males and 46 females, aged 60.9 ± 8.1 years) and 79 HCs (43 males and 34 females, aged 63.4 ± 7.3 years) enrolled according to the above criteria from March 2012 to June 2015. Two HCs with imaging quality score higher than two were excluded for further analysis (see the grading system in the section Image Reconstruction).

MRI Acquisition

All participants were scanned on a 3.0-T MRI system (Signa HDxt; GE Healthcare, Milwaukee, WI, USA) equipped with an eight-channel phased-array head coil. Foam pads and earplugs were applied to reduce head movement and scanner noise, respectively. A three-dimensional multi-echo gradient-echo (GRE) sequence was used to generate $T2^*$-weighted images with the following parameters: TR = 59.3 ms; TE1 = 2.7 ms, ΔTE = 2.9 ms, number of echoes = 16, flip angle = 12°, FOV = 22 × 22 cm^2, matrix = 256 × 256, resolution = 0.86 × 0.86 × 1.0 mm^3, bandwidth = 488.28 Hz/pixel, acceleration factor = 2, number of slices = 136, and acquisition time = 10 min 42 s. In addition, conventional sequences including T1-weighted images, T2-weighted fluid-attenuated inversion recovery (FLAIR) images, and diffusion-weighted imaging (DWI) were also acquired to screen for cerebrovascular diseases and other confounding diseases. The whole brain was covered axially parallel to the anterior commissure–posterior commissure (AC–PC) line.

Image Reconstruction

Before reconstruction of QSM, two neuroradiologists (ZC, NH) blindly assessed the quality of magnitude images using the following grading system (He et al., 2015): 1 = very good (little or no artifact); 2 = good (visible artifacts); 3 = poor (considerable motion artifacts); 4 = very poor (significant motion artifacts), 5 = non-diagnostic scan, and reached a consensus by discussion on the disagreements. Subjects with a score higher than two were excluded from further processing. Of the 79 HCs, 75 cases were scored 1; 2 and 2 cases were scored 2 and 3, respectively. Of the IPD cases, 85 cases scored 1 and the remaining 2 cases scored 2. QSM reconstructions were performed utilizing the method reported previously (Li et al., 2011). A brief description follows: (1) phase images were collected from each channel of the coil, and then averaged; (2) background phase variations were removed using SHARP (Sophisticated Harmonic Artifact Reduction for Phase) with a filter radius of 8 (Schweser et al., 2011); and (3) the susceptibility map was generated using an improved LSQR (iLSQR) method (Li et al., 2011, 2015) with the regularization threshold for Laplace filtering being set at 0.04.

Image Analysis and Segmentation

The STS is composed of the nigrosome-1 (hyperintensity on SWI) and its surrounding structures (hypointensity on SWI) (Schwarz et al., 2014). This usually appears at or below the caudal part of the red nucleus (RN) (Massey et al., 2017). This nigrosome-1-containing territory was set as the region of interest (ROI) and the relevant slices were manually drawn and segmented on the QSM data using ITK-SNAP (V3.4.0, http://www.itksnap.org) by one of the investigators (J.Z) blindly from the level of the inferior part of the RN to the inferior part of the SN (**Figure 1**). The most inferior part and the boundary regions of the SN were excluded to avoid partial volume effects. The lateral and medial parts of SN were delineated on a visible border of high-intensity signal of SN on QSM, with boundary voxels excluded. The final volume of interest (VOI) consisted of four slices unilaterally, and the bilateral VOIs were

confirmed by a neuroradiologist (Z.C., with 5 years' experience on neuroimaging diagnosis).

As a comparison, all the subjects were reviewed blindly on QSM for manifestation of STS or not by a neuroradiologist (ZC, with 5 years' experience on neuroimaging diagnosis). Bilateral and unilateral loss of the sign were classified as positive, and bilateral presence of the sign was classified as negative.

Feature Extraction

One hundred and five radiomic features including 18 first-order features, 13 shape-based features, and 74 texture features were extracted from the segmented nigrosome-1-containing SN images. The first-order features describe how the individual susceptibilities are distributed, consisting of mean, 10th percentile, variance, and 15 other features. The shape-based features describe the geometric characteristics of the VOI, such as volume, surface area, and diameter. The textural features express how individual susceptibilities distribute spatially, which are computed according to five defined matrices: Gray Level Co-occurrence Matrix (GLCM), Gray Level Dependence Matrix (GLDM), Gray Level Run Length Matrix (GLRLM), Gray Level Size Zone Matrix (GLSZM), and Neighboring Gray Tone Difference Matrix (NGTDM). The (i, j)th element of GLCM represents the frequency of the combination of gray levels i and j of two pixels, which are separated by a distance of δ along angle θ; GLDM quantifies gray-level dependencies in an image, and gray-level dependency is defined as the number of connected voxels that are dependent on the center voxel within distance δ along angle θ; GLRLM quantifies gray-level runs, which are defined as the length of consecutive pixels that have the same gray-level value within distance δ along angle θ; GLSZM quantifies gray-level zones in an image, and a gray-level zone is defined as the number of connected voxels that share the same gray-level intensity. The NGTDM quantifies the difference between a gray value and the average gray value of its neighbors within distance δ. In our study, we considered one neighboring pixel ($\delta = 1$). As for the parameter angle θ in GLCM, GLDM, GLRLM, we consider all 13 possible directions as $(1, 0, 0)$, $(0, 1, 0)$, $(0, 0, 1)$, $(1, 1, 0)$, $(1, 0, 1)$, $(0, 1, 1)$, $(1, -1, 0)$, $(-1, 1, 0)$, $(1, 0, -1)$, $(-1, 0, 1)$, $(0, -1, 1)$, $(0, 1, -1)$, and $(1, 1, 1)$ for our volumetric data, and the mean values of the features form different directions were used as the final features.

The meaning and equation of each feature from the above mentioned seven groups (first-order features, shape-based features, and textural feature-based GLCM, GLDM, GLRLM, GLSZM, and NGTDM) were described thoroughly in **Supplementary S1**, and 17 representative features were listed in **Table 1**. The feature extraction procedure was performed using the open-source python package pyradiomics (Jjm et al., 2017). All the feature descriptions are consistent to the illustration in their website (https://pyradiomics.readthedocs.io/en/latest/index.html).

Feature Selection

Feature selection was performed to keep the most relevant features. Ensemble feature selection can combine the results from
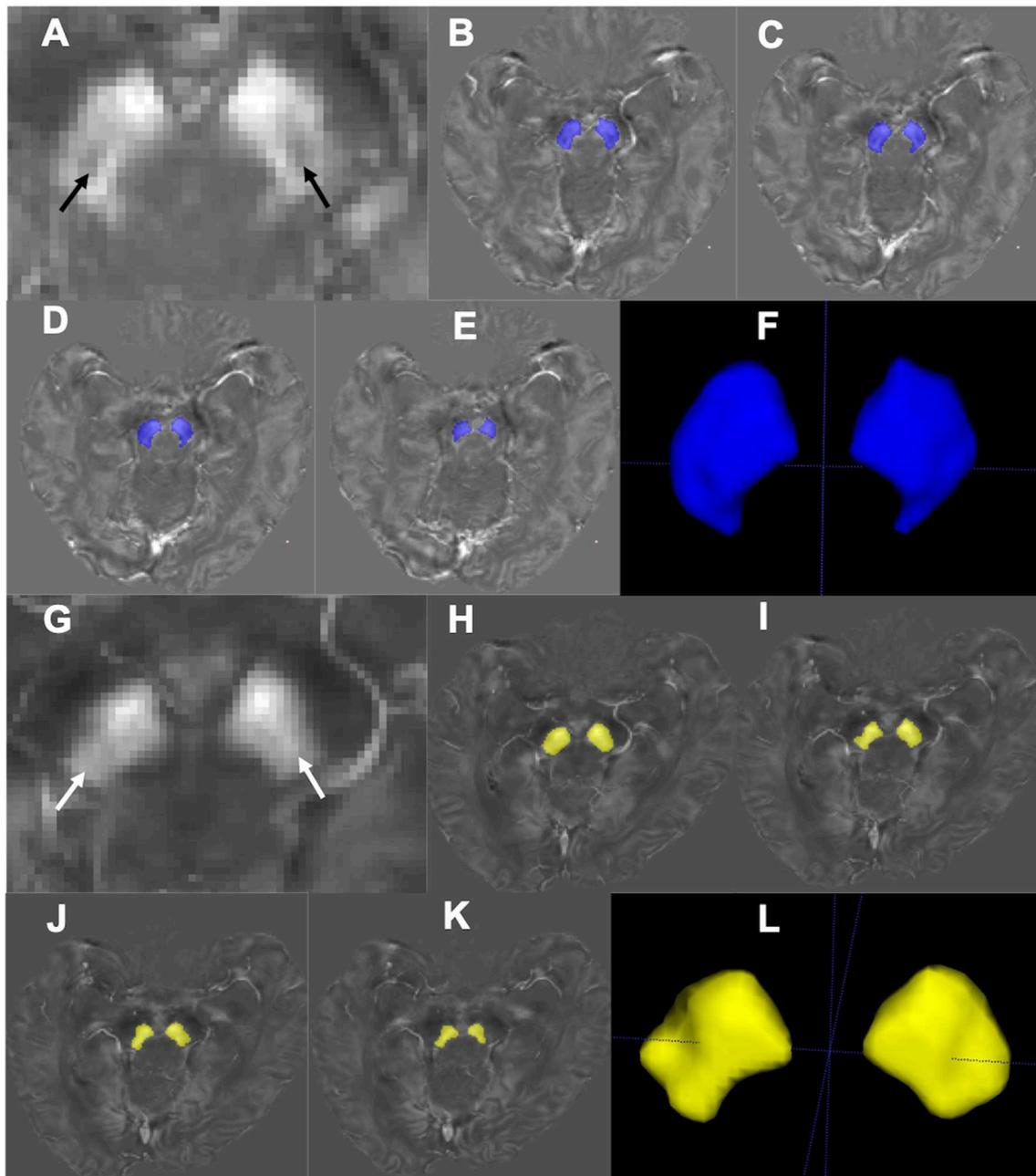
FIGURE 1 | Segmentation of the nigrosome-1-containing region of SN. Representative QSM images of the nigrosome-1 area for a HC (67 years old, male; **A**) and an IPD patient (69 years old, male, H&Y = 2; **G**). The nigrosome-1 area presents as the "swallow tail" sign bilaterally (**A**, black arrow) in the HC, while it cannot be seen in the IPD case (**G**, white arrow) (that is, there is a loss of the "swallow tail" sign). Regions of the SN below the RN were segmented slice by slice **(B–E; H–K)** to generate the three-dimensional nigrosome-1-containing SN **(F,L)**.

different feature selection methods to one major decision, and this can improve the robustness and stability of the final feature selection results (Jong et al., 2004; Abeel et al., 2010; Hoque et al., 2018). Therefore, three different methods were assembled in our feature selection procedure: analysis of variance (ANOVA), random forest, and recursive feature elimination (RFE). The average rank from different methods served as the final feature rank criterion. A schematic overview of our feature selection procedure can be found in **Figure 2**.

ANOVA was used to compare the features' mean values of IPD group with those of HC groups. The ANOVA F value (a measure of significance in F-test) was calculated and served as one criterion to rank features: features with a higher F value were ranked as more important.

TABLE 1 | Description and equation of the 13 representative features from different group.

Features	Description	Equation
Firstorder-minimum	The minimum susceptibility in the VOI	$\min(X)$
Firstorder-10 Percentile	The 10th percentile of the sorted susceptibility in the VOI	$10\text{th\_percentile}(X)$
Median	Median of the sorted susceptibility in the VOI	$\text{Median}(X)$
GLCM-Correlation	The image complexity	$\dfrac{\sum\limits_{i=1}^{N_g}\sum\limits_{j=1}^{N_g}p(i,j)ij-\mu_x\mu_y}{\sigma_x(i)\sigma_y(i)}$
GLCM-Informational Measure of Correlation 1 (InfMCor1)	Complexity of the texture	$\dfrac{HXY-HXY1}{max\{HX,HY\}}$
GLCM-SumEntropy (SumEntrp)	The sum of neighborhood intensity value differences	$\sum\limits_{k=2}^{2N_g}p_{x+y}(k)log_2(p_{x+y}(k))$
GLDM-Dependence Entropy (DepdEntrp)	The randomness of GLDM. Higher Dependence Entropy implies more complex texture	$p(i,j)\log_2(p(i,j))$
GLDM-Dependence Variance (DepdVar)	The variance in dependence size. Higher Dependence Variance implies more heterogeneity in local zone size.	$\sum\limits_{j=1}^{N_d}p(i,j)(j-\mu)^2,\text{where } \mu=\sum\limits_{i=1}^{N_g}\sum\limits_{j=1}^{N_d}jp(i,j)$
GLDM-GrayLevel Non-Uniformity (GryLvNon-Uni)	The similarity of gray-level intensity values in the image. A lower GLN value correlates with a greater similarity in intensity values	$\dfrac{\sum\limits_{i=1}^{N_g}\left(\sum\limits_{j=1}^{N_d}\mathbf{P}(i,j)\right)^2}{N_z}$
GLDM-Dependence Non-uniformity Normalized (DepdNonUni)	The similarity of dependence throughout the image, with a lower value indicating more homogeneity among dependencies of the image.	$\dfrac{\sum\limits_{j=1}^{N_d}\left(\sum\limits_{i=1}^{N_g}\mathbf{P}(i,j)\right)^2}{N_z^2}$
GLRLM-RunEntropy (RunEntrp)	The randomness of run lengths and gray levels. A higher value indicates more heterogeneity in the texture patterns	$p(i,j)\log_2(p(i,j))$
GLRLM-RunLengthNonUniformityNormalized (RunLthNonUni)	The similarity of run lengths throughout the image. A lower value indicates more homogeneity among run lengths of the image	$\dfrac{\sum\limits_{j=1}^{N_r}\left(\sum\limits_{i=1}^{N_g}\mathbf{P}(i,j)\right)^2}{N_r^2}$
GLRLM-HighGrayLevelRunEmphasis (HGLRunEmphs)	The distribution of the higher gray-level values, with a higher value indicating a greater concentration of high gray-level values in the image	$\dfrac{\sum\limits_{i=1}^{N_g}\sum\limits_{j=1}^{N_r}\mathbf{P}(i,j)i^2}{N_r}$
GLRLM-LongRunLowGrayLevelEmphasis (LRunLGREmphs)	The joint distribution of long run lengths with lower gray-level values	$\dfrac{\sum\limits_{i=1}^{N_g}\sum\limits_{j=1}^{N_r}\frac{\mathbf{P}(i,j)j^2}{i^2}}{N_r}$
GLSZM-ZoneEntropy (ZoneEntrp)	The randomness in the distribution of zone sizes and gray levels. A higher value indicates more heterogeneity in the texture patterns	$p(i,j)\log_2(p(i,j))$
GLSZM-GrayLevelNonUniformity (GryLvNonUniS)	The variability of gray-level intensity values in the image, with a lower value indicating more homogeneity in intensity values	$\dfrac{\sum\limits_{i=1}^{N_g}\left(\sum\limits_{j=1}^{N_s}\mathbf{P}(i,j)\right)^2}{N_z}$
Shape-Volume	The volume of the VOI	The voxel number in VOI

*The first part of the feature name is its group. The X is the susceptibility in the VOI; N_g is the number of discrete intensity values in the VOI; $P(i, j)$ is the value in corresponding matrix; N_d and N_z are the number of discrete dependency sizes and the number of dependency zones in GLDM; N_r is the number of discrete run lengths in the GLRLM; μ, σ, HX, HY, and HXY denote the mean, standard deviations, entropy of p_x, entropy of p_y, and entropy of $p(i,j)$ in GLCM, respectively. More details can be found in the **Supplementary S1** or the abovementioned pyradiomics website.*

Random forest is a classification method, which trains a series of classifying decision trees by randomly choosing sub-samples of the dataset and sub-features from all the features and then averages the results from different decision trees to avoid overfitting and improve the accuracy. When constructing the decision trees, random forest provides feature importance intrinsically. The feature's importance degree is measured by the impurity proposed by Breiman (2001). Higher impurity implies that the corresponding feature can influence the predictive results more obviously and will be reckoned more important. In our study, we constructed a random forest with 10 decision trees and used the provided importance degree as another criterion for features ranking.

The RFE method selects features by recursively pruning the least important features from current features set, and the estimator in our RFE method was specified as linear support vector machine (SVM). The estimator not only evaluated the importance of features but also gave feedback on training performance of different feature subsets, so that the curve between training accuracy and feature number (**Figure 3**) could be obtained to determine the optimal number of features. According to the curve, the number was set as 40 in order to reduce the feature dimensionality as much as possible.
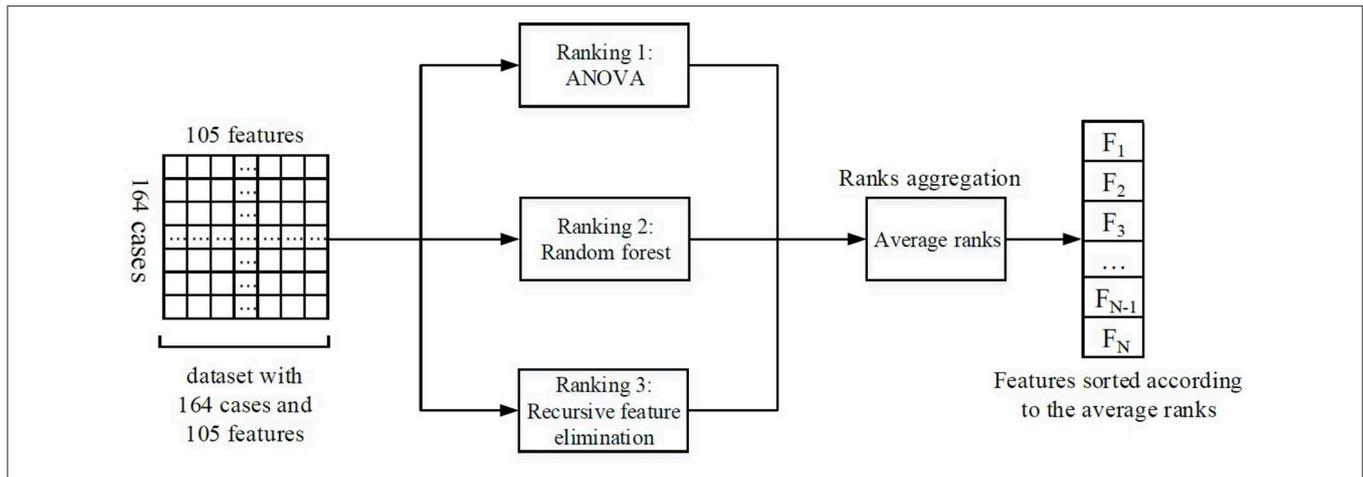
FIGURE 2 | Illustration of the feature selection process. Firstly, three feature selection methods ranked the features individually. Then, the ranks from different methods were averaged. Finally, the features were sorted according to the average ranks, and the most important N features were selected for the subsequent analysis (the number N could be determined by RFE).
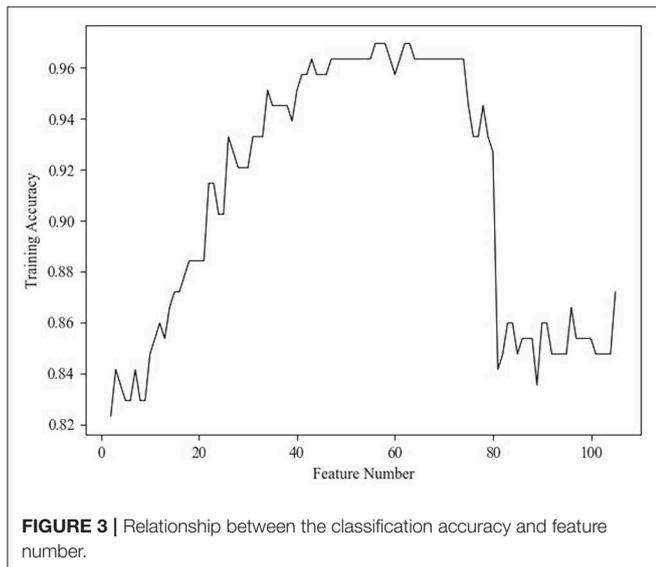

FIGURE 3 | Relationship between the classification accuracy and feature number.

Finally, the features were sorted according to the average feature ranks from the above three rankers and the top 40 features were selected for the subsequent classifiers. To illustrate the feature selection results, we would present the top five features selected by each feature selection method in **Table 2**. All these three methods were implemented on a python package named scikit-learn (Pedregosa et al., 2013). When ranking the features by random forest and RFE, we used all the 164 samples as training set because the validation/test accuracy did not matter when we concentrated on the features' importance in this procedure.

Classification

SVM is one of the most popular machine learning methods and has been used in PD diagnosis (Prashanth et al., 2016; Amoroso et al., 2018; Castillo-Barnes et al., 2018). It tries to find out the

optimal hyperplane that minimizes the classification error and maximizes the geometric margin on the training set, which leads to high generalization ability on the new cases (Burges, 1998). In practice, radial basis function (RBF) is usually used as the kernel function to non-linearly map the features to a higher dimensionality and improve the performance of SVM (Han et al., 2012). The RBF kernel function is:

$$K(x_i, x_j) = \exp(-\gamma \left\| x_i - x_j \right\|^2) \quad (1)$$

where x_i and x_j are the features of samples; γ is a constant parameter.

And the object function of the SVM is:

$$\begin{aligned} &\min_{\omega,b,\xi} \tfrac{1}{2} \left\| \omega \right\|^2 + C \sum_{i=1}^{N} \xi_i \\ &\text{s.t. } y_i(\omega x_i + b) \geq 1 - \xi_i, \ i = 1, 2, 3, \ldots, N \\ &\xi_i \geq 0, i = 1, 2, 3, \ldots, N \end{aligned} \quad (2)$$

where y_i is the label of the ith sample, and C is a constant parameter.

Then, the function can be solved by a Lagrangian function:

$$\begin{aligned} \min_{\omega,b,\xi} \ L(\omega, b, \xi, \alpha, \mu) = &-\frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j \exp(-\gamma ||x_i - x_j||^2) \\ &+ \sum_{i=1}^{N} \alpha_i \end{aligned} \quad (3)$$
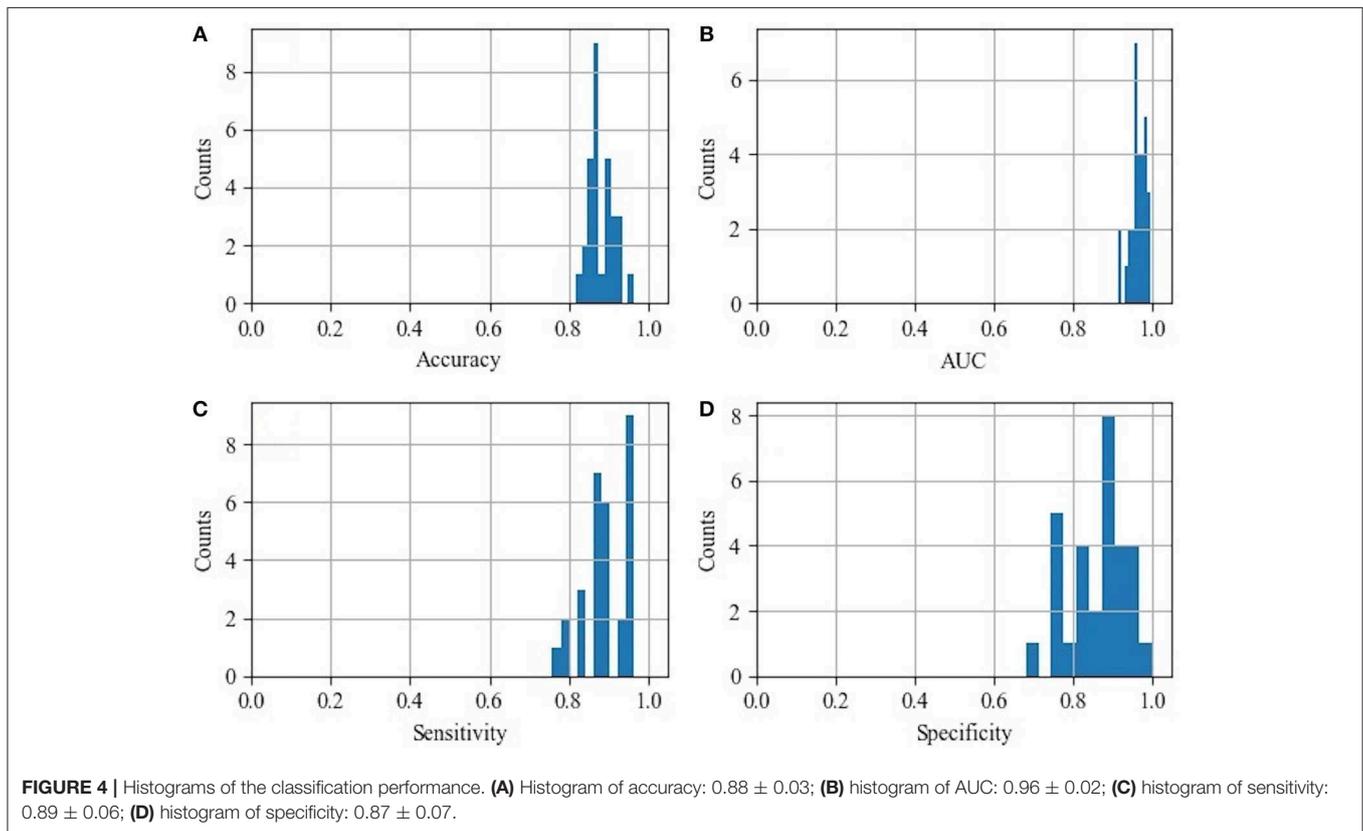
At last, the obtained SVM function is:

$$f(x) = \omega x + b \quad (4)$$

There are two parameters (C, γ) that can be adjusted according to a specific task and dataset. The parameter C is a penalty

TABLE 2 | Results of the ROC curve analyses of the top five features picked by each feature selection method.

	ANOVA		Random Forest		RFE	
	Features name	AUC	Features name	AUC	Features name	AUC
Top1	Minimum	0.83	GryLvNonUni	0.80	DepdEntrp	0.80
Top2	DepdNonUni	0.84	Correlation	0.78	SumEntrp	0.51
Top3	DepdVar	0.81	10 Percentile	0.75	RunEntrp	0.76
Top4	InfMCor1	0.81	Minimum	0.83	ZoneEntrp	0.62
Top5	GryLvNonUni	0.80	RunLthNonUni	0.78	HGLRunEmphs	0.50

FIGURE 4 | Histograms of the classification performance. **(A)** Histogram of accuracy: 0.88 ± 0.03; **(B)** histogram of AUC: 0.96 ± 0.02; **(C)** histogram of sensitivity: 0.89 ± 0.06; **(D)** histogram of specificity: 0.87 ± 0.07.

parameter, which determines the error tolerance when we train an SVM. The parameter γ adjusts the effect of RBF.

The order of the magnitudes of selected features was different; therefore, prior to being fed to the SVM classifier, the features were standardized to zero mean and unit variance. In addition, the standardized features were transformed into an orthogonal space to make the features more discriminating by principal components analysis (PCA) (Groth et al., 2013). Then, these features were input into the SVM classifier. the 3-fold cross-validation with 10 repetitions was used to reduce the influence of dataset partition during the procedure of training and testing. In other words, in each round of the training and testing, we randomly separated the 164 samples into 3-folds. Then, 2-folds (109 samples) were used as training set and the left fold (55 samples) was used as testing set. In our study, the kernel function of SVM was set as RBF conventionally.

Parameters C and γ were set as 30 and 0.001 by the grid search method.

The performance of the SVM classification was evaluated using the area under the receiver operating characteristic (ROC) curve, accuracy, specificity, and sensitivity.

Statistical Analysis

Two-tailed unpaired t-test was used to compare selected features between HCs and IPDs, and Bonferroni correction was used to correct the error of multiple comparisons. The adjusted p-value was set at $0.05/n$, where $n =$ time of comparison. ROC was used to further test the selected features in differentiating IPD from HC *via* the area under ROC (AUC).

Two-tailed Pearson linear correlation was also utilized to study the correlations between the selected features and the

TABLE 3 | Classification performance of 3-fold cross-validation in the first round.

Index	AUC	Accuracy	Sensitivity	Specificity
Threefold cross-validation 1	0.95	0.85	0.93	0.77
Threefold cross-validation 2	0.97	0.93	0.93	0.92
Threefold cross-validation 3	0.97	0.87	0.86	0.88
Mean ± SD	0.96 ± 0.01	0.88 ± 0.04	0.91 ± 0.04	0.86 ± 0.08

UPDRS-III scores of the IPDs. All the above statistical analyses were processed by SPSS Statistics version 25 for mac.

RESULTS

Classification Performance

The histograms of the results of the 3-fold cross-validation with 10 repetitions are shown in **Figure 4**, with the accuracy ranging from 0.80 to 0.95 (centralized at 0.85). The AUC was stable and approximately centralized at 0.95, while the sensitivity and specificity were relatively unstable, ranging from 0.8 to 1.0. Statistically, the AUC, accuracy, sensitivity, and specificity were 0.96 ± 0.02, 0.88 ± 0.03, 0.89 ± 0.06, and 0.87 ± 0.07 (mean ± standard deviation), respectively. To make the performance illustration more concrete and clearer, the cross-validation result in the first round was chosen as a representative example (**Table 3**; **Figure 5**). According to the classification performance, the selected features offered informative contents to the SVM and could produce the ROC with average AUC beyond 0.95. As a contrast, the STS was visualized by a radiologist in 67.52% (52/77) of the cases of HC and 9.20% (8/87) of the cases of IPD, respectively. The STS in differentiating HC from IPD was with an accuracy of 79.88%, a sensitivity of 67.53%, and a specificity of 90.80%, which was overall lower than that of the classification.

Top Five and Representative Features Analysis

ROC curve analysis of the top five picked by each feature selection is summarized in **Table 2**. The AUC of the top five features selected by ANOVA, random forest, and RFE varied from 0.80 to 0.84, 0.75 to 0.83, and 0.50 to 0.80, respectively.

There were 29 out of the 40 selected features showing significant difference (p-value <0.05/40) between IPD and HC groups, and five features were selected as representative in **Table 4** to show the details (all t-test results can be found in **supplementary S2**). The 10th percentile (0.023 ± 0.007 vs. 0.015 ± 0.009, $p < 0.0001$) and median (0.076 ± 0.016 vs. 0.066 ± 0.015, $p < 0.0001$) in IPD patients were higher than those in HCs; while, GLRLM-Long Run Low Gray Level Emphasis (0.420 ± 0.133 vs. 0.546 ± 0.312, $p = 0.001$), GLSZM-Gray Level Non-Uniformity (5.769 ± 2.442, 7.583 ± 2.707, $p < 0.0001$), and volume (519.514 ± 128.743 vs. 629.073 ± 129.558, $p < 0.0001$)

FIGURE 5 | Receiver operating characteristics curve obtained from the 3-fold cross-validation in the first round.

TABLE 4 | Unpaired t-test performances of the five representative features.

Features Name	IPD	HC	P-value	AUC
10 th percentile	0.023 ± 0.007	0.015 ± 0.009	1.49E−9	0.75
Median	0.076 ± 0.016	0.066 ± 0.015	4.10E−5	0.68
Volume	519.514 ± 128.743	629.073 ± 129.558	2.46E−7	0.73
LRunLGREmphs	0.420 ± 0.133	0.546 ± 0.312	1.00E−3	0.64
GryLvNonUniS	5.769 ± 2.442	7.583 ± 2.707	1.40E−5	0.71

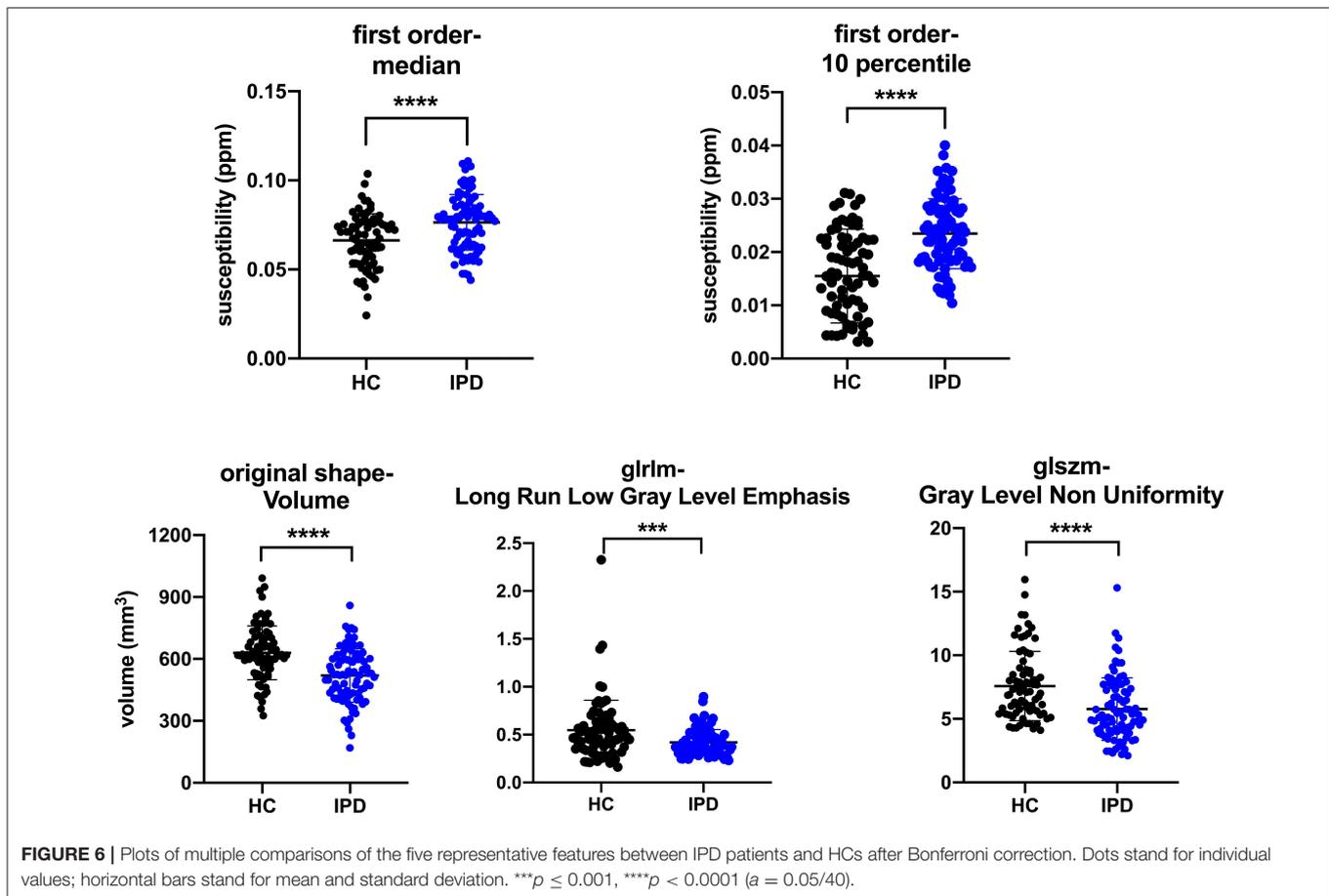*LRunLGREmphs, GLRLM-Long Run Low Gray Level Emphasis; GryLvNonUniS, GLSZM-Gray Level Non-Uniformity; the units of 10th Percentile and Median were parts per million (ppm); the unit of volume was mm^3; there were no units of the remaining two features.*

in IPD patients were lower than those in HCs. The AUC of these five features varied from 0.64 to 0.75 (**Table 4**; **Figure 6**).

The correlation between the five features and the UPDRS-III score of IPD was as follows: 10th percentile of the first order was positively correlated with UPDRS-III score ($r = 0.35$, $p = 0.001$); median of the first order, GLSZM-Gray Level Non-Uniformity, GLRLM-Long Run Low Gray Level Emphasis, and volume were not correlated with UPDRS-III score ($r = 0.15$, -0.02, -0.17, and -0.05 and $p = 0.19$, 0.87, 0.12, and 0.65, respectively).

DISCUSSION

Since iron accumulation in the SN plays a vital role in the progression of PD, visualization and quantification of increased iron are thought to have the potential to assist in the diagnosis and evaluation on PD (He et al., 2015; Noh et al., 2015; Reiter et al., 2015). Howerver, iron deposition is spatially heterogeneous and the nigrosome territory may individually vary (Schmidt et al., 2017). Single iron-sensitive nigrosome imaging for visualization or simple utilization of the mean of susceptibility might be subjective, inssufficient, and inaccurate.

FIGURE 6 | Plots of multiple comparisons of the five representative features between IPD patients and HCs after Bonferroni correction. Dots stand for individual values; horizontal bars stand for mean and standard deviation. ***$p \leq 0.001$, ****$p < 0.0001$ ($a = 0.05/40$).

Radiomics based on QSM could offer the means to overcome these deficiencies. Our preliminary results verified that radiomic features of the nigrosome-1 region of the SN were different between IPD patients and HCs. The performance of the SVM approach based on these radiomic features was better than that of the radiologist in our study, especially on accuracy and sensitivity, and was similar to that of a meta-analysis based on visualizing the STS at 3.0 T (Mahlknecht et al., 2017). The reason on one hand might be the limitation of radiologists' visual discernment on the sign especially between the early stage IPD and the HCs, and on the other hand, more HCs in our group showed loss of STS unilaterally or bilaterally as a result of individual nigrosome-1 variants or potential preclinical PD, which might lead to a similar performance to the previous visual assessment investigations.

As the first-order features describe the distribution of the voxel intensities within a VOI, they might have the potential to reflect the iron deposition spatially. Our study found that the representative features of the 10th percentile and median in the nigrosome-1 region of the SN in IPD patients were bigger than those of the HCs, and the 10th percentile positively correlated with the UPDRS-III, suggesting that the 10th percentile and median of intra-voxel susceptibility (iron content) might better differentiate IPD from HC and the 10th percentile might be a

biomarker for the progression of the IPD. The 10th percentile and median represent the first 10 and 50% proportion of voxels with positive order of susceptibility, respetively. The voxels with lower susceptibility can partially seperate the overlap of voxels between IPD and HC (**Figure 6**, first row) and might outperform the mean value, which is similar to the findings of a histogram analytical study that using the proportion of voxels with susceptibility lower than 70 ppb as the threshold could better differentiate early IPD from HC compared to the other thresholds including mean value (Kim et al., 2018). The volume, one of the shape-based features, was found smaller in IPD than that in HC, which is consistent with previous studies (Sasaki et al., 2006; Minati et al., 2007; Menke et al., 2009; Ziegler et al., 2013), suggesting that the iron deposition increases and dopaminergic neurons lose in IPD (Duguid et al., 1986). The second-order features describe the relationship of all voxel pairs within a segamented VOI. GLRLM-Long Run Low Gray Level Emphasis and GLSZM-Gray Level Non-Uniformity describe the joint distribution of long run lengths with lower gray-level values and the similarity of the gray-level intensity values, respectively, which were both found lower in the IPD patients compared to those in HCs in our study. This reflected the fact that iron accumulation of the nigrosome-1-containing part of SN was spatially more uniform in IPD, which was consistent with the loss of STS in IPD. Therefore,

radiomic features could be an objective means to assess iron accumulation and neuromelanin loss in SN spatially and could be a promising time-saving tool in assisting in the diagnosis of IPD.

Furthermore, apart from analyzing the selected radiomic features in this study, we also trained SVM to classify the subjects as IPD patients or HCs automatically. Recently, several studies have proved that it is effective to diagnose PD using SVM and different features (Prashanth et al., 2016; Amoroso et al., 2018; Castillo-Barnes et al., 2018) with an accuracy ranging from 0.70 to 0.96. However, few of them trained the SVM using the radiomic features as input. In our study, we found that the radiomic featues also provided discraminative information to the SVM classifier and could reach an accuracy of 0.88. The SVM classifier could ultilize high-dimensional features automatically, but its results intrinsically fluctuate according to different training datasets and testing datesets (**Figure 4**).

There are several limitations to this study. First, only the nigrosome-1 territory was segmented, which might bias the results. It is known that the nigrosome-1 is the first to be involved in the progression of PD (Sung et al., 2018). Since most of our IPD cases were at the the early stages (68 out of 77 cases were at Hoehn and Yahr stage 1–2), we expect that less bias was introduced. Second, instead of defining some specific features that might better characterize the nigrosome-1 pathological changes and subsequently well-differentiate the IPD from HCs, only 105 common radiomic features were extracted in the study, which might be insufficient. Third, we only used SVM as the classifier, rather than comparing or assembling the performances of different machine learning methods. In the future, we can improve the accuracy by exploring novel features and classifiers.

In conclusion, radiomic features of the SN based on QSM could be useful in the diagnosis of IPD and could serve as a surrogate marker for the qualitiative (visualization) and conventional quantitative evalaution of STS or nigrosome-1.

ETHICS STATEMENT

AUTHOR CONTRIBUTIONS

ZC and FY designed the study. NH, HX, ZJ, RT, and YL collected the MRI data. ZC and JZ segmented the QSM data. JZ, YW, and DQ analyzed the data and interpreted the radiomic features. ZC and JZ wrote the manuscript draft. EH revised the manuscript.

FUNDING

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnagi.2019.00167/full#supplementary-material

REFERENCES

Abeel, T., Helleputte, T., Van de Peer, Y., Dupont, P., Saeys, Y. (2010). Robust biomarker identification for cancer diagnosis with ensemble feature selection methods. *Bioinformatics* 26, 392–398. doi: 10.1093/bioinformatics/btp630

Amoroso, N., La Rocca, M., Monaco, A., Bellotti, R., and Tangaro, S. (2018). Complex networks reveal early MRI markers of Parkinson's disease. *Med. Image Anal.* 48, 12–24. doi: 10.1016/j.media.2018.05.004

Breiman, L. (2001). Random forests, machine learning 45. *J. Clin. Microbiol.* 2, 199–228. doi: 10.1023/A:1017934522171

Burges, C. J. C. (1998). A tutorial on support vector machines for pattern recognition. *Data Min. Knowl. Discov.* 2, 121–167. doi: 10.1023/A:1009715923555

Cameron, A., Khalvati, F., Haider, M. A., and Wong, A. (2016). MAPS: a quantitative radiomics approach for prostate cancer detection. *IEEE Trans. Biomed. Eng.* 63, 1145–1156. doi: 10.1109/TBME.2015.2485779

Castillo-Barnes, D., Ramirez, J., Segovia, F., Martinez-Murcia, F. J., Salas-Gonzalez, D., and Gorriz, J. M. (2018). Robust ensemble classification methodology for I123-Ioflupane SPECT images and multiple heterogeneous biomarkers in the diagnosis of Parkinson's disease. *Front. Neuroinform.* 12:53. doi: 10.3389/fninf.2018.00053

Damier, P., Hirsch, E. C., Agid, Y., and Graybiel, A. M. (1999). The substantia nigra of the human brain. II. Patterns of loss of dopamine-containing neurons in Parkinson's disease. *Brain* 122 (Pt 8), 1437–1448. doi: 10.1093/brain/122.8.1437

Du, G., Lewis, M. M., Styner, M., Shaffer, M. L., Sen, S., Yang, Q. X., et al. (2011). Combined R2* and diffusion tensor imaging changes in the substantia nigra in Parkinson's disease. *Mov. Disord.* 26, 1627–1632. doi: 10.1002/mds.23643

Du, G., Liu, T., Lewis, M. M., Kong, L., Wang, Y., Connor, J., et al. (2016). Quantitative susceptibility mapping of the midbrain in Parkinson's disease. *Mov. Disord.* 31, 317–324. doi: 10.1002/mds.26417

Duguid, J. R., De La Paz, R., and DeGroot, J. (1986). Magnetic resonance imaging of the midbrain in Parkinson's disease. *Ann. Neurol.* 20, 744–747. doi: 10.1002/ana.410200618

Feng, Q., Chen, Y., Liao, Z., Jiang, H., Mao, D., Wang, M., et al. (2018). Corpus callosum radiomics-based classification model in Alzheimer's disease: a case–control study. *Front. Neurol.* 9:618. doi: 10.3389/fneur.2018.00618

Groth, D., Hartmann, S., Klie, S., and Selbig, J. (2013). Principal components analysis. *Methods Mol. Biol.* 930, 527–547. doi: 10.1007/978-1-62703-059-5_22

Han, S. J., Cao, Q. B., and Han, M. (2012). "Parameter selection in SVM with RBF kernel function," in *2012 World Automation Congress (Wac)* (Puerto Vallarta).

He, N., Ling, H., Ding, B., Huang, J., Zhang, Y., Zhang, Z., et al. (2015). Region-specific disturbed iron distribution in early idiopathic Parkinson's

disease measured by quantitative susceptibility mapping. *Hum. Brain Mapp.* 36, 4407–4420. doi: 10.1002/hbm.22928

Hoque, N., Singh, M., and Bhattacharyya, D. K. (2018). EFS-MI: an ensemble feature selection method for classification. *Complex Intell. Syst.* 4, 105–118. doi: 10.1007/s40747-017-0060-x

Hughes, A. J., Daniel, S. E., Kilford, L., and Lees, A. J. (1992). Accuracy of clinical diagnosis of idiopathic Parkinson's disease: a clinico-pathological study of 100 cases. *J. Neurol. Neurosurg. Psychiatr.* 55, 181–184. doi: 10.1136/jnnp.55.3.181

Jjm, V. G., Fedorov, A., Parmar, C., Hosny, A., Aucoin, N., Narayan, V., et al. (2017). Computational radiomics system to decode the radiographic phenotype. *Cancer Res.* 77:e104– e107. doi: 10.1158/0008-5472.CAN-17-0339

Jong, K., Jérémie, M., Cornuéjols, A., Marchiori, E., and Sebag, M. (2004). "Ensemble Feature Ranking," in *European Conference on Principles and Practice of Knowledge Discovery in Databases* (Pisa). doi: 10.1007/978-3-540-30116-5_26

Kim, E. Y., Sung, Y. H., Shin, H. G., Noh, Y., Nam, Y., and Lee, J. (2018). Diagnosis of early-stage idiopathic Parkinson's disease using high-resolution quantitative susceptibility mapping combined with histogram analysis in the substantia nigra at 3 T. *J. Clin. Neurol.* 14, 90–97. doi: 10.3988/jcn.2018.14.1.90

Langkammer, C., Schweser, F., Krebs, N., Deistung, A., Goessler, W., Scheurer, E., et al. (2012). Quantitative susceptibility mapping (QSM) as a means to measure brain iron? A post mortem validation study. *Neuroimage* 62, 1593–1599. doi: 10.1016/j.neuroimage.2012.05.049

Lehericy, S., Bardinet, E., Poupon, C., Vidailhet, M., and Francois, C. (2014). 7 Tesla magnetic resonance imaging: a closer look at substantia nigra anatomy in Parkinson's disease. *Mov. Disord.* 29, 1574–1581. doi: 10.1002/mds.26043

Li, J., Chang, S., Liu, T., Wang, Q., Cui, D., Chen, X., et al. (2012). Reducing the object orientation dependence of susceptibility effects in gradient echo MRI through quantitative susceptibility mapping. *Magn. Reson. Med.* 68, 1563–1569. doi: 10.1002/mrm.24135

Li, W., Wang, N., Yu, F., Han, H., Cao, W., Romero, R., et al. (2015). A method for estimating and removing streaking artifacts in quantitative susceptibility mapping. *Neuroimage* 108, 111–122. doi: 10.1016/j.neuroimage.2014.12.043

Li, W., Wu, B., and Liu, C. (2011). Quantitative susceptibility mapping of human brain reflects spatial variation in tissue composition. *Neuroimage* 55, 1645–1656. doi: 10.1016/j.neuroimage.2010.11.088

Mahlknecht, P., Krismer, F., Poewe, W., and Seppi, K. (2017). Meta-analysis of dorsolateral nigral hyperintensity on magnetic resonance imaging as a marker for Parkinson's disease. *Mov. Disord.* 32, 619–623. doi: 10.1002/mds.26932

Martin, W. R., Wieler, M., and Gee, M. (2008). Midbrain iron content in early Parkinson disease: a potential biomarker of disease status. *Neurology* 70 (16 Pt 2), 1411–1417. doi: 10.1212/01.wnl.0000286384.31050.b5

Massey, L. A., Miranda, M. A., Al-Helli, O., Parkes, H. G., Thornton, J. S., So, P. W., et al. (2017). 9.4 T MR microscopy of the substantia nigra with pathological validation in controls and disease. *Neuroimage Clin.* 13, 154–163. doi: 10.1016/j.nicl.2016.11.015

Menke, R. A., Scholz, J., Miller, K. L., Deoni, S., Jbabdi, S., Matthews, P. M., et al. (2009). MRI characteristics of the substantia nigra in Parkinson's disease: a combined quantitative T1 and DTI study. *Neuroimage* 47, 435–441. doi: 10.1016/j.neuroimage.2009.05.017

Minati, L., Grisoli, M., Carella, F., De Simone, T., Bruzzone, M. G., and Savoiardo, M. (2007). Imaging degeneration of the substantia nigra in Parkinson disease with inversion-recovery MR imaging. *AJNR Am. J. Neuroradiol.* 28, 309–313.

Murakami, Y., Kakeda, S., Watanabe, K., Ueda, I., Ogasawara, A., Moriya, J., et al. (2015). Usefulness of quantitative susceptibility mapping for the diagnosis of Parkinson disease. *AJNR Am. J. Neuroradiol.* 36, 1102–1108. doi: 10.3174/ajnr.A4260

Nie, K., Shi, L., Chen, Q., Hu, X., Jabbour, S. K., Yue, N., et al. (2016). Rectal cancer: Assessment of neoadjuvant chemoradiation outcome based on radiomics of multiparametric MRI. *Clin. Cancer Res.* 22, 5256–5264. doi: 10.1158/1078-0432.CCR-15-2997

Noh, Y., Sung, Y. H., Lee, J., and Kim, E. Y. (2015). Nigrosome 1 detection at 3T MRI for the diagnosis of early-stage idiopathic Parkinson disease: assessment of diagnostic accuracy and agreement on imaging asymmetry and clinical laterality. *AJNR Am. J. Neuroradiol.* 36, 2010–2016. doi: 10.3174/ajnr.A4412

Pedregosa, F., Gaël, V., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2013). Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* 12, 2825–2830.

Prashanth, R., Dutta Roy, S., Mandal, P. K., and Ghosh, S. (2016). High-accuracy detection of early Parkinson's disease through multimodal features and machine learning. *Int. J. Med. Inform.* 90, 13–21. doi: 10.1016/j.ijmedinf.2016.03.001

Reiter, E., Mueller, C., Pinter, B., Krismer, F., Scherfler, C., Esterhammer, R., et al. (2015). Dorsolateral nigral hyperintensity on 3.0T susceptibility-weighted imaging in neurodegenerative Parkinsonism. *Mov. Disord.* 30, 1068–1076. doi: 10.1002/mds.26171

Sasaki, M., Shibata, E., Tohyama, K., Takahashi, J., Otsuka, K., Tsuchiya, K., et al. (2006). Neuromelanin magnetic resonance imaging of locus ceruleus and substantia nigra in Parkinson's disease. *Neuroreport* 17, 1215–1218. doi: 10.1097/01.wnr.0000227984.84927.a7

Schmidt, M. A., Engelhorn, T., Marxreiter, F., Winkler, J., Lang, S., Kloska, S., et al. (2017). Ultra high-field SWI of the substantia nigra at 7T: reliability and consistency of the swallow-tail sign. *BMC Neurol.* 17:194. doi: 10.1186/s12883-017-0975-2

Schwarz, S. T., Afzal, M., Morgan, P. S., Bajaj, N., Gowland, P. A., and Auer, D. P. (2014). 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* 9:e93814. doi: 10.1371/journal.pone.0093814

Schweser, F., Deistung, A., Lehr, B. W., and 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. doi: 10.1016/j.neuroimage.2010.10.070

Stezin, A., Naduthota, R. M., Botta, R., Varadharajan, S., Lenka, A., Saini, J., et al. (2018). Clinical utility of visualisation of nigrosome-1 in patients with Parkinson's disease. *Eur. Radiol.* 28, 718–726. doi: 10.1007/s00330-017-4950-5

Sung, Y. H., Lee, J., Nam, Y., Shin, H. G., Noh, Y., Shin, D. H., et al. (2018). Differential involvement of nigral subregions in idiopathic parkinson's disease. *Hum. Brain Mapp.* 39, 542–553. doi: 10.1002/hbm.23863

Wang, Y., and Liu, T. (2015). Quantitative susceptibility mapping (QSM): decoding MRI data for a tissue magnetic biomarker. *Magn. Reson. Med.* 73, 82–101. doi: 10.1002/mrm.25358

Zhang, B., Tian, J., Dong, D., Gu, D., Dong, Y., Zhang, L., et al. (2017). Radiomics features of multiparametric MRI as novel prognostic factors in advanced nasopharyngeal carcinoma. *Clin. Cancer Res.* 23, 4259–4269. doi: 10.1158/1078-0432.CCR-16-2910

Ziegler, D. A., Wonderlick, J. S., Ashourian, P., Hansen, L. A., Young, J. C., Murphy, A. J., et al. (2013). Substantia nigra volume loss before basal forebrain degeneration in early Parkinson disease. *JAMA Neurol.* 70, 241–247. doi: 10.1001/jamaneurol.2013.597