



Effect of Deep Learning-Based Artificial Intelligence on Radiologists' Performance in Identifying Nigrosome 1 Abnormalities on Susceptibility Map-Weighted Imaging

Jiyeon Park¹, Chae Young Lim¹, So Yeon Won¹, Han Kyu Na², Phil Hyu Lee², Sun-Young Baek^{1,3}, Yun Hwa Roh¹, Minjung Seong¹, Yongsik Sim¹, Eung Yeop Kim¹, Sung Tae Kim¹, Beomseok Sohn¹

¹Department of Radiology and Center for Imaging Science, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, Republic of Korea

²Department of Neurology, Severance Hospital, Yonsei University College of Medicine, Seoul, Republic of Korea

³Research Institute for Future Medicine, Samsung Medical Center, Seoul, Republic of Korea

Objective: To evaluate the effect of deep learning (DL)-based artificial intelligence (AI) software on the diagnostic performance of radiologists with different experience levels in detecting nigrosome 1 (N1) abnormalities on susceptibility map-weighted imaging (SMwI).

Materials and Methods: This retrospective diagnostic case-control study analyzed 139 SMwI scans of 59 patients with Parkinson's disease (PD) and 80 healthy participants. Participants were imaged using 3T MRI, and AI-generated assessments for N1 abnormalities were obtained using an AI model (version 1.0.1.0; Heuron Corporation, Seoul, Korea), which utilized YOLOX-based object detection and SparseInst segmentation models. Four radiologists (two experienced neuroradiologists and two less experienced residents) evaluated N1 abnormalities with and without AI in a crossover study design. Diagnostic performance metrics, inter-reader agreements, and reader responses to AI-generated assessments were evaluated.

Results: Use of AI significantly improved diagnostic performance compared with interpretation without it across three readers, with significant increases in specificity (0.86 vs. 0.94, $P = 0.004$; 0.91 vs. 0.97, $P = 0.024$; and 0.90 vs. 0.97, $P = 0.012$). Inter-reader agreement also improved with AI, as Fleiss's kappa increased from 0.73 (95% confidence interval [CI]: 0.61–0.84) to 0.87 (95% CI: 0.76–0.99). The net reclassification index (NRI) demonstrated significant improvement in three of the four readers. When grouped by experience level, less experienced readers showed greater improvement (NRI = 12.8%, 95% CI: 0.067–0.190) than experienced readers (NRI = 0.8%, 95% CI: -0.037–0.051). In the less experienced group, reader-AI disagreement was significantly higher in the PD group than in the normal group (8.1% vs. 3.8%, $P = 0.029$).

Conclusion: DL-based AI enhances the diagnostic performance in detecting N1 abnormalities on SMwI, particularly benefiting less experienced radiologists. These findings underscore the potential for improving diagnostic workflows for PD.

Keywords: Parkinson disease; Magnetic resonance imaging; Substantia nigra; Artificial intelligence

INTRODUCTION

Parkinson's disease (PD) is a progressive neurodegenerative disorder characterized by resting tremor, rigidity, and reduced voluntary movement. It affects approximately 1.0% of individuals aged ≥ 70 years in Asia and 2.3% of

individuals in Western countries [1]. While the clinical diagnostic accuracy of the Movement Disorder Society (MDS) Criteria reaches 85% [2-5], differentiating PD from atypical Parkinsonism remains challenging. Diagnosis often requires prolonged observation of medication response or symptom progression. Although recommended, dopamine transporter

Received: December 4, 2024 **Revised:** February 16, 2025 **Accepted:** May 20, 2025

Corresponding author: Beomseok Sohn, MD, PhD, Department of Radiology, Samsung Medical Center, Sungkyunkwan University School of Medicine, 81 Irwon-ro, Gangnam-gu, Seoul 06351, Republic of Korea

• E-mail: beomseoksohn@gmail.com

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<https://creativecommons.org/licenses/by-nc/4.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

(DaT) PET is costly and not widely accessible [2], driving the need for alternative diagnostic approaches.

PD pathology primarily affects the substantia nigra pars compacta (SNpc), particularly nigrosome 1 (N1), where dopaminergic neuron loss and iron overload occur [6,7]. Advanced 3T MRI techniques, including susceptibility-weighted imaging, gradient recall echo (GRE), and quantitative susceptibility mapping (QSM), allow N1 visualization and differentiation between PD and normal conditions [7-13]. Recently, multi-echo susceptibility map-weighted imaging (SMwI), which utilizes multi-echo GRE complex data to provide high susceptibility contrast, has enhanced N1 visibility by improving the contrast-to-noise ratio and spatial resolution [14,15].

However, detecting N1 abnormalities in SMwI remains challenging, as it requires precise localization and intensity assessment of this small structure [9,15-17]. Experienced radiologists often struggle to allocate sufficient time to review N1 because of the increasing workload resulting from the rising imaging demands [18]. Therefore, deep learning (DL)-based artificial intelligence (AI) technologies may facilitate faster and more accurate diagnoses. AI has already shown potential in detecting aneurysms on brain MR angiography [19] and cartilage lesions on knee MRI [20].

This study aimed to evaluate the diagnostic performance of commercially available DL-based AI [21] in detecting N1 abnormalities. Specifically, we assessed how the AI-

generated results affected the interpretations of radiologists with different levels of expertise.

MATERIALS AND METHODS

Participants

This diagnostic case-control study was conducted at a single tertiary center and was approved by our Institutional Review Board (IRB No. 2024-01-112). This study was a retrospective analysis of a cohort from a prospective study (NCT05513794). All participants provided informed consent. Those who withdrew consent during the study or were lost to follow-up were excluded. Initially, 142 participants were enrolled, including 62 patients with PD and 80 clinically normal (normal control, NC) patients; three patients with PD withdrew consent later. This study included 59 patients with clinically diagnosed PD and 80 intentionally enrolled NCs. Participants were conveniently sampled during the study period, with images acquired between November 7, 2022, and April 15, 2023, and then retrospectively analyzed (Fig. 1).

The inclusion criteria for patients with PD were as follows: 1) aged ≥ 19 years, 2) exhibited Parkinson's symptoms (e.g., tremor, rigidity, bradykinesia, gait disturbance) and scheduled for MRI, 3) significant reduction in dopamine uptake in the bilateral substantia nigra on DaT PET, 4) able to read the consent form and participate in a Q&A session,

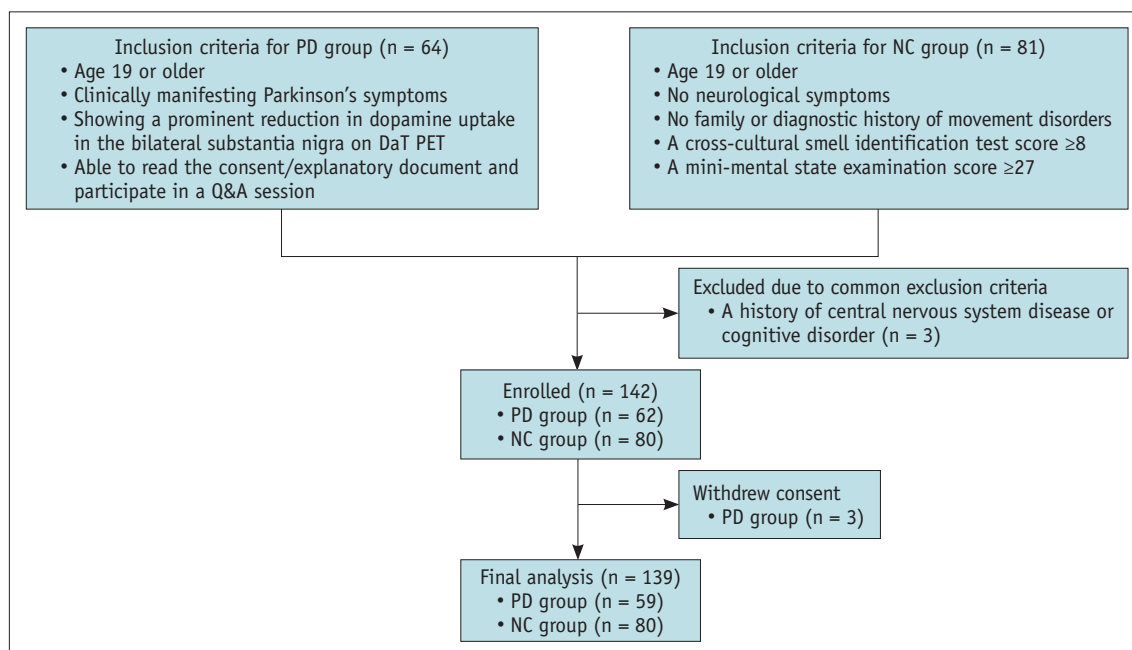


Fig. 1. Patient selection flow diagram. PD = Parkinson's disease, NC = normal control, DaT = dopamine transporter

and 5) provided written informed consent, or consent obtained from a legal guardian if cognitively impaired.

The inclusion criteria for normal participants were as follows: 1) aged ≥ 19 years, 2) no neurological symptoms, 3) no family history or prior diagnosis of movement disorders, 4) cross-cultural smell identification test score ≥ 8 , 5) mini-mental state examination score ≥ 27 , and 6) provided informed consent after a detailed explanation. No more than 15 participants were included in each group, to ensure age diversity.

The exclusion criteria for both groups were as follows: 1) history of central nervous system disease or cognitive disorder, 2) claustrophobia or mental illness, 3) metallic implants, 4) females unwilling to use contraception, 5) pregnant or lactating female, and 6) any condition deemed unsuitable by the investigator.

SMwI Data Acquisition

The imaging range was set to sufficiently include the midbrain to capture the N1 region, and 1 mm thin-section non-contrast brain MRI images were obtained. The scan plane was an oblique coronal plane aligned parallel to the plane from the posterior commissure (PC) to the anterosuperior border of the pons, approximately perpendicular to the midbrain. Compared to SMwI obtained using the conventional anterior commissure-posterior commissure (AC-PC) line commonly used in brain MRI, SMwI obtained in the oblique coronal plane has been shown to result in no significant difference in Parkinson's diagnostic accuracy [22,23]. Imaging was performed using a GRE sequence with three echoes. Subsequently, the QSM calculated using the iterative least-squares method was used

to create masks to reconstruct the SMwIs. The settings for the 3T scanner (Ingenia CX, Philips Healthcare, Best, the Netherlands) were as follows: repetition time = 48 ms; echo time = 14, 27, and 40 ms; echo spacing = 13 ms; flip angle = 20° ; slice thickness = 1 mm; matrix size = 384×384 ; and field of view = 192×192 .

AI Software

AI-generated interpretations were obtained using the Heuron IPD software (version 1.0.1.0; Heuron Corporation, Seoul, Korea) (Fig. 2). This software is based on the same core algorithm as described in a recent study by Suh et al. [24], which used an updated version trained with additional datasets. No data from the present study were used for training or adjustment. The system employs two AI models that operate independently to interpret SMwI images: 1) the YOLOX object detection model [25], which selects and analyzes five input slices covering the left or right substantia nigra and estimates the probability values for N1 abnormalities, and 2) the SparseInst segmentation model [26], which calculates the volumetric values of N1. For each slice, if the probability value exceeded a fixed threshold of 0.2, which was determined through internal validation to balance sensitivity and specificity while minimizing false negatives, the corresponding N1 volume was used as a weight. The system then summed the weighted probabilities across the five slices for both the 'normal' and 'abnormal' classes and assigned the class with the highest total as the final prediction for each N1 region (left and right) (Supplementary Fig. 1). This threshold was fixed and consistently applied across all datasets. The technical details and rationale for the AI model design are

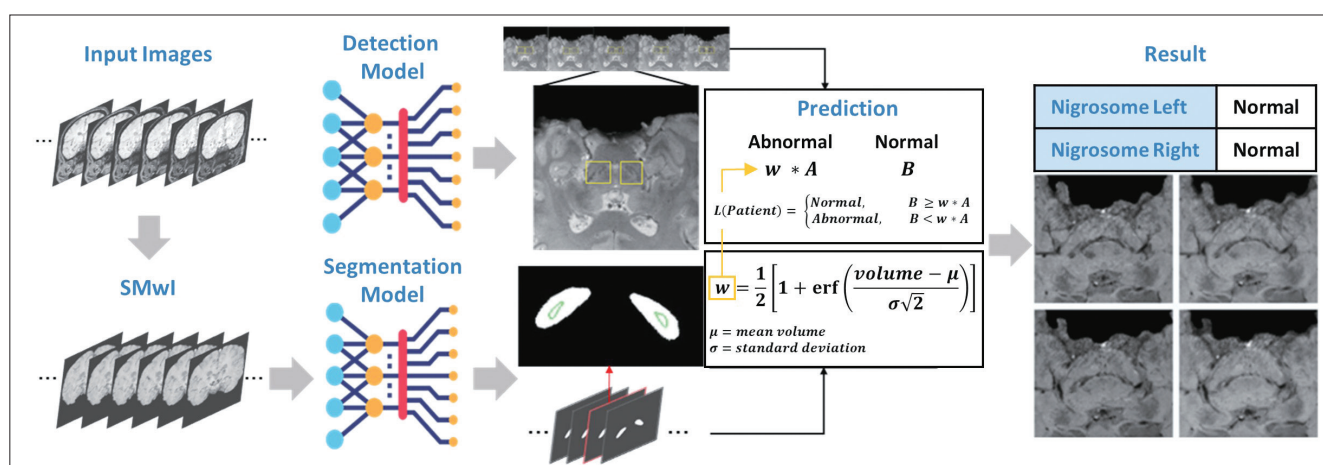


Fig. 2. Automatic detection and classification of nigrosome 1 abnormalities using deep learning models. Output from the deep learning model shows the analysis results for detecting abnormalities. SMwI = susceptibility map-weighted imaging

described in the Supplement. The software does not provide interpretive assistance tools, such as abnormality scores or saliency maps.

Image Analysis

Four radiologists participated in the study, two experienced neuroradiologists (B.S. with 14-year radiology experience and S.Y.W. with 10-year experience, referred to as readers A and B, respectively) and two less experienced radiology residents (C.Y.L. with 4-year experience and J.Y.P. with 3-year experience, referred to as readers C and D, respectively). Readers A and B had 3 years and 1 year of SMwI experience, respectively. While readers C and D had no clinical experience with SMwI, they completed a separate training with 10 SMwI cases that were not included as study participants before this study. The four radiologists were randomly divided into two groups, each including one experienced and one less experienced reader. Participants were divided into two equal groups. In the crossover design, reader group 1 initially evaluated half of the participants with AI and the other half without AI, whereas reader group 2 evaluated the same half in the opposite manner. After a 2-month washout period, the assignments and reading modes were reversed.

In the SMwI image, N1 was identified as a hyperintense area anterior to the inferior portion of the red nucleus and was visualized between the curvilinear hypointense bands. Abnormalities were defined as disruptions of the three-layer structure (inner, middle, and outer layers) (Fig. 3) [27]. The left and right N1 regions were evaluated separately as either 'normal' or 'abnormal.' In AI-assisted interpretation, readers reviewed the SMwI and AI-generated assessments simultaneously to make their final decisions.

Statistical Analyses

Clinical and demographic data were compared using the Wilcoxon rank-sum and chi-square tests. Diagnostic performance (i.e., sensitivity, specificity, and accuracy) was assessed by integrating the evaluations of the right and left N1 regions using clinical diagnosis and DaT PET results as reference standards. Generalized estimating equations were used for statistical analysis. Fleiss' kappa statistics measured the inter-reader agreement [28].

To better characterize the interaction between readers and the AI, the net reclassification index (NRI) [29] was calculated, and disagreements with AI-generated assessments were analyzed. Reader interpretations were compared with the AI results separately for PD and NC

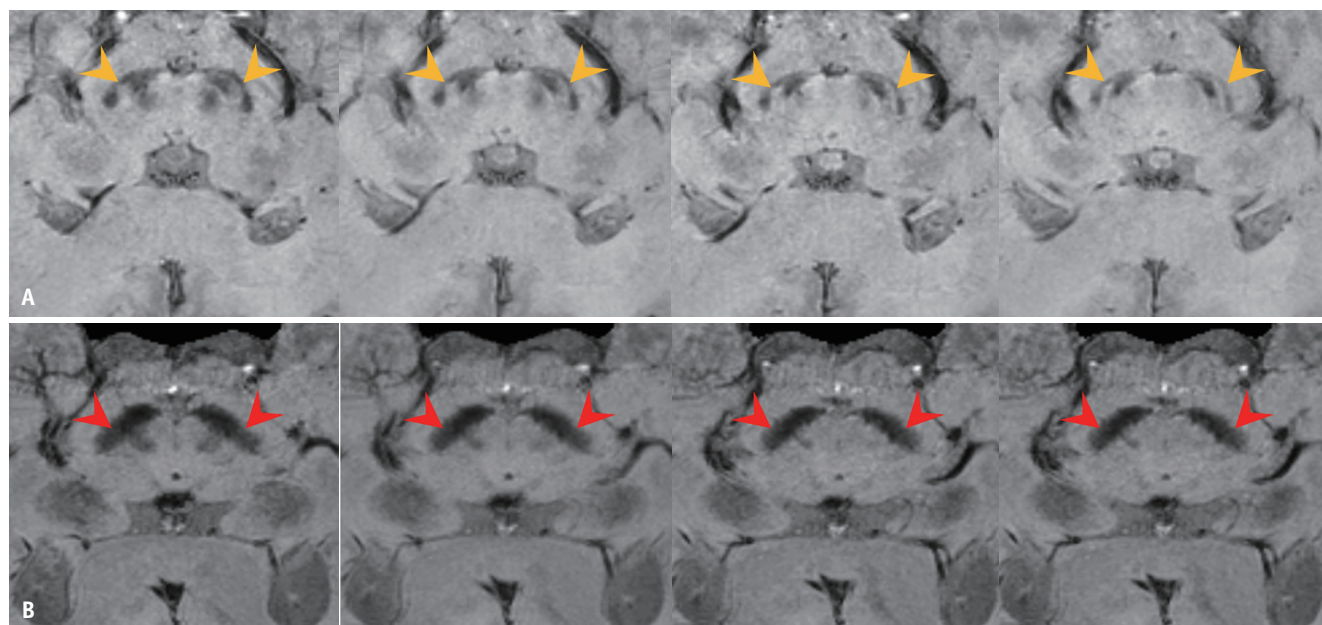


Fig. 3. Representative cases of normal and abnormal nigrosome 1. **A:** In a clinically normal 57-year-old female without neurological symptoms, the AI software detected bilateral normal nigrosome 1 region (arrowheads). The three-layer structure- hypointense inner, hyperintense middle, and hypointense outer- is clearly distinguishable anterior and inferior to the red nucleus. **B:** In a 67-year-old male clinically diagnosed with Parkinson's disease and evaluated with SMwI, the AI software detected bilateral abnormalities in the nigrosome 1 region, characterized by disruption of the three distinct layers (arrowheads). AI = artificial intelligence, SMwI = susceptibility map-weighted imaging

cases, with further stratification by reader experience level. For the NRI analysis, "PD up" and "NC up" referred to cases which AI corrected initial misclassifications into PD or NC, respectively, whereas "PD down" and "NC down" indicated cases where AI changed a correct initial classification to an incorrect one. The NRI was computed as the difference between the proportion of improved and worsened classifications based on the clinical diagnosis as a reference. Confidence intervals (CIs) were estimated using the bootstrap method with 1000 re-samples. For the disagreement analysis, the proportions of discordant interpretations between the readers and AI were compared using the chi-square test, and CIs were calculated.

Finally, the standalone performance of the AI model was assessed by comparison with a reference standard (i.e., clinical diagnosis and DaT PET results). Incorrect AI results were reviewed by an expert neuroradiologist (E.Y.K., with 5 years of SMwI experience) to examine their SMwI findings.

RESULTS

A total of 139 SMwI scans were analyzed retrospectively, including 59 patients with PD (PD group) (50–88 years, mean $\pm$ standard deviation 71.1 ± 8.9 years; 38 males and 21 females) and 80 clinically normal participants (NC group) (35–85 years, 60.5 ± 16.0 years; 27 males and 53 females). The PD group was significantly older than the NC group ($P < 0.001$) (Table 1). The average Hoehn and Yahr (H&Y) score in the PD group was 2.74 points, with the most common score being 3 (26/59, 44.1%).

First, the diagnostic performances of the four readers, neuroradiologists A and B and radiology residents C and D, were evaluated by comparing their individual interpretations of the left or right N1 to the clinical diagnosis, both with and without referring to AI. Three of the four readers (A, C, and D) showed improved specificity when aided by AI: 0.86 vs. 0.94, $P = 0.004$; 0.91 vs. 0.97, $P = 0.024$; and 0.90 vs. 0.97, $P = 0.012$). Further details are presented in Table 2.

Agreement among all four readers as measured with Fleiss's kappa with corresponding 95% CIs improved with AI assistance, from 0.73 (95% CI: 0.61–0.84) without AI to 0.87 (95% CI: 0.76–0.99) with AI, based on integrated evaluations of the right and left N1 regions.

To evaluate the readers' responses to AI suggestions, the NRI and disagreement rates were assessed (Tables 3, 4). Statistically significant improvements in classification with AI ($P < 0.05$) were observed for readers D (NRI = 15.3%, 95% CI: 0.074–0.241), C (NRI = 10.3%, 95% CI: 0.005–0.192),

Table 1. Study participants

Variable	PD group (n = 59)	NC group (n = 80)	P
Age, yr	71.1 $\pm$ 8.87 (50–88)	60.5 $\pm$ 15.97 (35–85)	<0.001*
Sex			<0.001†
Male	38 (64.4)	27 (33.8)	
Female	21 (35.6)	53 (66.3)	
H&Y score	2.7 $\pm$ 0.63 (1–5)	NA	NA

Data are presented as mean $\pm$ standard deviation with the range in parentheses or number of participants (%).

*Wilcoxon rank sum test, †Chi-square test.

PD = Parkinson's disease, NC = normal control, H&Y = Hoehn and Yahr scale for functional disability staging, NA = not applicable

Table 2. Diagnostic performances of readers with vs. without AI assistance

Performance parameter	Reader	Without AI	With AI	P
Sensitivity	A	0.93 (0.85–0.97) [110/118]	0.92 (0.83–0.97) [109/118]	0.759
	B	0.96 (0.87–0.99) [113/118]	0.91 (0.81–0.96) [107/118]	0.168
	C	0.80 (0.72–0.91) [94/118]	0.83 (0.74–0.90) [98/118]	0.483
	D	0.84 (0.72–0.91) [99/118]	0.92 (0.84–0.97) [109/118]	0.096
Specificity	A	0.86 (0.77–0.91) [137/160]	0.94 (0.87–0.97) [150/160]	0.004
	B	0.96 (0.89–0.98) [153/160]	0.95 (0.88–0.98) [152/160]	0.764
	C	0.91 (0.84–0.96) [146/160]	0.97 (0.88–0.99) [157/160]	0.024
	D	0.90 (0.82–0.95) [144/160]	0.97 (0.90–0.99) [155/160]	0.012
Accuracy	A	0.89 (0.83–0.93) [247/278]	0.93 (0.88–0.96) [259/278]	0.039
	B	0.96 (0.91–0.98) [266/278]	0.93 (0.88–0.96) [259/278]	0.201
	C	0.86 (0.81–0.90) [240/278]	0.92 (0.87–0.96) [255/278]	0.041
	D	0.87 (0.81–0.92) [243/278]	0.95 (0.90–0.97) [264/278]	0.005

Values are presented as the proportion with 95% confidence interval in parentheses and numerator/denominator in brackets. Statistical analysis was conducted using generalized estimating equations. Readers A and B are experienced neuroradiologists with 14 and 10 years of experience, respectively; Readers C and D are relatively less experienced radiology residents with 4 and 3 years of experience, respectively. AI = artificial intelligence

and A (NRI = 7.3%, 95% CI: 0.011–0.141), because their 95% CIs did not include zero. Reader B showed a negative and non-significant NRI (NRI = -5.7%, 95% CI: -0.120–0.004). When grouped, the less experienced readers (C and D) demonstrated a statistically significant NRI (NRI = 12.8%, 95% CI: 0.067–0.190), whereas the experienced group (A and B) did not (NRI = 0.8%, 95% CI: -0.037–0.051). The overall NRI across all readers was 6.8% (95% CI: 0.027–0.107), also indicating statistical significance. Disagreement

Table 3. Diagnostic reclassification analysis

		NRI (%)	95% CI
By reader	A	7.3	0.011–0.141
	B	-5.7	-0.120–0.004
	C	10.3	0.005–0.192
	D	15.3	0.074–0.241
By group	A & B	0.8	-0.037–0.051
	C & D	12.8	0.067–0.190
By total	A–D	6.8	0.027–0.107

NRI by reader and group. Positive values indicate improved diagnostic performance with artificial intelligence. 95% CIs were calculated using the bootstrap method. Readers A and B are experienced neuroradiologists with 14 and 10 years of experience, respectively. Readers C and D are less experienced radiology residents with 4 and 3 years of experience, respectively. NRI = net reclassification index, CI = confidence interval

Table 4. Disagreement rate analysis

			Disagreement rate (%)	95% CI	P
By reader	A	PD	4.2	0.006–0.079	0.766
		NC	5.0	0.016–0.084	
	B	PD	4.2	0.006–0.079	0.463
		NC	6.3	0.025–0.100	
	C	PD	11.9	0.060–0.177	0.004
		NC	3.1	0.004–0.058	
	D	PD	4.2	0.006–0.079	0.956
		NC	4.4	0.012–0.075	
By group	A & B	PD	4.2	0.017–0.068	0.460
		NC	5.6	0.031–0.082	
	C & D	PD	8.1	0.046–0.115	0.029
		NC	3.8	0.017–0.058	
By total	A–D	PD	6.0	0.039–0.082	0.319
		NC	4.7	0.031–0.063	

Disagreement rates between reader interpretations and artificial intelligence results, shown separately for PD and NC. 95% CIs and *P*-values were obtained using the chi-square test. Readers A and B are experienced neuroradiologists with 14 and 10 years of experience, respectively; readers C and D are less experienced radiology residents with 4 and 3 years of experience, respectively. PD = Parkinson's disease, NC = normal control, CI = confidence interval

rates between readers and AI were significantly higher in PD cases for reader C (PD: 11.9%, 95% CI: 0.060–0.177; NC: 3.1%, 95% CI: 0.004–0.058; *P* = 0.004) and in less experienced groups C and D (PD: 8.1%, 95% CI: 0.046–0.115; NC: 3.8%, 95% CI: 0.017–0.058; *P* = 0.029). No consistent trends were observed for the other readers.

Finally, AI-generated assessments were compared with the reference standard. Of the 278 cases, the AI-generated results were correct in 264 cases (264/278, 95.0%) and incorrect in 14 cases (14/278, 5.0%). Among the incorrect cases, 6 cases (6/278, 2.2%) were deemed apparent AI errors, as the SMwI findings were clearly abnormal (*n* = 3) or normal (*n* = 3). Six other cases (6/278, 2.2%) had ambiguous SMwI findings, which made interpretation by human experts challenging. The remaining 2 cases (2/278, 0.7%), bilateral N1 regions from a single patient, showed normal-appearing SMwI; however, the patient was confirmed to have PD, highlighting instances in which the disease may not be visually apparent on SMwI.

DISCUSSION

This study revealed that the DL-based AI software improved the diagnostic performance of radiologists in detecting PD-associated abnormalities when analyzing N1 on brain MRI. Previous studies have shown that patients with PD show decreased hyperintensity of N1, a subregion of the SNpc, on SMwI. However, a detailed SNpc analysis can be time-consuming and error-prone, particularly for those with less expertise. Recently, a DL-based AI software was developed to assist radiologists by improving object detection using the YOLOX model [24] and precise segmentation with SparseInst [25], aiding the classification and quantification of N1 abnormalities.

This study compared the diagnostic performances of two experienced neuroradiologists (A and B) and two less experienced radiology residents (C and D) with and without AI assistance. Three (neuroradiologist A and residents C and D) of the four readers demonstrated improvements in the diagnostic performance with AI. Specifically, AI significantly enhanced the specificity and accuracy of these readers (*P* < 0.05), suggesting its potential to reduce false-positives and improve diagnostic reliability. It also improved inter-reader agreement, as reflected by the increased Fleiss' kappa values and narrower CIs across both experienced and less experienced readers. In contrast, neuroradiologist B showed a slight decrease in diagnostic

performance with no statistically significant improvement in any metric, indicating that experienced neuroradiologists may benefit less from AI or may already operate near ceiling performance levels. In addition, the time required to review AI-generated outputs may present a potential disadvantage. Nonetheless, AI is expected to benefit radiologists across all levels of experience in real-world settings. Unlike in this study, in which readers focused solely on the N1 region, real-world practice requires interpreting the entire brain across multiple sequences, making it challenging to maintain the same level of attention in specific areas. In such settings, AI-assistance can enhance diagnostic performance. Further evaluations should include large-scale studies conducted in broader diagnostic environments including whole-brain interpretations.

Analysis of the NRI demonstrated that AI-assistance improved diagnostic performance in most readers, with a particularly notable effect in the less-experienced group. Readers D and C showed the highest NRI values (15.3% and 10.3%, respectively), and their combined performance yielded a statistically significant improvement (NRI = 12.8%, 95% CI: 0.067–0.190). In contrast, the experienced

neuroradiologist group (A and B) showed a minimal NRI gain (0.8%, 95% CI: -0.037–0.051), suggesting that AI had a more substantial effect on less experienced readers. Notably, in the less-experienced group, the disagreement rates in the PD cases were significantly higher than those in the NC cases (Reader C: 11.9% vs. 3.1%, $P = 0.004$; Group C and D: 8.1% vs. 3.8%, $P = 0.029$). This likely reflects the inherent difficulty in interpreting PD and greater variability in judgment among less experienced readers. The higher disagreement, along with the improved NRI, suggests that AI-assistance may have supported less experienced readers in making more accurate decisions in challenging cases, thereby contributing to the overall diagnostic improvement.

Discrepancies between AI-generated assessments and the reference standard offer insights into the limitations of AI software and SMwI. Although the AI results were concordant with clinical diagnoses in most cases, a small number of discrepancies (5%) were reviewed in detail. Among these, six were considered concordant with the SMwI findings of expert review, suggesting that the discrepancies may have stemmed from AI interpretation. Owing to the nature of DL-based AI, the rationale behind its assessment is often difficult to

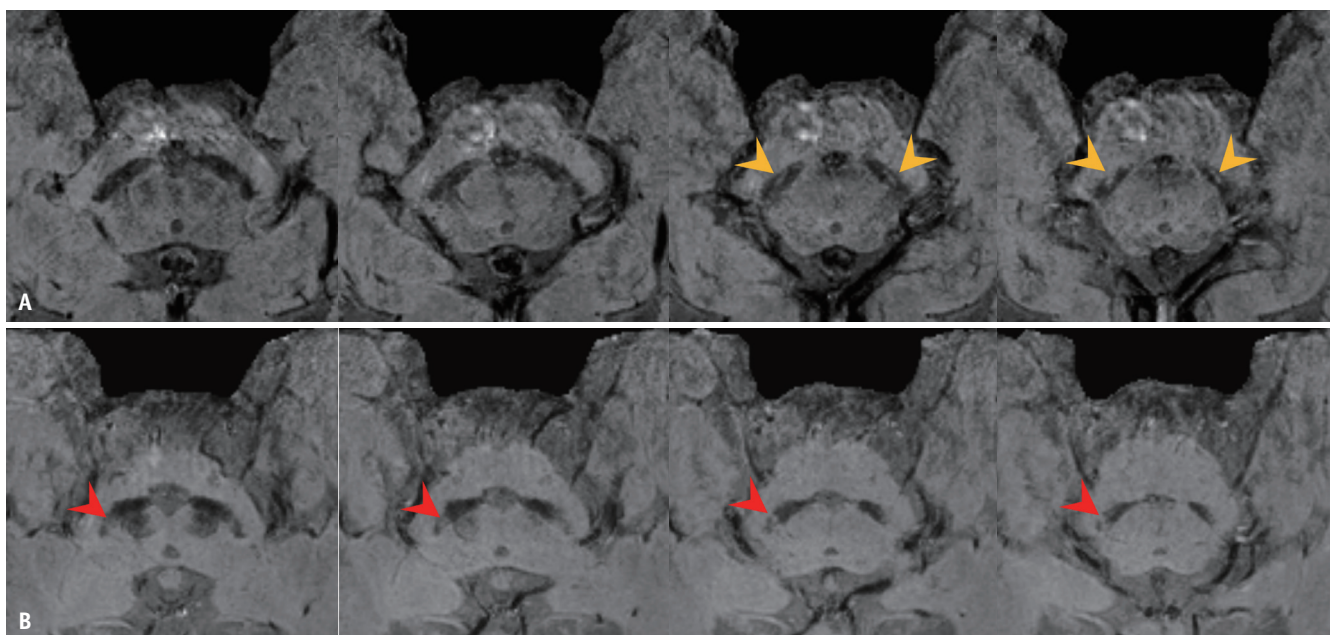


Fig. 4. Representative cases showing incorrect interpretations by the AI software. **A:** In the SMwI image of a 41-year-old male with no neurological symptoms, the AI software classified the right side as abnormal and the left side as normal. However, the SMwI clearly shows a well-defined hyperintense structure corresponding to nigrosome 1 on both sides (arrowheads). The reason for the AI's misclassification of the right side remains unclear. **B:** In the SMwI image of an 82-year-old female diagnosed with Parkinson's disease, the AI software classified the right side as normal and the left side as abnormal. A focal hyperintensity was observed in the right substantia nigra (arrowheads), which the AI likely interpreted as a normal nigrosome 1. However, it is considered a false-negative finding as it is located lateral rather than anterior to the red nucleus and lacks a clearly distinguishable three-layer structure. AI = artificial intelligence, SMwI = susceptibility map-weighted imaging

understand. The source of the discrepancy was unclear in most discordant cases (Fig. 4A). In one case in the PD group (Fig. 4B), the AI software classified the right side as normal. SMwI revealed a hyperintense region in the right substantia

nigra. However, this region appears before the red nucleus disappears, is located laterally rather than anteriorly, and lacks a clearly distinguishable three-layer structure, which may account for the observed discordance. The training and

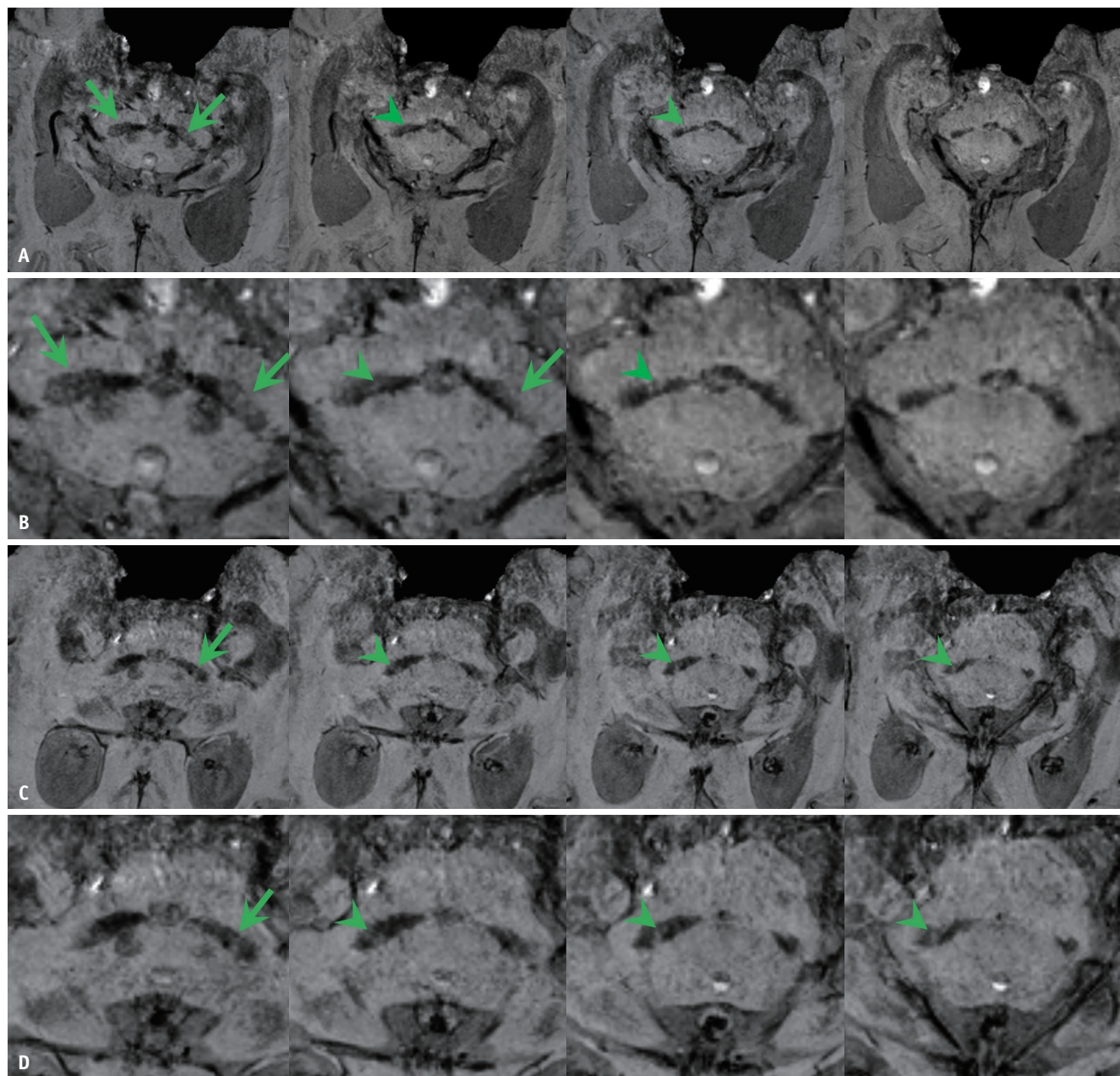


Fig. 5. Representative cases showing ambiguous SMwI findings for nigrosome 1. **A, B:** In the SMwI image of an 80-year-old male diagnosed with PD, the AI software classified both sides as normal. In the patient's SMwI (**A**) and the magnified images (**B**), the bilateral substantia nigra shows relatively heterogeneous intensity, making interpretation challenging (arrows). Notably, an indistinct linear hyperintensity band is observed in the right substantia nigra (arrowheads). Three out of four radiologists initially identified an abnormality on the right side without AI but later interpreted it as normal with AI assistance. **C, D:** In the SMwI image of an 88-year-old female diagnosed with PD, the AI software classified both sides as normal. In the patient's SMwI (**C**) and the magnified image (**D**), an indistinct band-like hyperintensity is observed in the right substantia nigra (arrowheads), suggesting an ambiguous finding, while a partially heterogeneous portion is seen in the left substantia nigra (arrow). Two out of four radiologists interpreted the findings as bilateral abnormalities without AI but as bilateral normal with AI assistance. SMwI = susceptibility map-weighted imaging, PD = Parkinson's disease, AI = artificial intelligence

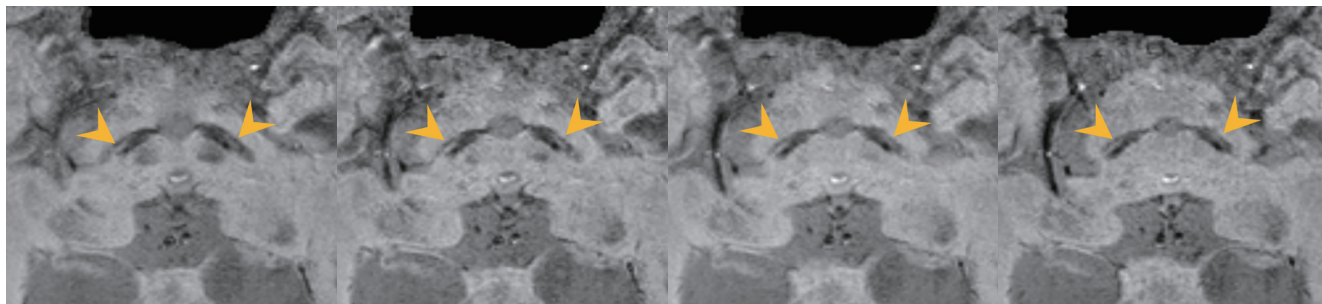


Fig. 6. Representative case showing a discrepancy between SMwI and DaT PET findings. A 80-year-old female clinically diagnosed with Parkinson's disease was interpreted by the AI software as having bilateral normal findings in the nigrosome 1 region (arrowheads). All four radiologists also assessed the findings as normal, both with and without AI. SMwI = susceptibility map-weighted imaging, DaT = dopamine transporter, AI = artificial intelligence

refinement of AI software with a large volume of additional data can help reduce such misinterpretations. However, since atypical-looking N1 cases [30] have been reported without the characteristic “swallowtail sign,” which is defined by two hypointense tails with a hyperintense middle, radiologists’ final judgment remains essential even when AI software achieves greater accuracy.

Six other cases were considered difficult to interpret because of the ambiguous SMwI findings. Figure 5 illustrates examples in which the indistinct and heterogeneous signal intensity of the N1 makes it challenging to determine its abnormality. These cases also showed high inconsistencies between the readers’ interpretations with and without AI. This is likely due to the limitations of the SMwI images. Addressing this issue could involve improving image quality through adjustments to MR parameters or the use of an image-quality-enhancing AI software [31]. Two cases involved bilateral N1 evaluation of a single patient with PD (Fig. 6), showing a discrepancy between the SMwI and DaT PET findings. Similar cases of PD with either normal SMwI [32] or DaT PET [33] have been reported. These findings underscore the need to integrate clinical symptoms with multiple diagnostic modalities and ensure sufficient follow-up for an accurate PD diagnosis.

This study has several limitations. First, the NC group did not undergo DaT PET for ethical reasons, which could have affected the accuracy of group classification. Additionally, other disease groups with negative DaT PET results, such as those with essential tremors, were not included, limiting the generalizability of the findings to real-world clinical settings. Including disease controls and normal controls with confirmed negative DaT PET findings could provide a more clinically relevant comparison in future studies. Second, there was an unbalanced age distribution between

the patient and healthy subject groups, which may have introduced bias and affected the generalizability of the findings. However, to date, no studies have reported that the diagnostic results of SMwI are influenced by the patient’s age or sex. Third, the number of study participants was relatively small ($n = 139$), limiting the statistical power of the analysis; the study was conducted at a single center, which may not fully represent broader clinical populations. These limitations highlight the need for further large-scale, multicenter, prospective studies to validate these findings and improve their generalizability across diverse settings.

In conclusion, this study demonstrates that DL-based AI software improves the diagnostic performance of radiologists in evaluating N1 abnormalities in SMwI, particularly benefiting less experienced readers. Despite some limitations, these findings highlight the potential of AI in enhancing diagnostic accuracy and reducing variability in the assessment of PD. Further large-scale studies are needed to validate these results and explore broader clinical applications.

Supplement

The Supplement is available with this article at <https://doi.org/10.3348/kjr.2025.0208>.

Availability of Data and Material

The datasets generated or analyzed during the study are not publicly available due to ethical restrictions and patient privacy concerns but are available from the corresponding author on reasonable request.

Conflicts of Interest

The authors have no potential conflicts of interest to

disclose.

Author Contributions

Conceptualization : Jiyeon Park, Beomseok Sohn. Data curation: Yongsik Sim, Beomseok Sohn. Formal analysis: Jiyeon Park, Chae Young Lim, So Yeon Won, Yun Hwa Roh, Sun-Young Baek. Investigation: Jiyeon Park. Methodology: Jiyeon Park, Beomseok Sohn. Project administration: Beomseok Sohn. Resources: Han Kyu Na, Phil Hyu Lee. Software: Jiyeon Park, Sun-Young Baek. Supervision: Eung Yeop Kim, Sung Tae Kim. Validation: Jiyeon Park, Sun-Young Baek. Visualization: Jiyeon Park, Beomseok Sohn. Writing—original draft: Jiyeon Park. Writing—review & editing: Jiyeon Park, Beomseok Sohn.

ORCID IDs

Jiyeon Park

<https://orcid.org/0000-0002-0353-0109>

Chae Young Lim

<https://orcid.org/0009-0000-7858-5430>

So Yeon Won

<https://orcid.org/0000-0003-0570-3365>

Han Kyu Na

<https://orcid.org/0000-0002-2093-1641>

Phil Hyu Lee

<https://orcid.org/0000-0001-9931-8462>

Sun-Young Baek

<https://orcid.org/0000-0001-6519-7770>

Yun Hwa Roh

<https://orcid.org/0000-0002-8041-1621>

Minjung Seong

<https://orcid.org/0000-0002-9257-4225>

Yongsik Sim

<https://orcid.org/0000-0003-2711-2793>

Eung Yeop Kim

<https://orcid.org/0000-0002-9579-4098>

Sung Tae Kim

<https://orcid.org/0000-0001-8185-0063>

Beomseok Sohn

<https://orcid.org/0000-0002-6765-8056>

Funding Statement

None

Acknowledgments

The authors would like to acknowledge Heuron Co., Ltd.

for their assistance in preparing the description of the AI algorithm in the Methods section and in the creation of Figure 2.

REFERENCES

1. Pringsheim T, Jette N, Frolkis A, Steeves TD. The prevalence of Parkinson's disease: a systematic review and meta-analysis. *Mov Disord* 2014;29:1583-1590
2. Postuma RB, Berg D, Stern M, Poewe W, Olanow CW, Oertel W, et al. MDS clinical diagnostic criteria for Parkinson's disease. *Mov Disord* 2015;30:1591-1601
3. Postuma RB, Poewe W, Litvan I, Lewis S, Lang AE, Halliday G, et al. Validation of the MDS clinical diagnostic criteria for Parkinson's disease. *Mov Disord* 2018;33:1601-1608
4. Malek N, Lawton MA, Grosset KA, Bajaj N, Barker RA, Ben-Shlomo Y, et al. Utility of the new Movement Disorder Society clinical diagnostic criteria for Parkinson's disease applied retrospectively in a large cohort study of recent onset cases. *Parkinsonism Relat Disord* 2017;40:40-46
5. Virameteekul S, Revesz T, Jaunmuktane Z, Warner TT, De Pablo-Fernández E. Clinical diagnostic accuracy of Parkinson's disease: where do we stand? *Mov Disord* 2023;38:558-566
6. Fearnley JM, Lees AJ. Ageing and Parkinson's disease: substantia nigra regional selectivity. *Brain* 1991;114(Pt 5):2283-2301
7. Blazejewska AI, Schwarz ST, Pitiot A, Stephenson MC, Lowe J, Bajaj N, et al. Visualization of nigrosome 1 and its loss in PD: pathoanatomical correlation and in vivo 7 T MRI. *Neurology* 2013;81:534-540
8. Kwon DH, Kim JM, Oh SH, Jeong HJ, Park SY, Oh ES, et al. Seven-tesla magnetic resonance images of the substantia nigra in Parkinson disease. *Ann Neurol* 2012;71:267-277
9. Cosottini M, Frosini D, Pesaresi I, Costagli M, Biagi L, Ceravolo R, et al. MR imaging of the substantia nigra at 7 T enables diagnosis of Parkinson disease. *Radiology* 2014;271:831-838
10. Noh Y, Sung YH, Lee J, Kim EY. 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* 2015;36:2010-2016
11. Haacke EM, Xu Y, Cheng YC, Reichenbach JR. Susceptibility weighted imaging (SWI). *Magn Reson Med* 2004;52:612-618
12. Rauscher A, Sedlacik J, Barth M, Mentzel HJ, Reichenbach JR. Magnetic susceptibility-weighted MR phase imaging of the human brain. *AJNR Am J Neuroradiol* 2005;26:736-742
13. Shmueli K, de Zwart JA, van Gelderen P, Li TQ, Dodd SJ, Duyn JH. Magnetic susceptibility mapping of brain tissue in vivo using MRI phase data. *Magn Reson Med* 2009;62:1510-1522
14. Gho SM, Liu C, Li W, Jang U, Kim EY, Hwang D, et al. Susceptibility map-weighted imaging (SMWI) for neuroimaging. *Magn Reson Med* 2014;72:337-346
15. Nam Y, Gho SM, Kim DH, Kim EY, Lee J. Imaging of nigrosome 1 in substantia nigra at 3T using multiecho susceptibility

- map-weighted imaging (SMWI). *J Magn Reson Imaging* 2017;46:528-536
16. 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
17. Sung YH, Lee J, Nam Y, Shin HG, Noh Y, Hwang KH, et al. Initial diagnostic workup of parkinsonism: dopamine transporter positron emission tomography versus susceptibility map-weighted imaging at 3T. *Parkinsonism Relat Disord* 2019;62:171-178
18. McDonald RJ, Schwartz KM, Eckel LJ, Diehn FE, Hunt CH, Bartholmai BJ, et al. The effects of changes in utilization and technological advancements of cross-sectional imaging on radiologist workload. *Acad Radiol* 2015;22:1191-1198
19. Sohn B, Park KY, Choi J, Koo JH, Han K, Joo B, et al. Deep learning-based software improves clinicians' detection sensitivity of aneurysms on brain TOF-MRA. *AJNR Am J Neuroradiol* 2021;42:1769-1775
20. Liu F, Zhou Z, Samsonov A, Blankenbaker D, Larison W, Kanarek A, et al. Deep learning approach for evaluating knee MR images: achieving high diagnostic performance for cartilage lesion detection. *Radiology* 2018;289:160-169
21. Shin DH, Heo H, Song S, Shin NY, Nam Y, Yoo SW, et al. Automated assessment of the substantia nigra on susceptibility map-weighted imaging using deep convolutional neural networks for diagnosis of idiopathic Parkinson's disease. *Parkinsonism Relat Disord* 2021;85:84-90
22. Lee S, Suh CH, Jo S, Chung SJ, Heo H, Shim WH, et al. Comparative performance of susceptibility map-weighted MRI according to the acquisition planes in the diagnosis of neurodegenerative parkinsonism. *Korean J Radiol* 2024;25:267-276
23. Sung YH, Noh Y, Lee J, Kim EY. Drug-induced Parkinsonism versus idiopathic Parkinson disease: utility of nigrosome 1 with 3-T imaging. *Radiology* 2016;279:849-858
24. Suh PS, Heo H, Suh CH, Lee M, Song S, Shin D, et al. Deep learning-based algorithm for automatic quantification of nigrosome-1 and Parkinsonism classification using susceptibility map-weighted MRI. *AJNR Am J Neuroradiol* 2025;46:999-1006
25. Ge Z, Liu S, Wang F, Li Z, Sun J. YOLOX: exceeding YOLO series in 2021. arXiv [Preprint]. 2021 [accessed on May 8, 2025]. Available at: <https://doi.org/10.48550/arXiv.2107.08430>
26. Cheng T, Wang X, Chen S, Zhang W, Zhang Q, Huang C, et al. Sparse instance activation for real-time instance segmentation arXiv [Preprint]. 2022 [accessed on May 8, 2025]. Available at: <https://doi.org/10.48550/arXiv.2203.12827>
27. Kim EY, Sung YH, Lee J. Nigrosome 1 imaging: technical considerations and clinical applications. *Br J Radiol* 2019;92:20180842
28. Han K, Ryu L. Statistical methods for the analysis of inter-reader agreement among three or more readers. *Korean J Radiol* 2024;25:325-327
29. Pencina MJ, D'Agostino RB Sr, D'Agostino RB Jr, Vasan RS. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. *Stat Med* 2008;27:157-172; discussion 207-212
30. Cheng Z, He N, Huang P, Li Y, Tang R, Sethi SK, et al. Imaging the nigrosome 1 in the substantia nigra using susceptibility weighted imaging and quantitative susceptibility mapping: an application to Parkinson's disease. *Neuroimage Clin* 2020;25:102103
31. Suh PS, Park JE, Roh YH, Kim S, Jung M, Koo YS, et al. Improving diagnostic performance of MRI for temporal lobe epilepsy with deep learning-based image reconstruction in patients with suspected focal epilepsy. *Korean J Radiol* 2024;25:374-383
32. Bae YJ, Song YS, Choi BS, Kim JM, Nam Y, Kim JH. Comparison of susceptibility-weighted imaging and susceptibility map-weighted imaging for the diagnosis of Parkinsonism with nigral hyperintensity. *Eur J Radiol* 2021;134:109398
33. Palermo G, Giannoni S, Depalo T, Frosini D, Volterrani D, Siciliano G, et al. Negative DAT-SPECT in old onset Parkinson's disease: an additional pitfall? *Mov Disord Clin Pract* 2022;9:530-534