AJNR

Deep learning-based algorithm for automatic quantification of nigrosome-1 and Parkinsonism classification using susceptibility map-weighted MRI

Pae Sun Suh, Hwan Heo, Chong Hyun Suh, Myeong Oh Lee, Soohwa Song, Dong Hoon Shin, Sung Yang Jo, Sun Ju Chung, Hwon Heo, Woo Hyun Shim, Ho Sung Kim, Sang Joon Kim and Eung Yeop Kim

AJNR Am J Neuroradiol published online 15 November 2024
http://www.ajnr.org/content/early/2024/11/15/ajnr.A8585

ORIGINAL RESEARCH

Deep learning-based algorithm for automatic quantification of nigrosome-1 and Parkinsonism classification using susceptibility map-weighted MRI

Pae Sun Suh*, Hwan Heo*, Chong Hyun Suh, Myeong Oh Lee, Soohwa Song, Dong Hoon Shin, Sung Yang Jo, Sun Ju Chung, Hwon Heo, Woo Hyun Shim, Ho Sung Kim, Sang Joon Kim, Eung Yeop Kim

ABSTRACT

BACKGROUND AND PURPOSE: To develop and validate a deep learning-based automatic quantification for nigral hyperintensity and a classification algorithm for neurodegenerative parkinsonism using susceptibility map-weighted imaging (SMwI).

MATERIALS AND METHODS: We retrospectively collected 450 participants (210 with idiopathic Parkinson's disease [IPD] and 240 individuals in the control group) for training data between November 2022 and May 2023, and 237 participants (168 with IPD, 58 with essential tremor, and 11 with drug-induced Parkinsonism) for validation data between July 2021 and January 2022. SMwI data were reconstructed from multi-echo GRE. Diagnostic performance for diagnosing IPD was assessed using deep learning-based automatic quantification (Heuron NI) and classification (Heuron IPD) model. Reference standard for IPD was based on ^{18}F-FP-CIT PET finding. Additionally, the correlation between the H&Y stage and volume of nigral hyperintensity in patients with IPD was assessed.

RESULTS: Quantification of nigral hyperintensity using Heuron NI showed AUC of 0.915 (95% CI, 0.872–0.947) and 0.928 (95% CI, 0.887–0.957) on the left and right, respectively. Classification of nigral hyperintensity abnormality using Heuron IPD showed AUC of 0.967 (95% CI, 0.881–0.991) and 0.976 (95% CI, 0.948–0.992) on the left and right, respectively. H&Y score ≥ 3 showed significant smaller nigral hyperintensity volume (1.43 ± 1.19 mm^3) compared to H&Y score 1–2.5 (1.98 ± 1.63 mm^3; p = 0.008).

CONCLUSIONS: Our deep learning-based model proves rapid, accurate automatic quantification of nigral hyperintensity, facilitating IPD diagnosis, symptom severity prediction, and patient stratification for personalized therapy. Further study is warranted to validate the findings across various clinical settings.

ABBREVIATIONS: IPD = Idiopathic Parkinson's disease; SN = substantia nigra; SMwI = susceptibility map weighted imaging; QSM = quantitative susceptibility mapping; CNN = convolutional neural network; ICV = intracranial volume.

Received April 04, 2024; accepted after revision October 21, 2024.
Department of Radiology and Research Institute of Radiology (P.S.S., C.H.S., H.H., W.H.S., H.S.K., S.J.K.), Asan Medical Center, University of Ulsan College of Medicine, Seoul, Republic of Korea; Department of Radiology and Research Institute of Radiological Science and Center for Clinical Imaging Data Science(P.S.S.), Yonsei University College of Medicine, Seoul, South Korea; Heuron Co.Ltd. (H.H., M.O.L., S.H.S., D.H.S.) Seoul, Republic of Korea; Department of Neurology (S.J.C., S.Y.J.), Asan Medical Center, University of Ulsan College of Medicine, Seoul, Republic of Korea; Department of Radiology (E.Y.K), Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, Republic of Korea
*Both authors(P.S.S., H.H.) contributed equally.

The authors of this manuscript declare relationships with the following companies: Heuron Co. Ltd.
This research was supported by a grant of the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare, Republic of Korea (grant number: RS-2023-00266948).

Please address correspondence to Chong Hyun Suh, MD, PhD, Department of Radiology and Research Institute of Radiology, University of Ulsan College of Medicine, Asan Medical Center, Olympic-ro 33, Seoul 05505, Korea; e-mail: chonghyunsuh@amc.seoul.kr

SUMMARY SECTION

PREVIOUS LITERATURE: Loss of nigrosome-1 in idiopathic Parkinson's disease (IPD) serves as an early diagnostic imaging biomarker. Although several deep-learning models have been proposed for diagnosing IPD and segmenting the substantia nigra or nigrosome-1 region, these models have not simultaneously segmented nigrosome-1 and diagnosed IPD. Moreover, previous studies have demonstrated relatively low diagnostic performance and lacked external validation. Therefore, we propose a fully automated deep-learning model using susceptibility map-weighted imaging (SMwI) to identify and determine abnormalities in the nigrosome-1 region (Heuron IPD), along with quantifying volume of nigral hyperintensity (Heuron NI).

KEY FINDINGS: Quantification of nigral hyperintensity using Heuron NI showed AUC of 0.915 and 0.928 on the left and right, respectively. Classification of nigral abnormalities using Heuron IPD showed AUC of 0.967 and 0.976. H&Y score ≥3 showed significant smaller nigral hyperintensity volume (1.43±1.19 mm^3) compared to H&Y score 1–2.5 (1.98±0.63 mm^3; p=0.008).

KNOWLEDGE ADVANCEMENT: Our deep learning-based model enables rapid and accurate automated quantification and classification of nigral hyperintensity, facilitating the diagnosis of IPD and predicting symptom severity.

Copyright 2024 by American Society of Neuroradiology.

INTRODUCTION

Idiopathic Parkinson's disease (IPD) is a chronic and progressive neurodegenerative disease causing Parkinsonian symptoms, including tremors, bradykinesia, rigidity, and postural instability.[1] The principal pathophysiological change associated with this disease is dopaminergic neuron degeneration in the substantia nigra (SN) pars compacta.[2] The nigrosome-1 is located in the dorsolateral portion of the SN. It is the initial site of dopaminergic neuron degeneration as the disease progresses.[3] The nigrosome-1 typically contains high neuromelanin and low iron concentrations. Therefore, this region usually identified by the hyperintense area at the dorsolateral aspect of the SN on susceptibility-weighted imaging (SWI), also known as the "swallow tail sign".[4]

Hyperintensity loss in nigrosome-1 occurs in IPD,[5-7] serving as an early imaging biomarker of IPD.[8] A recent meta-analysis demonstrated that detecting loss of nigrosome-1 yields excellent diagnostic accuracy, with pooled sensitivity and specificity of 0.96 and 0.95, respectively.[9] However, visualizing nigrosome-1 varies subjectively across imaging sequences, including T2* weighted imaging, SWI, and quantitative susceptibility mapping (QSM).[10] Among these imaging sequences, susceptibility map-weighted imaging (SMwI), which is similar to SWI but generated by combining the GRE magnitude image and a QSM-derived quantitative susceptibility mask, enhances susceptibility contrast of nigrosome-1.[11-14] Enhanced contrast in nigrosome-1 allows for better visualization of nigral hyperintensity structural information and improves diagnostic performance.

Recently, many deep learning models have been proposed for neurodegenerative diseases.[15-17] In the case of IPD, deep learning models have been developed not only for diagnosis but also for segmenting the SN and nigrosome-1 regions. Chen et al. developed hybrid model for predicting early-stage IPD that combines deep learning features and radiomics features, employing YOLO-v5 for initial brainstem detection and integrating both feature for enhanced predictive accuracy.[18] Shin et al. used a convolutional neural network (CNN)-based diagnostic algorithm on SMwI, exhibiting comparable performance to that of an experienced neuroradiologist.[19] Ariz et al. introduced an automatic segmentation model utilizing a 3D atlas-based pipeline on SWI, revealing that relative neuromelanin and iron quantification improved diagnostic performance.[20] Wang et al. used a CNN architecture on true SWI, demonstrating its capability to detect nigrosome-1 abnormalities and localize them.[21] However, these deep learning models have not been used to segment nigrosome-1 and diagnose Parkinson's Disease at the same time. Also, previous studies have demonstrated relatively low diagnostic performance for nigrosome-1 (AUC, 0.87–0.95), and no external validation was conducted except the model by Shin et al, whose data we incorporated into our training dataset. To our knowledge, the diagnostic performance of a fully automated deep learning-based model for simultaneously detecting and quantifying nigrosome-1 abnormality using SMwI and its external validation remains unexplored. Therefore, we propose a fully automated deep-learning model using SMwI to identify and determine abnormalities in the nigrosome-1 region, along with quantifying volume of nigral hyperintensity. This study aimed to develop and validate a deep learning-based automatic quantification for nigral hyperintensity and a classification algorithm for neurodegenerative parkinsonism using SMwI.

MATERIALS AND METHODS

This study follows the guidelines of Checklist for Artificial Intelligence in Medical Imaging (CLAIM)[1] (Supplemental Table 1). For data privacy, data was anonymized before entering the model, and the data entered was not used elsewhere.

Participants

Patients with Parkinsonism symptoms who underwent brain MRI, including SMwI, as well as N-3-fluoropropyl-2-β-carbomethoxy-3-β-(4-iodophenyl) nortropane (^{18}F-FP-CIT) PET imaging were included in the training and external validation datasets.

The training data comprised: 1) 151 subjects (56 patients with IPD and 95 healthy subjects in the control group) collected prospectively from three hospitals between November 2022 and May 2023, and 2) 299 subjects (154 patients with IPD, as well as a control group consisting of 125 patients with drug-induced parkinsonism [DIP] and 20 healthy subjects) from the training dataset of the previous study by Shin et al.[19] Overall, 450 participants (210 with IPD and 240 individuals in the control group) from four hospitals were used for training the detection and segmentation models.

For external validation, data was retrospectively collected from one hospital between July 2021 and January 2022. We excluded healthy subjects from the validation process and included only individuals with symptoms of Parkinsonism, which is crucial for differential diagnosis in identifying IPD. Out of 313 patients presenting with symptoms of parkinsonism and suspected IPD at clinics, eight were excluded due to inappropriate SMwI to interpretate due to image quality, artifact, and scan range. Patients with multiple system atrophy (n = 53), progressive supranuclear palsy (n = 9), and normal pressure hydrocephalus (n = 6) were also excluded. We enrolled 168 patients with IPD as a disease group with abnormality on PET imaging. As a disease control group, we enrolled 69 patients with DIP or ET who had normal PET imaging. Figure 1 shows the flowcharts detailing training and external validation dataset collection.

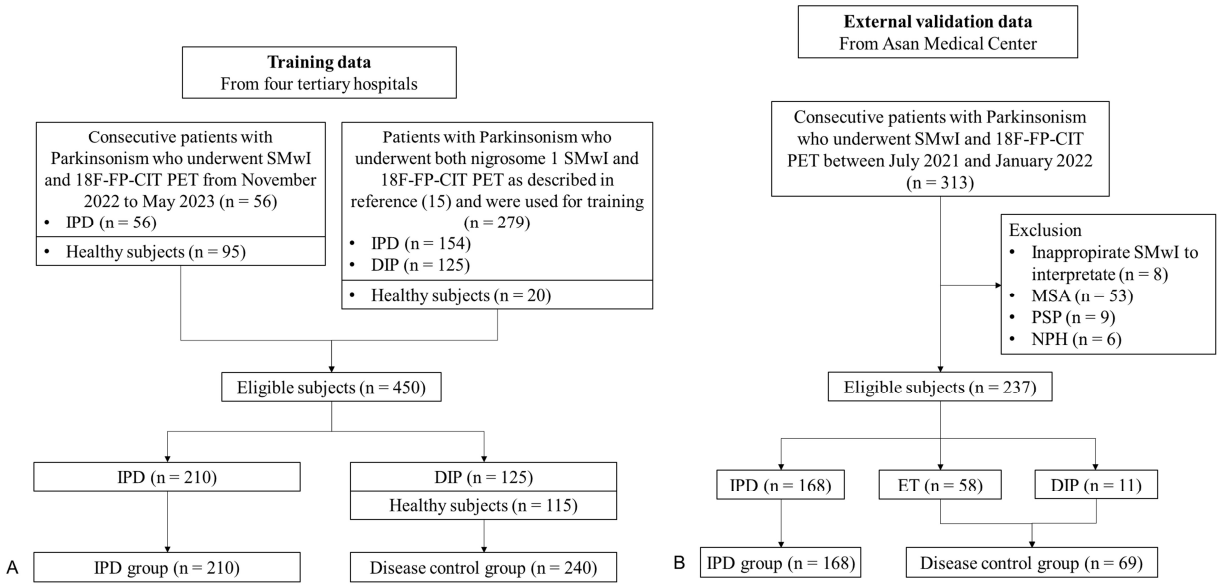


Figure 1. Flowcharts illustrating patient selection for (A) training and (B) validation data.

Abbreviations: IPD, Idiopathic Parkinson's disease; DIP, drug-induced Parkinsonism; ET, essential tremor

MR Imaging Protocol

Oblique coronal imaging was employed owing to the location of the nigrosome-1 region below the inferior pole of the red nucleus.[23] All participants underwent MR imaging on a 3T MRI with a 32-channel coil (Ingenia CX and Ingenia Elition X, Philips Medical Systems, Best, the Netherlands; Skyra, Siemens Healthineers, Forchheim, Germany). The imaging parameters of the 3T scanner are presented at Supplemental Table 2.

SMwI data reconstruction and reslicing

To visualize the nigrosome-1 region, a special sequence and image processing technique was employed called SMwI, derived from QSM.[24] This approach enhances the Swallow-tail sign visualization in the SN region by amplifying the susceptibility contrast resulting from iron.[25] Images were acquired using a 3D GRE 3-echo sequence. QSM, generated using the iterative least-square method, was further processed to create a QSM mask for susceptibility contrast weighting, based on a paramagnetic threshold value.[24,26] The threshold was determined for a optimal contrast-to-noise ratio. The mask was generated by calculating the quantitative susceptibility values of QSM using an equation that utilizes this threshold. SMwI was subsequently created by multiplying the multi-echo combined magnitude images with the QSM mask. The reconstructed SMwI data were resliced in 0.5 mm increments using the Antspy open-source medical imaging library (version 0.3.8, https://antspy.readthedocs.io/).

Development of deep learning-based automatic quantification for nigral hyperintensity

For SN segmentation and nigral hyperintensity quantification, we employed Heuron NI (version 1.0.0.19, Heuron, Seoul, Korea), which was commercially launched in the Republic of Korea in June 2021. This system is designed as a nigral hyperintensity segmentation model to evaluate quantitative information. It utilizes SMwI data as input and provides the measured volume of nigral hyperintensity as output (Figure 2).

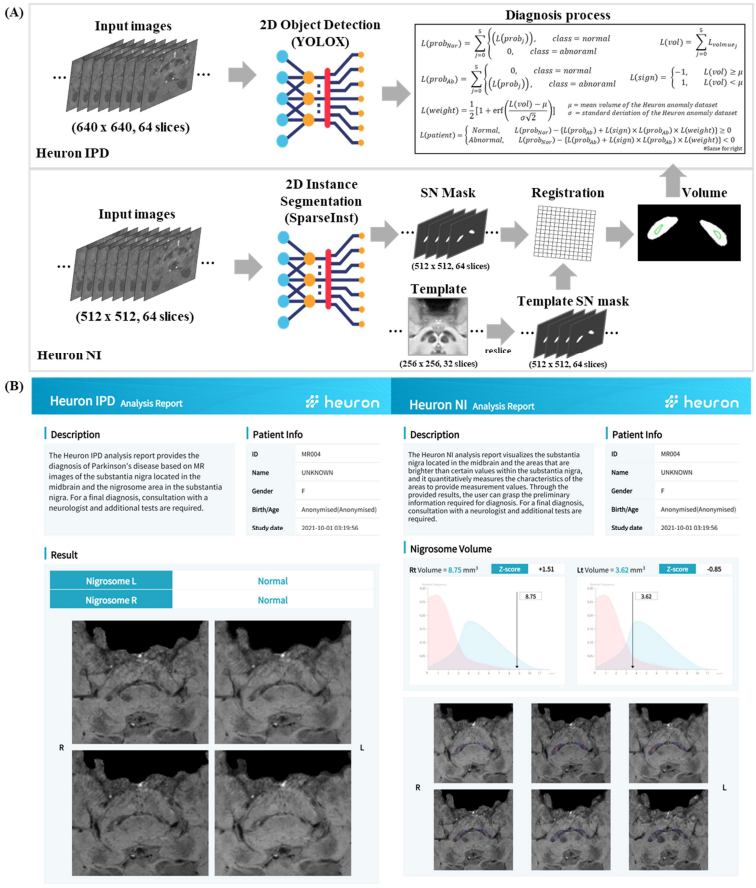


Figure 2. Automatic nigrostriatal degeneration classification and nigral hyperintensity volume measurement using deep learning models. (A) Classification progress of nigrostriatal degeneration and measurement of nigral hyperintensity volume. (B) Output from the Heuron IPD and Heuron NI systems. Heuron IPD shows the analysis results for nigrostriatal degeneration, while Heuron NI depicts nigral hyperintensity volume measurement results. The classification of the Heuron dataset graph is depicted in normal (blue) and abnormal (red) categories.

Abbreviations: IPD, Idiopathic Parkinson's disease; CNN, convolutional neural network; SN, substantia nigra

Initially, the input images underwent preprocessing, consists of reslicing and resizing to produce 64 slices of 512*512 2D SMwI images. These preprocessed images were then used as input for the 2D instance segmentation model, SparseInst,[27] to segment the SN region. We used SN mask-based image registration for nigral hyperintensity quantification. The SN segmentation model generates predicted SN mask information for each slice and binarized it by excluding background or extraneous data from the image to extract the hyperintensity region. In this process, input and target data for registration were generated. Inputdata was defined by binarized SN mask. Target data was defined by the SMwI template[8] SN mask. An experienced neurologist delineated the SN and nigral hyperintensity in the SMwI template and annotated hyperintensity region as 'processed nigrosome-1'. After registering the input data (binarized SN mask) and the target data (template SN mask), the overlapping area between the hyperintensity region in the binarized SN mask and processed nigrosome-1 region in the template SN mask was designated as 'nigrosome-1'. To account for interindividual variation, normalization was conducted using the intracranial volume, calculated using SynthSeg.[28] The details of development and progress of our deep-learning model is demonstrated in Supplemental materials.

Development of deep learning-based automatic classification for nigral hyperintensity

Heuron IPD, a deep learning-based system (version 1.0.1.10, Heuron Corporation, Seoul, Korea), has been commercially available in Korea since April 2021. This system is designed to detect nigrostriatal degeneration and includes a CNN-based model for object detection and segmentation model, integrating outputs of each model to classify nigrostriatal degeneration (Fig 2).

Initially, the input images underwent preprocessing to produce 64 slices of 640*640 2D SMwI images. Heuron IPD first performed object detection using YOLOX model,[29] utilizing the preprocessed images to predict the nigrosome-1 region and classify abnormalities. The model was trained on a dataset annotated with nigrosome-1 location by an experienced neurologist, with ^{18}F-FP-CIT PET imaging serving as the ground truth for the presence of nigrostriatal degeneration.

For the classification of nigrostriatal degeneration, five slices from the scanned MR images containing SN and nigral hyperintensity structures were utilized for analysis.[19,25] The object detection model returned a probability value indicating the likelihood of a specific object based on its training. The quantification information of nigral hyperintensity volume, measured by Heuron NI, informed the object

detection model and acted as a weight for the class probability value. These weighted class probability values were employed to determine the abnormalities in the nigral hyperintensity region. The details of development and progress of our deep-learning model is demonstrated in Supplemental materials. The diagnostic performance of Heuron IPD was evaluated and compared with the model using only the CNN, without volume weight, to determine the effectiveness of volume weighting.

Evaluation of diagnostic performance

18F FP-CIT PET (Biograph TruePoint; Siemens, Erlangen, Germany) was performed in all subjects and served as the reference standard to determine whether the patient had IPD, with findings assessed by experienced nuclear medicine physicians. The performance of both Heuron IPD and Heuron NI was evaluated per SN and participants. The performance of Heuron IPD was assessed by classifying IPD based on the reference standard. Although Heuron NI was employed to quantify the nigral hyperintensity and provide volume measurement for the left and right sides, its diagnostic performance for diagnosing IPD was also evaluated using the cut-off value as a threshold. The cut-off value was calculated based on the highest AUC from the ROC curve of the volume and reference standard. For the per participant analysis, a participant was considered to exhibit an abnormality if abnormalities were on either side of the SN.

Statistical analysis

Determining diagnostic accuracy involved calculating several metrics, including AUC, sensitivity, specificity, F1-score, and accuracy for each result set. Furthermore, 95% CIs were determined using Clopper–Pearson intervals. Correlations with Hoehn and Yahr (H&Y) Scale were analyzed using Kruskal–Wallis, Mann–Whitney, and Hanley&McNeil tests. $p<0.05$ was considered statistically significant. All statistical analyses were conducted using MedCalc version 22.007 (MedCalc Software, Ostend, Belgium).

RESULTS

Participant enrolment

Overall, 237 participants were enrolled for external validation of the performance of the model. The external validation data included 168 patients with IPD, as well as a control group consisting of 58 with essential tremor and 11 with drug-induced Parkinsonism. Demographics of external validation data are shown in Table 1 and training data in Supplemental Table 3.

Table 1. Demographic findings and clinical characteristics of external validation data.

	IPD (n=168)	Control (n=69)	P-value
Age, yr	66 ± 10	66 ± 12	.659*
Sex, male:female	79:89	26:43	.189†
Disease duration, yr	5.7 ± 5.9	4.0 ± 6.2	.507*
H&Y stage	2.4 ± 0.8	2.0 ± 0.6	.004*
Disease involvement			
Right	4	0	
Left	2	0	
Both	162	0	

*Comparison of two groups using Mann-Whitney test, †Comparison of two groups using chi-square test.

Diagnostic performance of Heuron NI and Heuron IPD per SN and participant

The diagnostic performance of Heuron NI was assessed by comparing the volume measured through Heuron NI and the reference standard for all participants using the AUC. The AUC were 0.915 (95% CI, 0.872–0.947) and 0.928 (95% CI, 0.887–0.957) on the left and right, respectively (Figure 3). No significant difference in diagnostic performance was observed between the left (sensitivity 87.8%, specificity 83.6%) and right (sensitivity 89.2%, specificity 85.9%) directions ($p = .65$). The diagnostic performance per participant exhibited 96.4% sensitivity, 73.9% specificity, and 93.1% F1-score. The same analysis was repeated by dividing the measured volume by the ICV **(Online Supplemental Data)**, yielding nearly identical diagnostic performance as before dividing the volume by ICV.

When using only the CNN model in Heuron IPD without incorporating volume weight from quantification information, the AUC was 0.942 (95% CI 0.904–0.968) and 0.933 (95% CI 0.893–0.961) on the left on right, respectively **(Online Supplemental Data)**. Using Heuron IPD, applying weighted class probability values from quantification information, the AUC increased to 0.967 (95% CI, 0.881–0.991) and 0.976 (95% CI, 0.948–0.992) on the left and right, respectively (Figure 3). Consistent with the volume-reference standard analysis, no significant difference was observed in the diagnostic performance between the left (sensitivity 95.1%, specificity 95.9%) and right (sensitivity 94.6%, specificity 95.8%) directions ($p=0.60$). The per-participant diagnostic performance was as follows: sensitivity 97.0%, specificity 95.7%, and F1-score 97.6% **(Online Supplemental Data)**. The representative cases of Heuron NI and Heuron IPD for both normal and IPD patient are shown in Figure 4 and Figure 5, respectively.

Heuron IPD also demonstrated significant performance in differentiating IPD from ET (AUC, 0.965–0.980) and DIP (AUC, 0.959–0.968) (Supplemental Table 4, 5). Additionally, there was no significant performance difference by patient gender (Supplemental Table 6, 7).

Heuron IPD also demonstrated significant performance in differentiating IPD from ET (AUC, 0.965–0.980) and DIP (AUC, 0.959–0.968) (Supplemental Table 4, 5). Additionally, there was no significant performance difference by patient gender (Supplemental Table 6, 7).

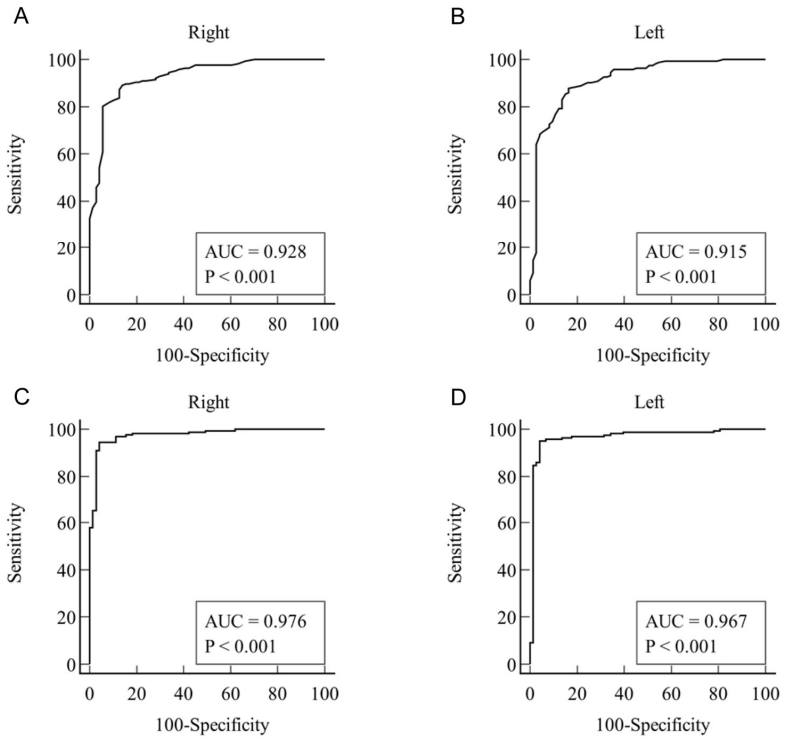


Figure 3. ROC curves for each nigrosome volume and Heuron IPD for PET visual assessment. ROC curves for the volume of nigrosome and PET for the right (A) and left (B) SN. ROC curves for the classification results of Heuron IPD and PET for the right (C) and left (D) SN.

Abbreviation: IPD, Idiopathic Parkinson's disease; SN, substantia nigra; PET, positron emission tomography

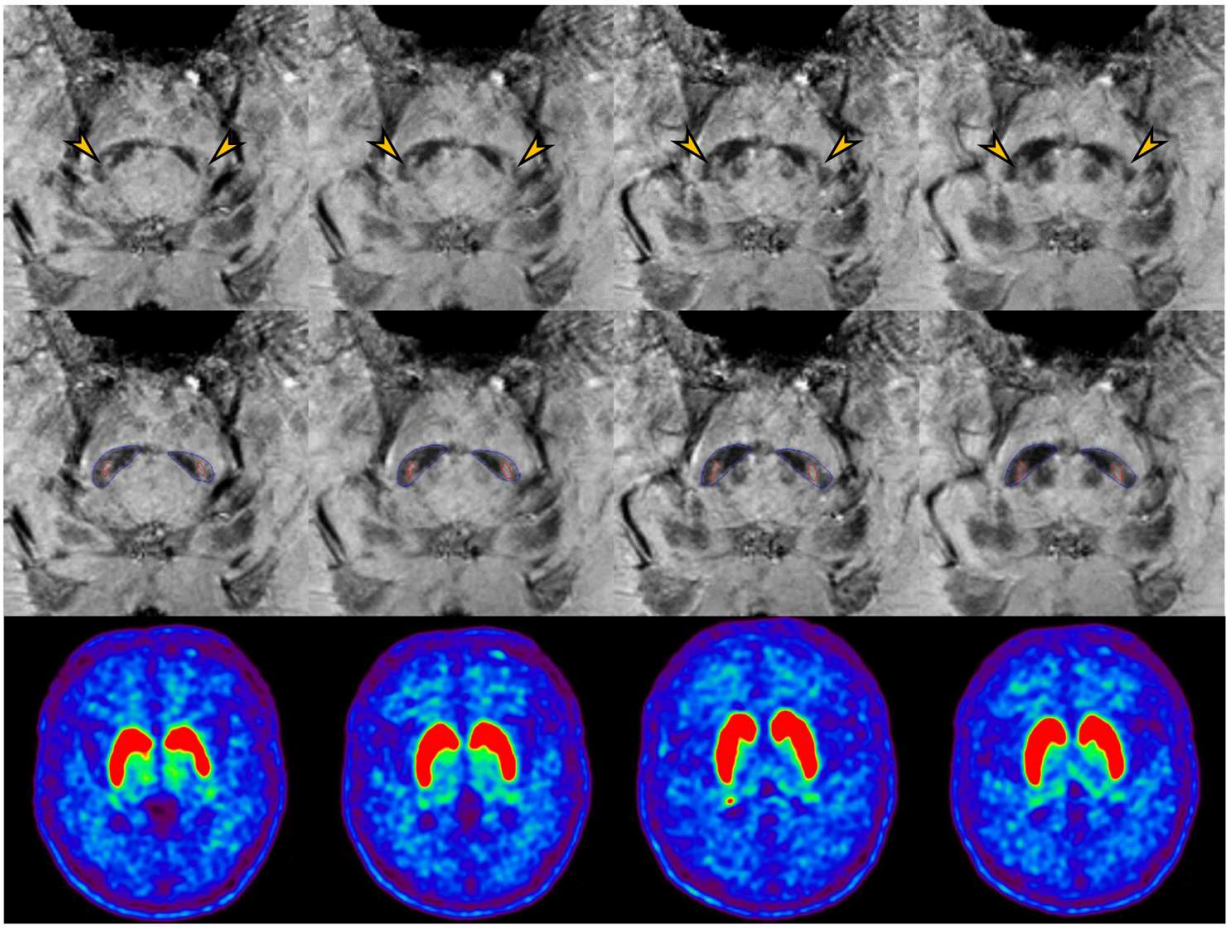


Figure 4. A 69-year-old female patient with essential tremor, showing nigral hyperintensity within the substantia nigra (arrow). The analysis results (middle row) display the segmentation of the substantia nigra (blue) and nigral hyperintensity (red). The nigral hyperintensity volume is 10.13 mm^3 on the left and 9 mm^3 on the right. ^{18}F-FP-CIT imaging (lower row) shows normal dopamine transporter binding in the bilateral basal ganglia.

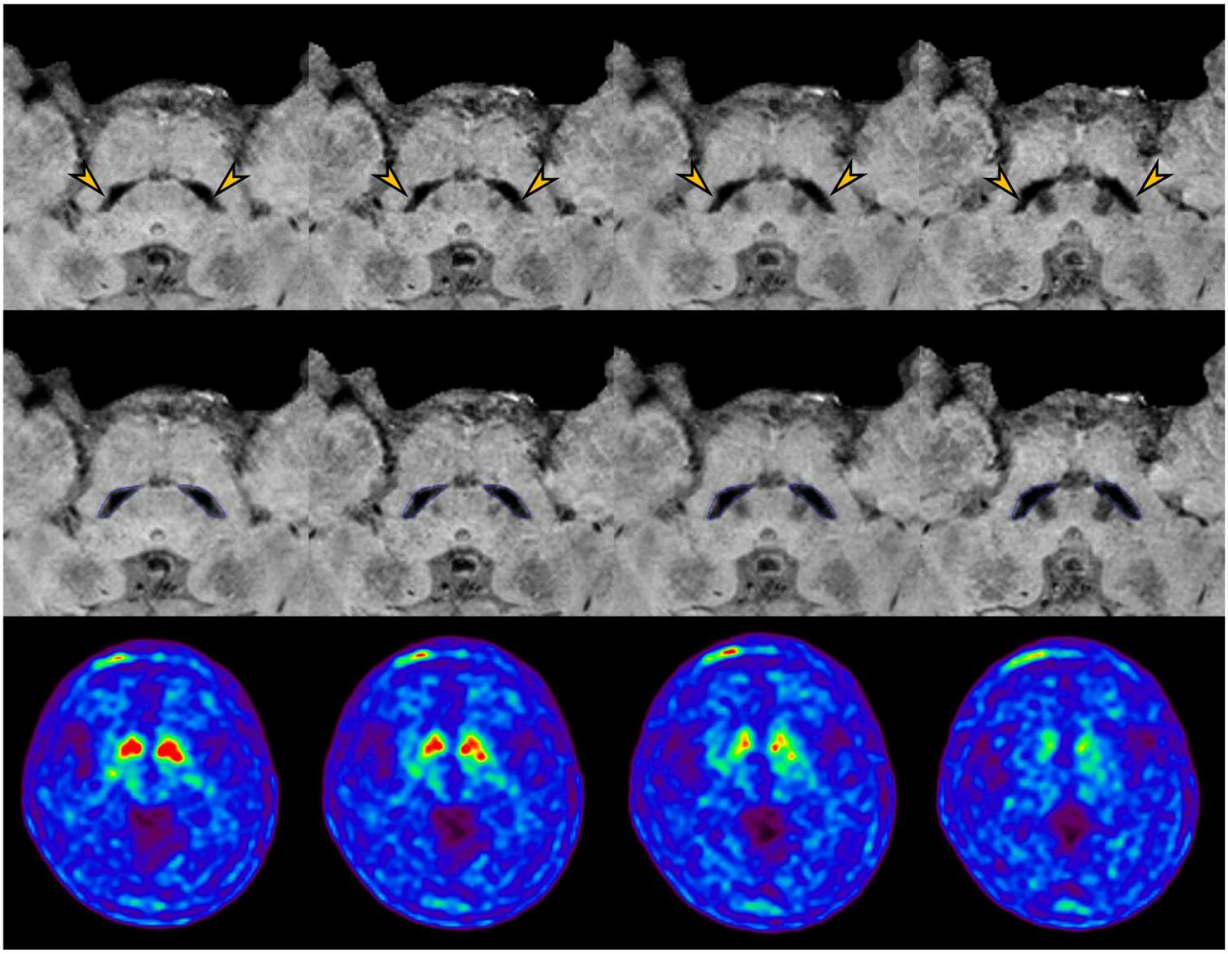


Figure 5. A 63-year-old female patient with idiopathic Parkinson's disease was found to have abnormal nigrosome-1 regions (arrow) on both sides. The analysis results (middle row) display the segmentation of the substantia nigra (blue) and nigral hyperintensity (red). The nigral hyperintensity volume is 0 mm^3 on the left and 0.13 mm^3 on the right. ^{18}F-FP-CIT imaging (lower row) shows decreased dopamine transporter binding in the bilateral basal ganglia

Diagnostic performance of Heuron IPD across H&Y score groups

We compared diagnostic performance in different groups of H&Y score in patients with IPD. H&Y scores were categorized into two distinct groups. Group 1 comprised four categories: 1 (1–1.5), 2 (2–2.5), 3 (3–3.5), and 4 (4–5), while group 2 consisted of two categories: 1 (1–2.5) and 2 (3–5). Both group 1 and 2 showed higher AUC in lower H&Y stage group. The AUC was 1.000 in H&Y score 1 to 1.5 in group 1, and 0.957 in H&Y score 1 to 2.5 in group 2 **(Online Supplemental Data)**. However, the diagnostic performance within each group revealed no significant statistical difference (Group 1: category 1 vs. category 2, p = 0.11; category 1 vs. category 3, p = 0.29; category 2 vs. category 3, p = 0.61; Group 2: category 1 vs. category 2, p = 0.55).

Volume-H&Y score correlation using Heuron NI

We assessed the correlation between the H&Y stage and volume in patients with IPD. We evaluated the relationship between the quantitative information of the nigral hyperintensity volume measured by Heuron NI and disease progression. In the analysis (Supplemental Table 8 and Supplemental Figure 1), group 1 showed no statistical correlation between H&Y stages and volume (p = .06). However, group 2 exhibited a significant correlation (p = .008). This pattern remained consistent even after adjusting the volume for intracranial volume (group 1, p= .07; group 2, p = .01).

DISCUSSION

We developed and validated a deep learning-based model for the automatic quantification of nigral hyperintensity (Heuron NI) and diagnosis of IPD (Heuron IPD). Both Heuron IPD and Heuron NI demonstrated successful diagnosis of IPD in patients with clinical suspicion of Parkinsonism, achieving an AUC of 0.915–0.976 per SN and accuracy of 89.9–96.6% per patient. Quantifying nigral hyperintensity, H&Y stage 3–5 showed significantly lower nigral hyperintensity volume compared to H&Y stage 1–2.5, suggesting more severe motor symptoms.

Nigrosome-1 is affected in early-stage IPD, and its loss serves as an imaging biomarker for early IPD diagnosis.[8,30,31] Previous studies

predominantly employed manual drawing of region of interest for nigrosome-1 and performed voxelwise or quantitative analyses.[8,30,32] Ariz et al. recently introduced the first fully automated segmentation model for quantifying nigrosome-1.[20] Their segmentation model relied on a 3D atlas-based pipeline,[33] generating two atlases from neuromelanin MRI and SWI images. However, their model requires atlas reconstruction and it is difficult to utilize SMwI. Creating an atlas from SMwI poses challenges owing to inconsistent imaging slabs. Consequently, our model initially segmented the SN and created an SN mask from the input images, then it quantified nigral hyperintensity within the segmented SN area using the SMwI template. Shin et al. introduced an automated nigrosome-1 identification method, but it did not provide quantification.[19] Our model had advancements in simultaneously detecting and quantifying nigrosome-1 abnormality. In addition, our approach differs from previous studies as we evaluated diagnostic performance in patients with clinical suspicion of Parkinsonism rather than comparing it with a normal control group. Our model exhibited an AUC ranging from 0.915 to 0.928 for diagnosing IPD based on nigral hyperintensity volume, suggesting significant diagnostic utility in clinical practice. However, a small number of false positives and false negatives could lead to misinterpretation. Therefore, deep-learning models should play a supportive role and be used in conjunction with comprehensive clinical and radiological evaluations when diagnosing IPD.

Loss of nigral hyperintensity resulting from dopaminergic cell degeneration and iron deposition,[34,35] suggesting a potential association between nigrosome-1 and IPD symptom severity. A negative correlation has been observed between the UPDRS part III score and signal-to-noise ratio of neuromelanin in the posterolateral SN [36] or nigrosome-1 volume.[37] Hence, we evaluated the correlation between nigral hyperintensity volume and H&Y stage, revealing that the H&Y stage 1–2.5 group showed significantly higher nigral hyperintensity volume than the H&Y stage 3–5 group. Advanced-staged IPD, characterized by an H&Y stage of ≥3, patients typically exhibit a lower quality of life,[38] an increased risk of medication-refractory disease progression,[39] and a shorter overall survival time.[40] Therefore, accurately quantifying nigral hyperintensity holds clinical value in assessing patient outcomes. Furthermore, automated quantification can possibly be utilized as an imaging biomarker for monitoring the efficacy of disease-modifying drugs, and further studies are warranted.

We developed a model (Heuron IPD) using a deep learning-based algorithm called YOLOX for classifying nigral hyperintensity, as normal or abnormal. The YOLOX model has been improved with a decoupled-head approach, optimizing layer arrangement to enhance classification and localization.[19] Moreover, YOLOX employs an anchor-free design facilitated by the multi-positive approach, enhancing its flexibility and performance through structural modifications. This modification eliminates the need for heuristic tuning associated with anchor optimization while concurrently diminishing computational requirements. Consequently, our model (AUC: 0.967–0.976, per SN) outperformed the diagnostic performance of the previous model based on YOLOv3 (AUC: 0.921–0.937) and even visual interpretation by a neuroradiologist (AUC: 0.937–0.945).[19] It also showed higher performance than other pre-existing deep-learning model such as Pvtv2[41] and Detectors[42] (Supplement Table 9). Our study demonstrated the advancement of the YOLO series and its practical utility in clinical practice. Additionally, Heuron IPD showed enhanced diagnostic performance following volume-weighting application to the class probability value. In patients with IPD, the volume of the nigral hyperintensity is notably small or even absent. Therefore, comprehensively analyzing SN and nigral hyperintensity volume is essential for improving diagnostic accuracy.

This study has some limitations. First, we opted for 18F-FP-CIT PET imaging as the reference standard for diagnosing IPD. Consequently, atypical Parkinsonism, including MSA and PSP, which can also exhibit decreased dopamine transporter binding similar to IPD, were excluded. Therefore, when interpreting the quantification and classification of nigral hyperintensity in a clinical context using our model, caution is advised, and a comprehensive assessment of other clinical features is essential. Second, we assessed nigral hyperintensity, which was considered as nigrosome-1 region. However, these hyperintensity areas were not pathologically confirmed and may not be definitely correlated with nigrosome-1. Third, we did not evaluate the use of SWI or neuromelanin imaging as input for our model. There is also a limitation in using SMwI reconstructed from QSM, as it is not yet a widely generalized method. However, SMwI is probably more suitable for diagnosing IPD due to its higher signal-to-noise ratio and contrast-to-noise ratio compared to SWI imaging, and its more distinct margins than neuromelanin imaging.[7,14] Lastly, this is a retrospective study and external validation was limited to data from a single hospital, which may not represent demographic diversity. Furthermore, the use of a limited number of MRI scanners—Ingenia CX, Ingenia Elition X, and Skyra—along with the need for high-resolution images to visualize the “swallow tail sign” may limit the generalizability of our model, necessitating a real-world validation study.

CONCLUSIONS

In conclusion, our deep learning-based model proves valuable for rapid and accurate automatic quantification of nigral hyperintensity, facilitating IPD diagnosis and symptom severity prediction.

ACKNOWLEDGMENTS

This research was supported by a grant of the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare, Republic of Korea (grant number: RS-2023-00266948).

REFERENCES

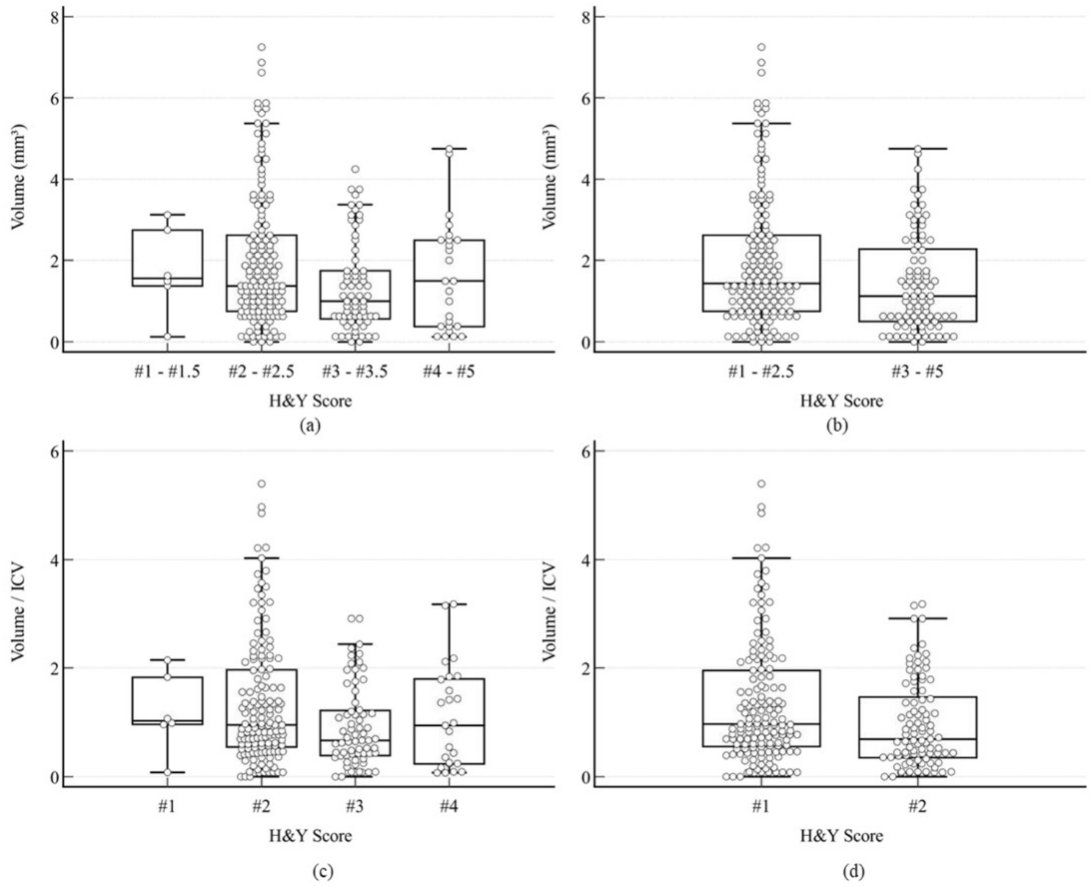
1. Lang AE, Lozano AM. Parkinson's disease. First of two parts. N Engl J Med 1998;339:1044-53
2. Blazejewska AI, Schwarz ST, Pitiot A, et al. Visualization of nigrosome 1 and its loss in PD: pathoanatomical correlation and in vivo 7 T MRI. Neurology 2013;81:534-40

3. Damier P, Hirsch EC, Agid Y, et al. The substantia nigra of the human brain. I. Nigrosomes and the nigral matrix, a compartmental organization based on calbindin D(28K) immunohistochemistry. Brain 1999;122 (Pt 8):1421-36
4. Schwarz ST, Afzal M, Morgan PS, et al. 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
5. Cosottini M, Frosini D, Pesaresi I, et al. MR imaging of the substantia nigra at 7 T enables diagnosis of Parkinson disease. Radiology 2014;271:831-8
6. Cosottini M, Frosini D, Pesaresi I, et al. Comparison of 3T and 7T susceptibility-weighted angiography of the substantia nigra in diagnosing Parkinson disease. AJNR Am J Neuroradiol 2015;36:461-6
7. Lee DH, Heo H, Suh CH, et al. Improved diagnostic performance of susceptibility-weighted imaging with compressed sensing-sensitivity encoding and neuromelanin-sensitive MRI for Parkinson's disease and atypical Parkinsonism. Clin Radiol 2024;79:e102-e11
8. Sung YH, Noh Y, Kim EY. Early-stage Parkinson's disease: Abnormal nigrosome 1 and 2 revealed by a voxelwise analysis of neuromelanin-sensitive MRI. Hum Brain Mapp 2021;42:2823-32
9. Kim PH, Lee DH, Suh CH, et al. Diagnostic performance of loss of nigral hyperintensity on susceptibility-weighted imaging in parkinsonism: an updated meta-analysis. Eur Radiol 2021;31:6342-52
10. Cheng Z, He N, Huang P, 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
11. Kim HK, Kim T, Baek MS, et al. Nigrosome 1 visibility and its association with nigrostriatal dopaminergic loss in Parkinson's disease. Eur J Neurol 2023;30:1639-47
12. Sung YH, Kim JS, Yoo SW, et al. A prospective multi-centre study of susceptibility map-weighted MRI for the diagnosis of neurodegenerative parkinsonism. Eur Radiol 2022;32:3597-608
13. Bae YJ, Song YS, Choi BS, et al. Comparison of susceptibility-weighted imaging and susceptibility map-weighted imaging for the diagnosis of Parkinsonism with nigral hyperintensity. Eur J Radiol 2021;134:109398
14. Lee S, Suh CH, Jo S, et al. Comparative Performance of Susceptibility Map-Weighted MRI According to the Acquisition Planes in the Diagnosis of Neurodegenerative Parkinsonism. Korean J Radiol 2024;25:267-76
15. Suh CH, Shim WH, Kim SJ, et al. Development and Validation of a Deep Learning-Based Automatic Brain Segmentation and Classification Algorithm for Alzheimer Disease Using 3D T1-Weighted Volumetric Images. AJNR Am J Neuroradiol 2020;41:2227-34
16. Doss DJ, Johnson GW, Narasimhan S, et al. Deep Learning Segmentation of the Nucleus Basalis of Meynert on 3T MRI. AJNR Am J Neuroradiol 2023;44:1020-25
17. Fan S, Ponisio MR, Xiao P, et al. AmyloidPETNet: Classification of Amyloid Positivity in Brain PET Imaging Using End-to-End Deep Learning. Radiology 2024;311:e231442
18. Chen H, Liu X, Luo X, et al. An automated hybrid approach via deep learning and radiomics focused on the midbrain and substantia nigra to detect early-stage Parkinson's disease. Front Aging Neurosci 2024;16:1397896
19. Shin DH, Heo H, Song S, 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
20. Ariz M, Martinez M, Alvarez I, et al. Automatic Segmentation and Quantification of Nigrosome-1 Neuromelanin and Iron in MRI: A Candidate Biomarker for Parkinson's Disease. J Magn Reson Imaging 2023
21. Wang C, He N, Zhang Y, et al. Enhancing Nigrosome-1 Sign Identification via Interpretable AI using True Susceptibility Weighted Imaging. J Magn Reson Imaging 2024
22. Tejani AS, Klontzas ME, Gatti AA, et al. Checklist for Artificial Intelligence in Medical Imaging (CLAIM): 2024 Update. Radiol Artif Intell 2024;6:e240300
23. Sung YH, Noh Y, Lee J, et al. Drug-induced Parkinsonism versus Idiopathic Parkinson Disease: Utility of Nigrosome 1 with 3-T Imaging. Radiology 2016;279:849-58
24. Nam Y, Gho SM, Kim DH, et al. Imaging of nigrosome 1 in substantia nigra at 3T using multiecho susceptibility map-weighted imaging (SMWI). J Magn Reson Imaging 2017;46:528-36
25. Sung YH, Lee J, Nam Y, et al. Differential involvement of nigral subregions in idiopathic parkinson's disease. Hum Brain Mapp 2018;39:542-53
26. Gho SM, Liu C, Li W, et al. Susceptibility map-weighted imaging (SMWI) for neuroimaging. Magn Reson Med 2014;72:337-46
27. Cheng T, Wang X, Chen S, et al. Sparse instance activation for real-time instance segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022
28. Billot B, Greve DN, Puonti O, et al. SynthSeg: Segmentation of brain MRI scans of any contrast and resolution without retraining. Med Image Anal 2023;86:102789
29. Ge Z, Liu S, Wang F, et al. Yolox: Exceeding yolo series in 2021. arXiv preprint arXiv:2107.08430 2021
30. Kim EY, Sung YH, Shin HG, et al. 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 2018;14:90-97
31. He N, Ghassaban K, Huang P, et al. Imaging iron and neuromelanin simultaneously using a single 3D gradient echo magnetization transfer sequence: Combining neuromelanin, iron and the nigrosome-1 sign as complementary imaging biomarkers in early stage Parkinson's disease. Neuroimage 2021;230:117810
32. Chau MT, Agzarian M, Wilcox RA, et al. Simple quantitative planimetric measurement of nigrosome-1 for clinical settings. J Neurol Sci 2023;454:120857
33. Ariz M, Abad RC, Castellanos G, et al. Dynamic Atlas-Based Segmentation and Quantification of Neuromelanin-Rich Brainstem Structures in Parkinson Disease. IEEE Trans Med Imaging 2019;38:813-23
34. Bae YJ, Kim JM, Kim E, et al. Loss of Nigral Hyperintensity on 3 Tesla MRI of Parkinsonism: Comparison With (123) I-FP-CIT SPECT. Mov Disord 2016;31:684-92
35. Bae YJ, Kim JM, Sohn CH, et al. Imaging the Substantia Nigra in Parkinson Disease and Other Parkinsonian Syndromes. Radiology 2021;300:260-78
36. Biondetti E, Gaurav R, Yahia-Cherif L, et al. Spatiotemporal changes in substantia nigra neuromelanin content in Parkinson's disease. Brain 2020;143:2757-70
37. Chu YT, Yu CF, Fan SP, et al. Substantia nigra nigrosome-1 imaging correlates with the severity of motor symptoms in Parkinson's disease. J Neurol Sci 2023;451:120731
38. Evans JR, Mason SL, Williams-Gray CH, et al. The natural history of treated Parkinson's disease in an incident, community based cohort. J Neurol Neurosurg Psychiatry 2011;82:1112-8

39. Goetz CG, Stebbins GT, Blasucci LM. Differential progression of motor impairment in levodopa-treated Parkinson's disease. Mov Disord 2000;15:479-84
40. Roos RA, Jongen JC, van der Velde EA. Clinical course of patients with idiopathic Parkinson's disease. Mov Disord 1996;11:236-42
41. Wang W, Xie E, Li X, et al. Pyramid vision transformer: A versatile backbone for dense prediction without convolutions. In: Proceedings of the IEEE/CVF international conference on computer vision, 2021
42. Qiao S, Chen L-C, Yuille A. Detectors: Detecting objects with recursive feature pyramid and switchable atrous convolution. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2021

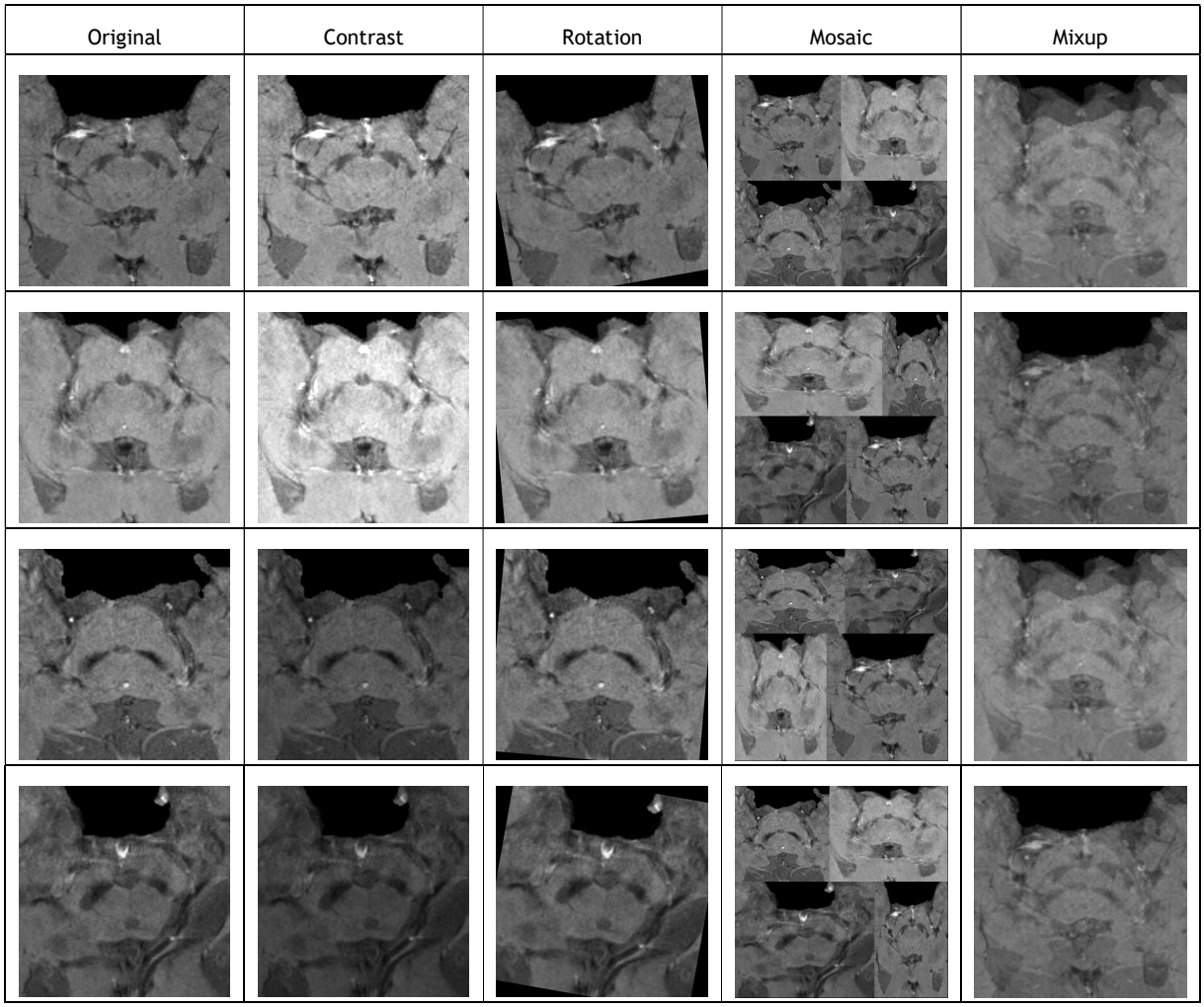
SUPPLEMENTAL FILES

Supplemental Figure 1. Volume analysis results for the two groups. (a) and (b) show the analysis results between the volume and the two groups while (c) and (d) depict the analysis results between the volume/ICV and the two groups. The normalized nigrosome volume was calculated using the following formula: total nigrosome volume (mm^3) / total intracranial volume (mm^3) × 1,000,000.



Abbreviation: ICV, intracranial volume

Supplementary Figure 2. Examples demonstrating the results of image augmentation on the input image. In the Yolox model, multi-scale(480 ~ 800, 32 range), mosaic, mixup, color jitter, random flip, and resize were used. In the SparseInst model, multi-scale(416 ~ 608, 32 range), random brightness, contrast, flip, and rotation(± 10°) were utilized. Examples of the main augmentation methods used are shown in the figure below.



Online Supplemental Data. Diagnostic performance of Heuron NI and Heuron IPD in all participants

	Sensitivity, % [95 % CI]	Specificity, % [95 % CI]	Accuracy, % [95 % CI]	F1-score, % [95 % CI]	AUC [95 % CI]
Heuron NI					
Right	89.2 [83.4, 93.5]	85.9 [75.6, 93.0]	88.2 [83.4, 92.0]	91.4 [87.0, 94.6]	0.928 [0.887, 0.957]
Left	87.8 [81.8, 92.4]	83.6 [73.1, 91.2]	86.5 [81.5 90.6]	90.0 [85.5, 93.5]	0.915 [0.872, 0.947]
Participant	96.4 [92.4, 98.7]	73.9 [69.9, 83.8]	89.9 [85.3, 93.4]	93.1 [89.1, 96.0]	
Heuron NI / ICV					
Right	88.0 [82.0, 92.5]	88.7 [79.0, 95.0]	88.2 [83.4, 92.0]	91.3 [86.9, 94.5]	0.928 [0.887, 0.957]
Left	87.8 [81.8, 92.3]	83.6 [73.1, 91.2]	86.5 [81.5, 90.6]	90.0 [85.5, 93.5]	0.915 [0.872, 0.947]
Participant	96.4 [92.4, 98.7]	75.4 [63.5, 85.0]	90.3 [85.8, 93.8]	93.4 [89.4, 96.2]	
Heuron IPD (only CNN)					
Right	92.2 [87.0, 95.8]	94.4 [86.2, 98.4]	92.8 [88.8, 95.8]	94.7 [91.1, 97.2]	0.933 [0.893, 0.961]
Left	93.9 [89.1, 97.0]	94.5 [86.6, 98.5]	94.1 [92.2, 97.9]	95.7 [90.3, 96.7]	0.942 [0.904, 0.968]
Participant	97.0 [93.2, 99.0]	91.3 [82.0, 967]	95.4 [93.6, 98.6]	96.7 [91.9, 97.7]	
Heuron IPD					
Right	94.6 [90.0, 97.5]	95.8 [88.1, 99.1]	94.9 [91.3, 97.4]	96.3 [93.1, 98.3]	0.976 [0.948, 0.992]
Left	95.1 [90.6, 97.9]	95.9 [88.5, 99.1]	95.4 [91.9, 97.7]	96.6 [93.4, 98.5]	0.967 [0.936, 0.986]
Participant	97.0 [93.2, 99.0]	95.7 [87.8, 99.1]	96.6 [93.5, 98.5]	97.6 [94.8, 99.2]	

Online Supplemental Data. Diagnostic performance of Heuron IPD across different groups

H&Y score	Sensitivity, % [95 % CI]	Specificity, % [95 % CI]	Accuracy, % [95 % CI]	F1-score, % [95 % CI]	AUC [95 % CI]
Group 1					
#1 (1~1.5)	100 [79.4, 100]	100 [63.1, 100]	100 [85.8, 100]	100 [85.8, 100]	1.000 [0.858–1.000]

#2 (2~2.5)	97.3 [90.5, 99.6]	91.3 [72.0, 98.9]	95.8 [89.7, 98.8]	97.3 [91.7, 99.5]	0.943 [0.876-0.980]
#3 (3~3.5)	97.1 [84.7, 99.9]	80.0 [28.4, 99.5]	94.9 [82.7, 99.3]	97.1 [85.9, 99.9]	0.885 [0.742-0.964]
#4 (4~5)	100 [75.3, 100]	-	100 [75.3, 100]	100 [75.3, 100]	-
Group 2					
#1 (1~2.5)	97.8 [92.1 - 99.7]	93.6 [78.6, 99.2]	96.7 [91.7, 99.1]	97.8 [93.2, 99.6]	0.957 [0.903, 0.985]
#2 (3~5)	97.9 [88.7, 100]	80.0 [28.4, 99.5]	96.2 [86.8, 99.5]	97.9 [89.4, 99.9]	0.889 [0.772, 0.959]

1. Tejani AS, Klontzas ME, Gatti AA, et al. **Checklist for Artificial Intelligence in Medical Imaging (CLAIM): 2024 Update.** *Radiol Artif Intell* 2024;6:e240300. DOI: https://doi.org/10.1148/ryai.240300

Supplemental materials

Development of deep learning-based automatic quantification for nigral hyperintensity

For SN segmentation and nigral hyperintensity quantification, we used Heuron NI (version 1.0.0.19, Heuron, Seoul, Korea), launched commercially in Republic of Korea in June 2021. The process leading to clinical validation and regulatory approval includes pilot studies in clinical settings and user feedback. The Heuron NI system comprises a segmentation model, utilizing SMwI data as input and providing the measured volume of the nigral hyperintensity as output (Fig 2).

This nigral hyperintensity segmentation model was developed to evaluate quantitative information effects on the nigral hyperintensity area on classification and disease progression. Heuron NI system performs segmentation of the SN, followed by quantification of the nigrosome-1 within the segmented SN area. The 2D instance segmentation model SparseInst[1] was used to segment the SN region, and the training data annotation was generated by an experienced neurologist. This model, built with a fully convolutional encoder-decoder structure, comprised a backbone, context-encoder, and decoder, generating an instance activation map (IAM). ResNet[2] served as the backbone to extract multi-scale features from the input image. These context-encoders processed these features, yielding single-level features enriched with contextual information. The decoder takes the output of the encoder as input and generates an IAM and a mask feature. The IAM highlighted discriminative regions associated with objects and enhanced instance-level feature representation by grouping multiple IAMs for a single object. This was accomplished using the Hungarian algorithm[3] and bipartite matching during the training phase. Each target was assigned one IAM through bipartite matching, ensuring IAM singular activation for each object highlighting more discriminative regions. This obviated the need for a separate non-maximum suppression algorithm in post-processing, enhancing computational efficiency. The preprocessing consists of reslicing and resizing, resulting in 512*512 2D SMwI images. These preprocessed images are used as input for the instance segmentation model, which predicts SN mask information. The model was trained using an AdamW optimizer with a learning rate of 5e-5 and batch size of 16. Multi-scale training, random brightness, contrast, horizontal flip, and rotation were used for data augmentation techniques. The loss functions utilized for training include focal,[4] dice, and binary cross-entropy losses.

We performed SN mask-based image registration for nigral hyperintensity quantification. This method prioritized SN mask information, excluding background or extraneous data from the image. The SN segmentation model generates predicted SN mask information for each slice. This data serves as the moving data for registration, with the fixed data defined by the SMwI template[5] SN mask. Annotations were generated by an experienced neurologist, who delineated the SN and nigral hyperintensity in the SMwI template. Registration between these datasets was conducted using the SyNRA registration from Antspy. Upon completing the registration, the inverse registration field was used to transform the nigrosome-1 label information in the SMwI template from fixed to moving space. These areas are defined as the processed nigrosome-1. The SN region pixel values were binarized to extract the hyper-intensity region, defining it as the binarized SN mask. We identified areas where the processed nigrosome-1 region overlapped with the binarized SN mask. These overlapping areas were then designated as nigrosome-1. We determined the optimal cut-off using the analyzed nigrosome-1 volume and subsequently evaluated the diagnostic performance.

To account for interindividual variation, normalization was conducted using the intracranial volume. SynthSeg was utilized to quantify the intracranial volume from three-dimensional T1-weighted imaging. SynthSeg—utilizing a CNN approach—is specifically designed to segment brain structures regardless of MRI contrast and resolution. This tool is integrated into the FreeSurfer 7.3.2 package, as described in the SynthSeg-FreeSurfer Wiki (harvard.edu).[6] In the version employed here, SynthSeg 2.0 can accurately segment native 3D-T1 to acquire parcellation volume and intracranial volume (ICV).[7,8]

Development of deep learning-based automatic classification for nigral hyperintensity

Heuron IPD, a deep learning-based system (version 1.0.1.10, Heuron Corporation, Seoul, Korea), commercially available in Korea since April 2021, was utilized to detect nigrostriatal degeneration. The system comprises a CNN-based object detection and segmentation model, integrating each model output to classify nigrostriatal degeneration (Fig 2). Heuron IPD first performs object detection using YOLOX model, followed by segmentation using our previously introduced Heuron NI model, which utilizes the SparseInst segmentation model. SMwI served as the input data for the Heuron IPD system, with the output indicating nigrostriatal degeneration.

In previous study, a system using YOLOv3[9] can classify nigrostriatal degeneration.[10] In this study, we used the object detection model YOLOX for classifying nigrostriatal degeneration.[11] This model was designed based on YOLOv3 and included features comprising anchor-free, decoupled head, multi-positive, and simOTA alongside the original YOLOv3. YOLOX utilizes SMwI data preprocessed to 640*640

as 2D images to predict the nigrosome-1 region and classify abnormalities. The model was trained using a dataset annotated with nigrosome-1 location by an experienced neurologist, with 18F-FP-CIT PET imaging used as the ground truth to determine the presence of nigrostriatal degeneration.

We captured the input images adhering to a consistent protocol. We observed the anchor-free approach was more efficient, as the images were not significantly influenced by anchor clustering. The model was optimized using simOTA[11] to address challenges inherent in anchor-free methods that categorize the center of the ground truth box as positive. While the Optimal Transport Assignment (OTA) can be employed to address the transport problem through the Sinkhorn-Knopp iteration, it introduces an increased learning overhead.[12] YOLOX reduced the learning cost by using simOTA with a dynamic top-K strategy.[11] The model employed a CSP-Darknet backbone, learning rate 0.01, batch size 8, and the SGD optimizer. For classification, the binary cross entropy loss function was used, while the IOU loss function was used for localization. For image augmentation, we used multi-scale learning (480 ~ 800, 32 range), mosaic,[13] mixup,[14] color jitter, random flip, and resize techniques. We performed 5-fold cross-validation on the training dataset to select the optimal model. All implementations were conducted using the open-source framework mmDetection (v.2.27.0, https://github.com/open-mmlab/mmdetection).

To classify the nigrostriatal degeneration, five slices from the scanned MR images containing SN and nigral hyperintensity structures were utilized for analysis.[10,15] Following image analysis, the object detection model returns a probability value indicating the likelihood of a specific object based on its training. The object detection model was improved by incorporating quantification information because the hyperintensity area of nigrosome-1 is lost in patients with IPD in SMwI. The derived volume from the segmentation model serves as a weight for the class probability value (Fig 2). Consequently, these weighted class probability values were employed to determine the abnormalities in the nigral hyperintensity region. The diagnostic performances of Heuron IPD were evaluated and compared with the model using only the CNN, without volume weight, to determine the volume weight effectiveness.

References

1. Cheng T, Wang X, Chen S, et al. Sparse instance activation for real-time instance segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022
2. He K, Zhang X, Ren S, et al. Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition, 2016
3. Kuhn HW. The Hungarian method for the assignment problem. Naval research logistics quarterly 1955;2:83-97
4. Lin TY, Goyal P, Girshick R, et al. Focal Loss for Dense Object Detection. IEEE Trans Pattern Anal Mach Intell 2020;42:318-27
5. Sung YH, Noh Y, Kim EY. Early-stage Parkinson's disease: Abnormal nigrosome 1 and 2 revealed by a voxelwise analysis of neuromelanin-sensitive MRI. Hum Brain Mapp 2021;42:2823-32
6. Billot B, Greve DN, Puonti O, et al. SynthSeg: Segmentation of brain MRI scans of any contrast and resolution without retraining. Med Image Anal 2023;86:102789
7. Billot B, Magdamo C, Cheng Y, et al. Robust machine learning segmentation for large-scale analysis of heterogeneous clinical brain MRI datasets. Proc Natl Acad Sci U S A 2023;120:e2216399120
8. Suh PS, Jung W, Suh CH, et al. Development and validation of a deep learning-based automatic segmentation model for assessing intracranial volume: comparison with NeuroQuant, FreeSurfer, and SynthSeg. Front Neurol 2023;14:1221892
9. Redmon J, Farhadi A. Yolov3: An incremental improvement. arXiv preprint arXiv:1804.02767 2018
10. Shin DH, Heo H, Song S, 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
11. Ge Z, Liu S, Wang F, et al. Yolox: Exceeding yolo series in 2021. arXiv preprint arXiv:2107.08430 2021
12. Ge Z, Liu S, Li Z, et al. Ota: Optimal transport assignment for object detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021
13. Wei Z, Duan C, Song X, et al. Amrnet: Chips augmentation in aerial images object detection. arXiv preprint arXiv:2009.07168 2020
14. Zhang H, Cisse M, Dauphin YN, et al. mixup: Beyond empirical risk minimization. arXiv preprint arXiv:1710.09412 2017
15. Sung YH, Lee J, Nam Y, et al. Differential involvement of nigral subregions in idiopathic parkinson's disease. Hum Brain Mapp 2018;39:542-53

Supplemental Table 1. Checklist for Artificial intelligence In Medical Imaging (CLAIM)

Section / Topic	No.	Item	Our work	Comment
TITLE / ABSTRACT				
	1	Identification as a study of AI methodology, specifying the category of technology used (e.g., deep learning)	O	Page 1, 2
ABSTRACT				
	2	Summary of study design, methods, results, and conclusions	O	Page 2
INTRODUCTION				
	3	Scientific and/or clinical background, including the intended use and role of the AI approach	O	Page 4, 5
	4	Study aims, objectives, and hypotheses	O	Page 4
METHODS				
Study Design	5	Prospective or retrospective study	O	Page 1, Page 6.
	6	Study goal	O	Page 4.
Data	7	Data sources	O	Page 5~6
	8	Inclusion and exclusion criteria	O	Page 5~6, Figure 1.
	9	Data pre-processing	O	Page 8, 10
	10	Selection of data subsets	O	Page 7.
	11	De-identification methods	X	
	12	How missing data were handled	O	Page 5~6, Figure 1.
	13	Image acquisition protocol	O	Page 6.
Reference Standard	14	Definition of method(s) used to obtain reference standard	O	
	15	Rationale for choosing the reference standard	O	
	16	Source of reference standard annotations	O	
	17	Annotation of test set	O	
	18	Measures of inter- and intra-rater variability of features described by the annotators	O	
Data Partitions	19	How data were assigned to partitions	O	Page 5, 6
	20	Level at which partitions are disjoint	O	Page 5, 6
Testing Data	21	Intended sample size	O	Page 6.
Model	22	Detailed description of model	O	Page 7~10
	23	Software libraries, frameworks, and packages	O	Page 7, 9, 10, 11
	24	Initialization of model parameters	O	Page 7~10
Training	25	Details of training approach	O	Page 8, 10
	26	Method of selecting the final model	O	Page 10
	27	Ensembling techniques	X	
Evaluation	28	Metrics of model performance	O	Table 2
	29	Statistical measures of significance and uncertainty	O	Page 11
	30	Robustness or sensitivity analysis	O	Supplement Table 3~7
	31	Methods for explainability or interpretability	O	Page 11
	32	Evaluation on internal data	X	
	33	Testing on external data	O	Page 6, Table 2
	34	Clinical trial registration	X	
Result				
Data	35	Numbers of patients or examinations included and excluded	O	Page 6, Figure 1
	36	Demographic and clinical characteristics of cases in each partition	O	Figure 1
Model performance	37	Performance metrics and measures of statistical uncertainty	O	Table 2, Supplement Table 3~7
	38	Estimates of diagnostic performance and their precision	O	Table 2, Supplement Table 3~7
	39	Failure analysis of incorrect results	O	
DISCUSSION				
	40	Study limitations	O	Page 15, 16
	41	Implications for practice, including intended use and/or clinical role	O	Page 14, 15, 16
OTHER INFORMATION				
	42	Provide a reference to the full study protocol or to additional technical details	O	Supplementary Material
	43	Statement about the availability of software, trained model, and/or data	X	
	44	Sources of funding and other support; role of funders	O	

Supplemental Table 2. SMwI protocol for three different scanners. Participants in the previous study (Shin et al.) underwent MR imaging on a 3T MRI with a 32-channel coil (Skyra, Siemens Healthineers, Forchheim, Germany). Participants from three hospitals also underwent MR imaging. The MRI scanners used were as follows: 1) Ingenia CX, Philips Medical System, Best, the Netherlands 2) Ingenia Elition X, Philips Medical System, Best, the Netherlands 3) Skyra, Siemens Healthineers, Forchheim, Germany.

Participants for the external validation also underwent MR imaging (Ingenia CX, Philips Medical System, Best, the Netherlands)

H&Y score	Siemens	Philips
Matrix	384 x 384	384 x 384
FOV (mm^3)	192 x 192 x 32	192 x 192 x 32
Slice thickness(mm)	1	1
voxel size (mm^3)	0.5 x 0.5 x 1	0.5 x 0.5 x 1
Flip Angle	20 deg	20 deg
Number of Echo	3	3
TR (ms)	48	48
TE (ms)	14 / 26.5 / 39	14.38 / 26.99 / 39.6
Bandwidth	100 Hz / pixel	100 Hz / pixel

Supplement Table 3. Demographic findings and clinical characteristics of training data

	IPD (n=56)	Control (n=95)	P-value
Age, yr	70 ± 10	56 ± 17	.0001*
Sex, male:female	32:24	36:59	.022†

*Comparison of two groups using Mann-Whitney test, †Comparison of two groups using chi-square test.

Supplemental Table 4. Diagnostic performance of HeuronIPD in differentiating IPD from ET and DIP

	Sensitivity, % [95 % CI]	Specificity, % [95 % CI]	Accuracy, % [95 % CI]	F1-score, % [95 % CI]	AUC [95 % CI]
ET vs IPD					
Right	94.6 [90.0, 97.5]	96.7 [88.5, 99.6]	95.1 [91.5, 97.6]	96.6 [93.4, 98.6]	0.980 [0.951, 0.994]
Left	95.1 [90.6, 97.9]	95.2 [86.5, 99.0]	95.1 [91.5, 97.6]	96.6 [93.3, 98.6]	0.965 [0.932, 0.985]
Subject	97.0 [93.2, 99.0]	96.6 [88.1, 99.6]	96.9 [93.7, 98.8]	97.9 [95.1, 99.3]	
DIP vs IPD					
Right	94.6 [90.0, 97.5]	92.3 [64.0, 99.8]	94.4 [90.0, 97.3]	96.9 [93.2, 98.9]	0.959 [0.919, 0.983]
Left	95.1 [90.6, 97.9]	93.3 [68.1, 99.8]	95.0 [90.7, 97.7]	97.2 [93.6, 99.1]	0.968 [0.931, 0.989]
Subject	97.0 [93.2, 99.0]	90.9 [58.7, 99.8]	96.7 [92.9, 98.8]	98.2 [95.0, 99.6]	

Abbreviations: IPD, Idiopathic Parkinson's disease; ET, essential tremor; DIP; drug-induced parkinsonism

Supplemental Table 5. Comparisons of AUC in differentiating IPD from ET and DIP using Hanley & McNeil method

	ET vs IPD	DIP vs IPD	p-value
Right	0.980	0.959	.462
Left	0.965	0.968	.887

Supplementary Table 6. Diagnostic performance of HeuronIPD system by gender

	Sensitivity, % [95 % CI]	Specificity, % [95 % CI]	Accuracy, % [95 % CI]	F1-score, % [95 % CI]	AUC [95 % CI]
Female					
Right	95.5 [88.9, 98.8]	97.6 [87.7, 99.9]	96.2 [91.4, 98.8]	97.1 [92.7, 99.3]	0.985 [0.946, 0.998]
Left	90.7 [82.5, 95.9]	100 [92.3, 100]	93.9 [88.4, 97.3]	95.1 [89.9, 98.1]	0.967 [0.921, 0.990]
Subject	95.5 [88.9, 98.8]	100 [91.8, 100]	97.0 [92.4, 99.2]	97.7 [93.5, 99.5]	
Male					
Right	96.1 [89.0, 99.2]	92.9 [76.5, 99.1]	95.2 [89.2, 98.4]	96.7 [91.3, 99.2]	0.963 [0.907, 0.990]
Left	100 [95.4, 100]	88.9 [70.8, 97.6]	97.1 [91.9, 99.4]	98.1 [93.3, 99.8]	0.958 [0.900, 0.987]
Subject	98.7 [93.1, 100]	88.5 [69.8, 97.6]	96.2 [90.5, 99.0]	97.5 [92.4, 99.6]	

Supplementary Table 7. Comparison of AUC by gender using Hanley & McNeil method

	ET vs IPD	DIP vs IPD	p-value
Right	0.985	0.963	.331
Left	0.967	0.958	.794

Supplemental Table 8. Comparison of nigrosome-1 volume among the two groups of H&Y scores.

	Nigrosome volume (mm^3)	P-value	Nigrosome volume/ICV ($mm^3 \times 10^6$)	P-value
Group 1		0.06		0.074
1–1.5	1.75 ± 1.07 (0.62–2.88)		1.18 ± 0.73	
2–2.5	1.99 ± 1.65 (1.72–2.26)		1.35 ± 1.15	
3–3.5	1.36 ± 1.12 (1.09–1.63)		0.92 ± 0.76	
≥ 4	1.62 ± 1.38 (1.05–2.18)		1.09 ± 0.94	
Group 2		0.008		0.011
1–2.5	1.98 ± 1.63 (1.72–2.24)		1.34 ± 1.13	
3–5	1.43 ± 1.19 (1.18–1.67)		0.96 ± 0.81	

Supplement Table 9. Comparative analysis with other deep learning models

	Sensitivity, % [95 % CI]	Specificity, % [95 % CI]	Accuracy, % [95 % CI]	F1-score, % [95 % CI]
Yolox				
Right	93.9 [89.1, 97.0]	94.5 [86.6, 98.5]	94.1 [92.2, 97.9]	95.7 [90.3 ,96.7]
Left	92.2 [87.0, 95.8]	94.4 [86.2, 98.4]	92.8 [88.8, 95.8]	94.7 [91.1 ,97.2]
Subject	97.0 [93.2, 99.0]	91.3 [82.0 ,96.7]	95.4 [93.6 ,98.6]	96.7 [91.9, 97.7]
Yolov3				
Right	89.0 [83.2, 93.4]	97.3 [90.5, 99.7]	91.6 [87.3, 94.8]	93.6 [90.8, 96.1]
Left	89.2 [83.4, 93.4]	95.8 [88.1, 99.1]	91.1 [87.8, 94.4]	93.3 [90.4, 96.1]
Subject	92.9 [87.8, 99.6]	97.1 [89.9, 99.6]	94.1 [90.3, 96.7]	95.7 [93.2, 97.9]
Pvtv2				
Right	89.6 [83.9, 93.8]	94.5 [86.6, 98.5]	91.1 [86.8, 94.4]	93.3 [90.3, 96.1]
Left	90.1 [85.5, 94.9]	94.4 [86.2, 98.4]	92.0 [87.8, 95.1]	94.1 [91.4, 96.7]
Subject	93.5 [88.6, 96.7]	94.2 [85.8, 98.4]	93.7 [89.8, 96.4]	95.4 [93.1, 97.4]
Detectors				
Right	87.2 [81.1, 91.9]	90.4 [81.2, 96.1]	88.2 [83.4, 92.0]	91.1 [89.4, 94.0]
Left	89.8 [84.1, 93.9]	88.7 [79.0, 95.0]	89.5 [84.8, 93.1]	92.2 [89.0, 95.1]
Subject	94.6 [90.0, 97.5]	87.0 [76.7, 93.9]	92.4 [88.3, 95.4]	94.6 [92.0, 96.9]