European Radiology (2019) 29:6891–6899
https://doi.org/10.1007/s00330-019-06327-0

IMAGING INFORMATICS AND ARTIFICIAL INTELLIGENCE

Deep learning to differentiate parkinsonian disorders separately using single midsagittal MR imaging: a proof of concept study

Shigeru Kiryu[1] · Koichiro Yasaka[2] · Hiroyuki Akai[2] · Yasuhiro Nakata[3] · Yusuke Sugomori[4] · Seigo Hara[4] · Maria Seo[4] · Osamu Abe[5] · Kuni Ohtomo[6]

Received: 26 April 2019 / Revised: 4 June 2019 / Accepted: 13 June 2019 / Published online: 1 July 2019

Abstract

Objectives To evaluate the diagnostic performance of deep learning with the convolutional neural networks (CNN) to distinguish each representative parkinsonian disorder using MRI.

Methods This clinical retrospective study was approved by the institutional review board, and the requirement for written informed consent was waived. Midsagittal T1-weighted MRI of a total of 419 subjects (125 Parkinson's disease (PD), 98 progressive supranuclear palsy (PSP), and 54 multiple system atrophy with predominant parkinsonian features (MSA-P) patients, and 142 normal subjects) between January 2012 and April 2016 was retrospectively assessed. To deal with the overfitting problem of deep learning, all subjects were randomly divided into training (85%) and validation (15%) data sets with the same proportions of each disease and normal subjects. We trained the CNN to distinguish each parkinsonian disorder using single midsagittal T1-weighted MRI with a training group to minimize the differences between predicted output probabilities and the clinical diagnoses; then, we adopted the trained CNN to the validation data set. Subjects were classified into each parkinsonian disorder or normal condition according to the final diagnosis of the trained CNN, and we assessed the diagnostic performance of the CNN.

Results The accuracies of diagnostic performances regarding PD, PSP, MSA-P, and normal subjects were 96.8, 93.7, 95.2, and 98.4%, respectively. The areas under the receiver operating characteristic curves for distinguishing each condition from others (PD, PSP, MSA-P, and normal subjects) were 0.995, 0.982, 0.990, and 1.000, respectively.

Conclusion Deep learning with CNN enables highly accurate discrimination of parkinsonian disorders using MRI.

Key Points

- *Deep learning convolution neural network achieves differential diagnosis of PD, PSP, MSA-P, and normal controls with an accuracy of 96.8, 93.7, 95.2, and 98.4%, respectively.*
- *The areas under the curves for distinguishing between PD, PSP, MSA-P, and normality were 0.995, 0.982, 0.990, and 1.000, respectively.*
- *CNN may learn important features that humans not notice, and has a possibility to perform previously impossible diagnoses.*

Keywords Artificial intelligence · Parkinson disease · Magnetic resonance imaging · ROC curve · Deep learning

✉ Shigeru Kiryu
kiryu-tky@umin.ac.jp

[1] Department of Radiology, International University of Health and Welfare Hospital, 537-3, Iguchi, Nasushiobara, Tochigi 329-2763, Japan

[2] Department of Radiology, The Institute of Medical Science, The University of Tokyo, 4-6-1 Shirokanedai, Minato-ku, Tokyo 108-8639, Japan

[3] Department of Neuroradiology, Tokyo Metropolitan Neurological Hospital, 2-6-1 Musashidai, Fuchu, Tokyo 183-0042, Japan

[4] MICIN, Inc., Nihon Building 13F, 2-6-2 Otemachi, Chiyoda-ku, Tokyo 100-0004, Japan

[5] Department of Radiology, Graduate School of Medicine, The University of Tokyo, 7-3-1, Hongo, Bunkyo-ku, Tokyo 113-0033, Japan

[6] International University of Health and Welfare, 2600-1 kitakanamaru, Otawara, Tochigi 324-8501, Japan

Abbreviations

CNN	Convolutional neural network
CNS	Central nervous system
DWI	Diffusion-weighted imaging
JPEG	Joint photographic experts group
MR	Magnetic resonance
MSA-P	Multiple system atrophy with predominant parkinsonian features
PD	Parkinson's disease
PSP	Progressive supranuclear palsy
ReLU	Rectified linear unit
ROC	Receiver operating characteristic

Introduction

Parkinsonian disorders, including Parkinson's disease (PD), progressive supranuclear palsy (PSP), and multiple system atrophy with predominant parkinsonian features (MSA-P), exhibit differences in their responses to treatment and prognosis. Therefore, distinguishing among the diseases is critical for clinical management [1–3]. However, this can be difficult because these diseases frequently share clinical features, particularly in the early stages, which is further complicated by increasing evidence of mixed pathologies [4].

MRI provides important biomarkers for the diagnosis of neurodegenerative disorders. One important role of MRI is to exclude secondary causes of parkinsonism, such as vascular disease, and several applications of MRI in differential diagnosis have been studied for diagnosis of parkinsonian disorders [5]. Among the several MRI techniques available, midsagittal brainstem measurement is both convenient and straightforward [5–7]. The pontine areas of MSA patients are less prominent than those of non-MSA parkinsonian patients, and the midbrain areas are less prominent in PSP patients than in non-PSP parkinsonian patients [6]. However, no information is available at present to allow differentiation of PD from other parkinsonian disorders or to differentiate each parkinsonian disorder separately using midsagittal brainstem measurement [8].

Recently, deep learning with convolutional neural networks (CNN) has been developed for use in medical imaging research [9, 10]. CNN has been used to evaluate lesions on CT, mammography, and MRI [11–16]. Conventional machine learning or visual assessment requires advance knowledge of specific imaging features, but CNN uses imaging features extracted via the convolutional process to learn. Thus, CNN may learn important features that human operators may not notice, and may be able to reduce the number of equivocal diagnoses.

We hypothesized that CNN may be able to distinguish among parkinsonian disorders using MRI. Here, we assessed the diagnostic accuracy of the CNN to distinguish among PD, PSP, MSA-P patients, and normal subjects using single midsagittal MRI of the brainstem.

Materials and methods

This retrospective study was approved by our institutional research ethics committee as retrospective data analysis for medical imaging-based diagnoses, and the requirement for informed consent was waived.

Subjects

Of the 22,723 patients who underwent brain MRI examinations at a single hospital from January 2012 to April 2016, a total of 3330 underwent brain MRI according to the parkinsonian disorder protocol on suspicion of a parkinsonian disorder (Fig. 1). Of these scans, 372 were diagnosed by neurologists as PD, PSP, or MSA-P. All patients underwent full neurological and general examinations and basic laboratory tests by neurologists, and their medical records were reviewed by four of the authors. The clinical diagnoses were based on the criteria of the National Institute of Neurological Diseases and Stroke (for PSP) [17], the UK Brain Bank Criteria (for PD) [18], and the Consensus Statement on the Diagnosis of Multiple-System Atrophy (for MSA-P) [19]. If more than one MRI examination for a given patient was available, we used only the earliest one; thus, 95 examinations were excluded. Hence, 125 patients with PD, 98 with PSP, and 54 with MSA-P were included in the study. From a series of patients who had undergone the parkinsonian disorder protocol during the same period, a total of 142 age- and sex-matched normal subjects with no history of CNS disease and normal neurological examinations were also enrolled. Ultimately, a total of 419 subjects were included in the study (mean age, 70.9 ± 9.5 years; range, 41–92 years; 192 males and 227 females). Demographic data are shown in Table 1. Overall, 19, 18, and 88 subjects underwent MRI within 3 years, from 3 to 5 years, and more than 5 years after onset of PD, respectively. There were 30, 20, and 48 for PSP, and 14, 15, and 25 for MSA-P, respectively. The average intervals between disease onset and MRI examination were 11.07 ± 8.18 years for PD, 5.71 ± 4.01 years for PSP, and 4.87 ± 2.22 years for MSA-P.

To address the overfitting problem associated with deep learning, the study was performed in a hold-out manner as in other deep learning studies [9, 10]. All subjects were randomly divided into training (85%) and validation (15%) datasets, each with the same proportion of each condition. Ultimately, 356 (107 PD, 85 PSP, and 45 MSA-P patients; 119 normal subjects) and 63 (18 PD, 13 PSP, and 9 MSA-P patients; 23 normal subjects) subjects constituted the training and validation datasets, respectively.

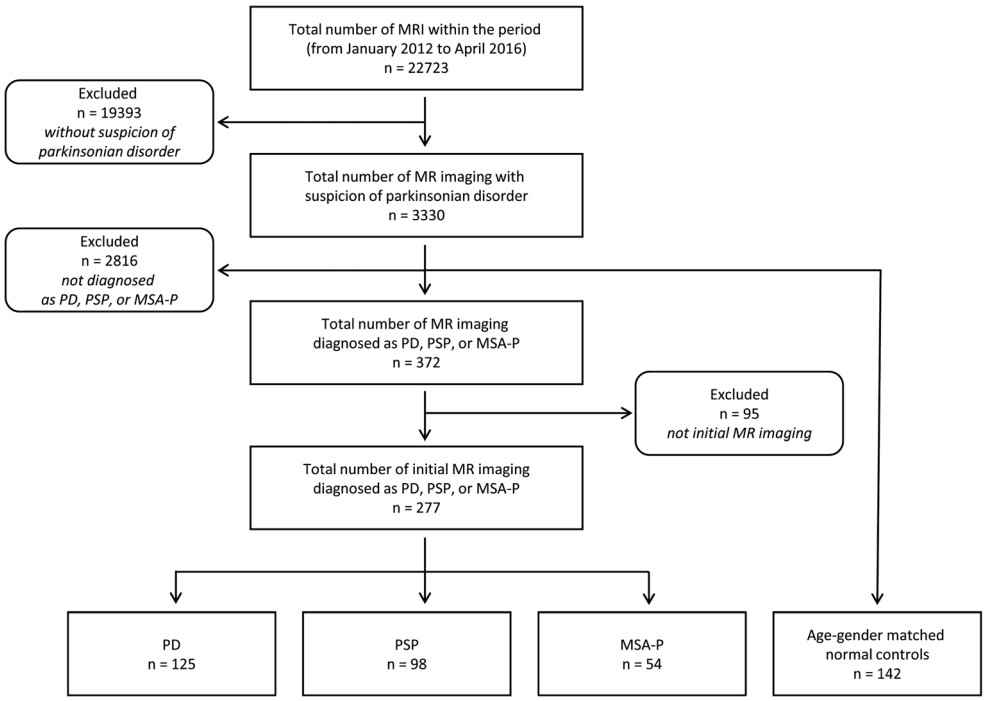

Fig. 1 Patient flow diagram. PD, Parkinson disease; PSP, progressive supranuclear palsy; MSA-P, multiple system atrophy with predominant parkinsonian features

MRI examination of parkinsonian protocol

Single, midsagittal, T1-weighted MR images of the brainstem were acquired per protocol when assessing parkinsonian disorders. To ensure the reproducibility of the scanning plane, the image passed through the centers of the interpeduncular cistern and the aqueduct as revealed using a pre-acquired axial image at the level of the midbrain (Fig. 2) as described previously [7]. We used two MR scanners from the same vendor. MRI scans of 154 subjects were acquired on a 1.5 T scanner (Signa CV/I HDxt; GE Healthcare) and the remaining 265 subjects underwent MRI using a 3 T scanner (Discovery MR750; GE Healthcare). The imaging parameters are summarized in Table 2. Overall, 223 (63%) and 133 (37%) subjects for the training dataset and 42 (67%) and 21 subjects (33%) for the validation dataset were scanned using the 3 T and 1.5 T scanners, respectively.

Image processing

MR images were displayed using a picture archiving and communication system (Centricity Radiology RA 1000; GE Healthcare) at the default window-level setting (automatically optimized by the MRI scanner) and we converted the midsagittal brain slice to 8-bit joint photographic experts group (JPEG) format using the image capture function. The images were further processed using a computer fitted with a GeForce GTX 1080Ti graphics processing unit (NVIDIA), an Intel Core i7-6800K 3.40 GHz central processing unit, and 32.0 GB of RAM. The images were converted to gray-scale images and augmented to render the CNN model robust against random zooming (5%), random rotation (5 degree), random parallel shift (2%), and slight differences in noise (5%). Thus, a total of 32 images were generated from each original image.

Table 1 Subject numbers and ages

	PD	PSP	MSA-P	Normal
Total subjects	125	98	54	142
Male	64	53	19	56
Female	61	45	35	86
Age (years)	70.2 ± 10.4	74.5 ± 7.2	66.1 ± 8.7	71.0 ± 9.6
Average interval between disease onset and MRI examination (years)	11.07 ± 8.18	5.71 ± 4.01	4.87 ± 2.22	N/A

The data are subject numbers or averages ± standard deviations

PD Parkinson's disease, *PSP* progressive supranuclear palsy, *MSA-P* multiple system atrophy with predominant parkinsonian features, *N/A* not applicable

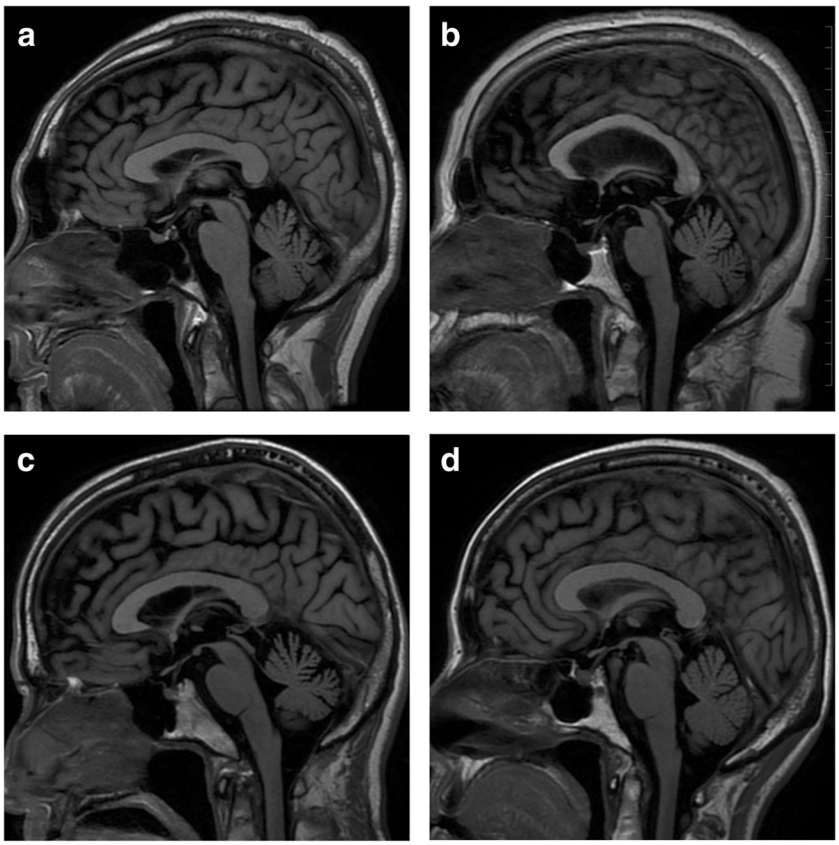

Fig. 2 Example of midsagittal T1-weighted MR images of (**a**) Parkinson disease, (**b**) progressive supranuclear palsy, (**c**) multiple system with predominant parkinsonian features, and (**d**) normal subjects. All cases were diagnosed correctly by the deep convolutional neural network

Training and validation phase

The supervised training dataset was fed to the deep learning process to create a model to differentiate each parkinsonian disorder group and normal subjects using the same hardware as used for image processing. The Python programming language and Keras Python deep learning library 2.0.4 (https://keras.io/) were used to train the CNN model. The CNN model repeated two convolutional layers and one max pooling layer set twice, followed by two fully connected layers. The convolutional layer convolutes the previous layer with multiple convolution filters (3 × 3 pixels) and generates multiple feature maps to detect distinctive local motif-like edges, lines, and other visual elements [10]. For each convolutional layer, the padding and stride were set as 1 pixel. The max pooling layer, in which the maximum gray-scale levels of 2 × 2 pixels were generated, effectively reduces the dimension of the feature maps by the max function to make them less sensitive to small shifts or distortions of the target object. The final feature maps were unrolled into a long feature vector and fed into a fully connected layer. The fully connected layer integrated all of the feature responses and provided the final results. A 50% dropout was used to prevent overfitting [20], which is a potential problem with CNN. The rectified linear unit (ReLU) function served as the activation function to make the model nonlinear. It accelerated the convergence rate of the deep network. ReLU was placed after the convolutional layer except for the first convolutional layer because this configuration resulted in better performance in a preliminary experiment. Figure 3 presents detailed information about the size of the convolutional and max pooling layer and the insertion of dropout and the ReLU function into the

Table 2 Summary of MRI parameters

	1.5 T	3 T
Sequence	Spin-echo sequence	T1-fluid, attenuated inversion recovery
Repetition time	400	2500
Echo time (ms)	10	16
Flip angle (°)	90	111
Inversion time (ms)	N/A	982
Matrix	256 × 192	256 × 224
Slice thickness (mm)	4	3
Number of slices	21	24
Inter-slice gap (mm)	3	2
Field-of-view (mm)	220 × 220	220 × 220
Number of excitations	2	2

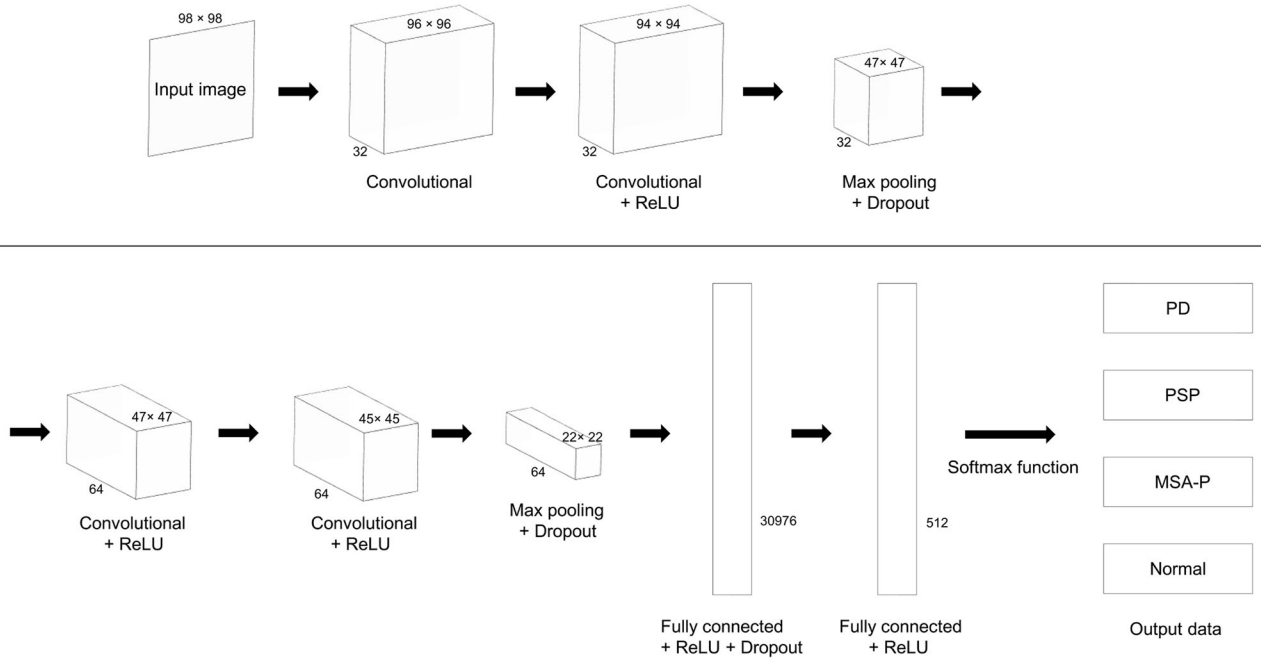

Fig. 3 The architecture of the deep convolutional neural network

model. Finally, the predicted probabilities for each group were obtained as outputs using the softmap function, calculated to ensure that the sum of each group was unity. The output data were compared with the clinical diagnoses (reference standards), and errors were back-propagated to update the CNN parameters, minimizing the errors between the output data and the reference standards. Mini-batches of 32 samples and a total of 117 epochs were used to train the CNN model.

The validation data set was fed to the trained model, and the output data was obtained as the predicted probability for each group. In result, a total of four predicted probabilities (PD, PSP, MSA-P, and normal subjects) were acquired for each condition. The highest predicted probability served as the final diagnosis. The same CNN model was applied in training and validating group.

Statistics

We used EZR software ver. 1.33 for statistical analyses [21]; this is a graphical user interface for the R software package (ver. 3.2.2; R Development Core Team). All subjects were divided into each condition or others using the final diagnosis of CNN, and the accuracy, sensitivity, and specificity associated with identification of each group were measured. Receiver operating characteristic (ROC) analyses were performed to determine the diagnostic performance of the CNN. All subjects were divided into each condition or others, and the areas under the curves (AUC) in ROC analyses distinguishing each group (via predicted probability) were assessed.

Results

A total of 92.1% of subjects were diagnosed correctly in the validation dataset. A total of five out of 63 cases, including one PD, two PSP, and two MSA-P patients, were misdiagnosed. The details are shown in Table 3. In cases diagnosed correctly by CNN, there were PSP cases where atrophy of the midbrain was unclear (Fig. 4), and MSA-P cases where atrophy of the pons was not clear (Fig. 5). These findings are not typical for PSP or MSA-P [8].

The accuracies of PD, PSP, MSA-P, and normal diagnoses were 96.8%, 93.7%, 95.2%, and 98.4%, respectively. The sensitivities and specificities are also shown in Table 4. The average (± standard deviation) predicted probabilities in cases correctly diagnosed were 0.915 (± 0.126) for PD, 0.819 (± 0.173) for PSP, 0.788 (± 0.146) for MSA-P, and 0.981 (± 0.034) for normal subjects. The predicted probabilities for a correct diagnosis in the five misdiagnosed cases by CNN were 0.260 for PD, 0.453 and 0.484 for PSP, and 0.021 and 0.087 for MSA-P. The AUCs for discriminating among PD, PSP, MSA-P, and normal subjects were 0.995, 0.982, 0.990, and 1.000, respectively (Fig. 6). Training took 30 min and validation took 2 min.

Discussion

We found that the CNN distinguished among parkinsonian disorders, affording high diagnostic performance using single midsagittal MRI. The definitive diagnosis of parkinsonian disorders is neuropathological in nature. Clinical diagnoses

Table 3 Confusion matrix of the clinical diagnoses and CNN predictions

		Clinical diagnosis			
		PD	PSP	MSA-P	Normal
CNN prediction	PD	17	1	0	0
	PSP	1	11	1	0
	MSA-P	0	1	7	0
	Normal	0	0	1	23

The data are subject numbers

PD Parkinson's disease, *PSP* progressive supranuclear palsy, *MSA-P* multiple system atrophy with predominant parkinsonian features

are suboptimal [4], and various biomarkers have been used to increase the in vivo discrimination of parkinsonian disorders [22]. Our results may increase the utility of MRI in the differential diagnosis of parkinsonian disorders. Due to the small number of cases in each diagnosis group in the validation set, the excellent performance results of the trained CNN may not be robust. However, the possibility of the usefulness of deep learning in diagnostic performance shown in this study may justify further validation on a larger dataset.

Midsagittal measurement of the brainstem is simple and feasible and is used to distinguish PSP and MSA from other conditions [6]. A large multicenter retrospective study showed that the sensitivities to differentiate PSP and MSA-P from normal subjects were 83.0% and 63.3%, respectively [23]. In addition to the midsagittal brainstem structure, other structures, such as the middle cerebellar peduncle, putamen, ventricle, and so forth, show specific changes in parkinsonian disorders [8]. However, it was not possible to differentiate among these diseases (i.e., PD, PSP, and MSA-P) [8].

Fig. 4 Brainstem images of progressive supranuclear palsy cases in same scale. Both cases were diagnosed correctly by the deep convolutional neural network. Atrophy of midbrain is shown in a 70-year-old female case (**a**), however, it is unclear in 78-year-old male case (**b**)

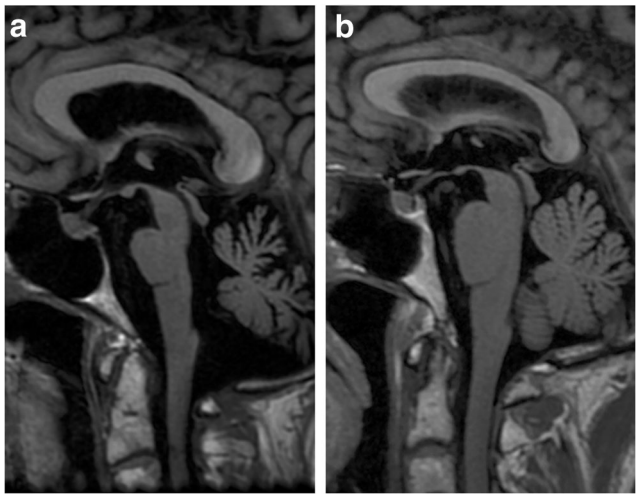

Fig. 5 Brainstem images of multiple system atrophy with predominant parkinsonian features cases in same scale. Both cases were diagnosed correctly by the deep convolutional neural network. Atrophy of pons is shown in a 61-year-old female case (**a**), however, it is unclear in 75-year-old female case (**b**)

Although MRI is one of the best sources of biomarkers for diagnosing neurodegenerative disorders, the utility of MRI for differentiating among parkinsonian disorders is limited at present. Our observation that CNN can be used to differentiate between each condition directly with high accuracy suggests that CNN increases the utility of MRI.

The T1-weighted imaging used in this study is a core MRI pulse sequence and has robust performance with regard to hardware and imaging parameters, where other sequences are not. This simplified our imaging technique, which may be applicable to other institutions.

We converted MRI into JPEG format, and thus some contrast information may have been lost. We assume that the CNN prioritized morphological over contrast information. It is not possible to understand how the CNN "thought"; the CNN is not formatted to this end. The midsagittal MRI contains not only the brainstem but also the corpus callosum, pituitary gland, and other structures. Therefore, it is possible that the CNN also analyzed structures other than the brainstem, such as the corpus callosum, changes in which are associated with parkinsonian disorders [24–26]. Attempts have been made to determine the decision processes of deep learning [27–29]. Although this work is in its early stages, if the CNN process becomes clearer in the future, the differentiation of parkinsonian disorders may improve further, along with diagnostic radiology in general.

Our single-slice procedure using CNN is useful for differentiating among parkinsonian disorders, and we consider it practical given the present state of CNN research. As the number of imaging slices increases, assessment using CNN requires more training and more powerful computer systems, but it is a promising method to obtain better results. This point should be assessed in future studies.

Table 4 Summary of the diagnostic performance of the CNN

	PD	PSP	MSA-P	Normal
Sensitivity (%)	94.4 (17/18)	84.6 (11/13)	77.8 (7/9)	100.0 (23/23)
Specificity (%)	97.8 (44/45)	96.0 (48/50)	98.1 (53/54)	97.5 (39/40)
Accuracy (%)	96.8 (61/63)	93.7 (59/63)	95.2 (60/63)	98.4 (62/63)

The data are percentages of subject numbers

PD Parkinson's disease, *PSP* progressive supranuclear palsy, *MSA-P* multiple system atrophy with predominant parkinsonian features

The reference standard was a clinical diagnosis based on established criteria [17–19], which had a high but not a 100% accuracy in the diagnosis of parkinsonian disorders. A neuropathological diagnosis by brain autopsy in all cases is not realistic in the clinical settings. In this situation, there is a possibility that the label based on clinical diagnosis may be erroneous in the cases that the determination of trained CNN is judged as incorrect. The diagnostic accuracy of the clinical diagnosis of parkinsonian disorders had been improved by time, during follow-up. The average interval of most subjects between disease onset and MRI examination is several years in this study; therefore, to consider the trained CNN diagnosis as inaccurate in cases of divergent diagnoses may be appropriate. There is another concern that the patients incorrectly diagnosed by the trained CNN may be treated inappropriately. Parkinsonian disorders are chronic diseases and long-term follow-up is performed usually. Similar to clinical practice based on clinical diagnosis, misdiagnosis by AI is expected to be corrected during this follow-up period.

Although we need more validation cases, the accuracy for normal subjects tended to be high in this study. Thus, further examinations may be omitted for cases diagnosed

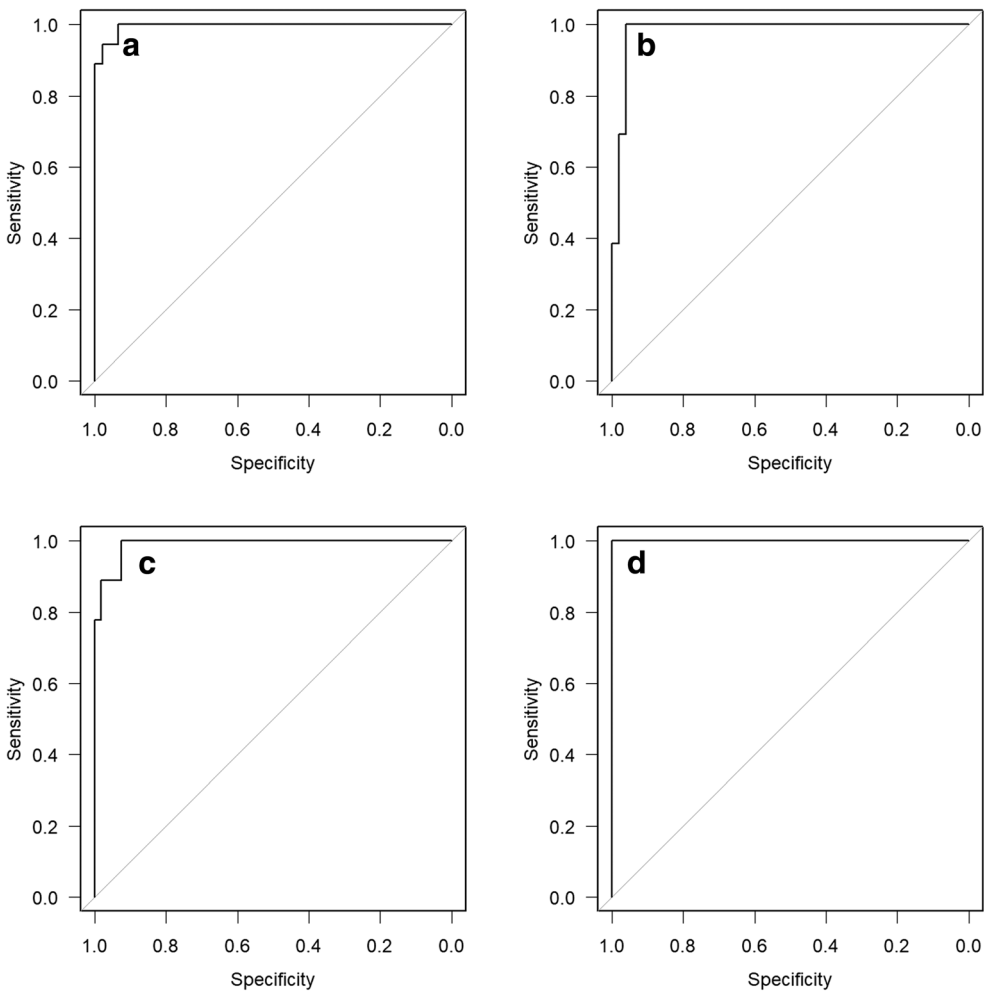

Fig. 6 Receiver operating characteristic curves of (**a**) Parkinson disease, (**b**) progressive supranuclear palsy, (**c**) multiple system with predominant parkinsonian features, and (**d**) normal subjects

as normal by the trained CNN, an approach that is likely to be cost-effective. Due to the limited number of neurologists and dedicated neurology hospitals, correct diagnosis of parkinsonian disorders is made in confined facilities specialized in neurological diseases. Since the AI-based diagnosis using MRI is possible wherever MRI scanner is available, a higher rate of correct diagnosis can be expected by using the trained CNN. This is one of its highly anticipated advantages to AI.

Our study had some limitations. First, it was a retrospective study, and therefore had intrinsic selection bias. Second, it was a single-institution study, and our results require confirmation at other institutes. As mentioned above, as we used a single T1-weighted image, we believe that our method would be robust at other institutions. Third, MRI was performed using two MRI scanners with different magnetic field strengths (1.5 T and 3 T). As midsagittal imaging is performed at the center of the magnetic field, any geometrical distortion is theoretically small. Moreover, the two scanners were produced by the same vendor. However, there were differences in the imaging sequence and scanning parameters because these were optimized for each MRI scanner. These differences may have caused mislabeling. Problems associated with differences in imaging parameters and MRI scanners may occur when conducting deep learning research on larger numbers of cases at other institutions. We expect that this may be resolved by increasing the number of cases. Fourth, the number of cases of each disease was not balanced. We gathered cases comprehensively over a period of time, and therefore the number of MSA-P cases tended to be lower than the others. It is possible that this imbalance could have affected each predicted probability. From a statistical perspective, the small number of cases for each diagnosis in the validation group was a limitation of this study. We will collect more cases in collaboration with other institutions and investigate the effects of the proportion of parkinsonian disorders in a future study. The inclusion of some advanced cases in the dataset after disease onset was also a limitation of this study. The sensitivity of the imaging features of each disease by CNN may be high in advanced cases, and we cannot extrapolate the current high performance of CNN in this study to the differentiation of parkinsonian diseases in the early stages.

In conclusion, we assessed the diagnostic performance of deep learning for distinguishing among PD, PSP, MSA-P, and normal status using single midsagittal MRI. As we validated the performance of CNN on a limited number of cases, the excellent performance of the trained CNN may not be robust. However, further validation on a larger dataset may be justified. Although the imaging features used by the CNN to distinguish the conditions remain unknown, the diagnostic performance was high in this limited dataset.

Funding The authors state that this work has not received any funding.

Compliance with ethical standards

Guarantor The scientific guarantor of this publication is Dr. Shigeru Kiryu.

Conflict of interest The authors of this manuscript declare no relationships with any companies, whose products or services may be related to the subject matter of the article.

Statistics and biometry No complex statistical methods were necessary for this paper.

Informed consent Written informed consent was waived by the Institutional Review Board.

Ethical approval Institutional Review Board approval was obtained.

Methodology
- retrospective
- diagnostic or prognostic study
- performed at one institution

References

1. Seppi K, Yekhlef F, Diem A et al (2005) Progression of parkinsonism in multiple system atrophy. J Neurol 252:91–96
2. Dickson DW, Rademakers R, Hutton ML (2007) Progressive supranuclear palsy: pathology and genetics. Brain Pathol 17:74–82
3. Marras C, Lang A (2008) Invited article: changing concepts in Parkinson disease: moving beyond the decade of the brain. Neurology 70:1996–2003
4. Rizzo G, Copetti M, Arcuti S, Martino D, Fontana A, Logroscino G (2016) Accuracy of clinical diagnosis of Parkinson disease: a systematic review and meta-analysis. Neurology 86:566–576
5. Hotter A, Esterhammer R, Schocke MF, Seppi K (2009) Potential of advanced MR imaging techniques in the differential diagnosis of parkinsonism. Mov Disord 24:S711–S720
6. Cosottini M, Ceravolo R, Faggioni L et al (2007) Assessment of midbrain atrophy in patients with progressive supranuclear palsy with routine magnetic resonance imaging. Acta Neurol Scand 116: 37–42
7. Oba H, Yagishita A, Terada H et al (2005) New and reliable MRI diagnosis for progressive supranuclear palsy. Neurology 64:2050–2055
8. Heim B, Krismer F, De Marzi R, Seppi K (2017) Magnetic resonance imaging for the diagnosis of Parkinson's disease. J Neural Transm (Vienna) 124:915–964
9. Yasaka K, Akai H, Kunimatsu A, Kiryu S, Abe O (2018) Deep learning with convolutional neural network in radiology. Jpn J Radiol 36:257–272
10. Lee JG, Jun S, Cho YW et al (2017) Deep learning in medical imaging: general overview. Korean J Radiol 18:570–584
11. van der Burgh HK, Schmidt R, Westeneng HJ, de Reus MA, van den Berg LH, van den Heuvel MP (2016) Deep learning predictions of survival based on MRI in amyotrophic lateral sclerosis. Neuroimage Clin 13:361–369
12. Laukamp KR, Thiele F, Shakirin G et al (2019) Fully automated detection and segmentation of meningiomas using deep learning on routine multiparametric MRI. Eur Radiol 29:124–132

13. Lin W, Tong T, Gao Q et al (2018) Convolutional neural networks-based MRI image analysis for the Alzheimer's disease prediction from mild cognitive impairment. Front Neurosci 12:777
14. Xin J, Zhang Y, Tang Y, Yang Y (2019) Brain differences between men and women: evidence from deep learning. Front Neurosci 13:185
15. Korfiatis P, Erickson B (2019) Deep learning can see the unseeable: predicting molecular markers from MRI of brain gliomas. Clin Radiol 74:367–373
16. Gong E, Pauly JM, Wintermark M, Zaharchuk G (2018) Deep learning enables reduced gadolinium dose for contrast-enhanced brain MRI. J Magn Reson Imaging 48:330–340
17. Litvan I, Agid Y, Calne D et al (1996) Clinical research criteria for the diagnosis of progressive supranuclear palsy (Steele-Richardson-Olszewski syndrome): report of the NINDS-SPSP International Workshop. Neurology 47:1–9
18. Hughes AJ, Daniel SE, Kilford L, Lees AJ (1992) Accuracy of clinical diagnosis of idiopathic Parkinson's disease: a clinico-pathological study of 100 cases. J Neurol Neurosurg Psychiatry 55:181–184
19. Gilman S, Wenning GK, Low PA et al (2008) Second consensus statement on the diagnosis of multiple system atrophy. Neurology 71:670–676
20. Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R (2014) Dropout: a simple way to prevent neural networks from overfitting. J Mach Learn Res 15:1929–1958
21. Kanda Y (2013) Investigation of the freely available easy-to-use software "EZR" for medical statistics. Bone Marrow Transplant 48:452–458
22. Rizzo G, Zanigni S, De Blasi R et al (2016) Brain MR contribution to the differential diagnosis of parkinsonian syndromes: an update. Parkinsons Dis 2016:2983638
23. Möller L, Kassubek J, Südmeyer M et al (2017) Manual MRI morphometry in parkinsonian syndromes. Mov Disord 32:778–782
24. Goldman JG, Bledsoe IO, Merkitch D, Dinh V, Bernard B, Stebbins GT (2017) Corpus callosal atrophy and associations with cognitive impairment in Parkinson disease. Neurology 88:1265–1272
25. Boelmans K, Bodammer NC, Suchorska B et al (2010) Diffusion tensor imaging of the corpus callosum differentiates corticobasal syndrome from Parkinson's disease. Parkinsonism Relat Disord 16:498–502
26. Rosskopf J, Müller HP, Huppertz HJ, Ludolph AC, Pinkhardt EH, Kassubek J (2014) Frontal corpus callosum alterations in progressive supranuclear palsy but not in Parkinson's disease. Neurodegener Dis 14:184–193
27. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D (2017) Grad-CAM: visual explanations from deep networks via gradient-based localization. Available via https://arxiv.org/abs/1610.02391. Accessed 2 Feb 2019
28. Samek W, Binder A, Montavon G, Lapuschkin S, Muller KR (2017) Evaluating the visualization of what a deep neural network has learned. IEEE Trans Neural Netw Learn Syst 28:2660–2673
29. Philbrick KA, Yoshida K, Inoue D et al (2018) What does deep learning see? Insights from a classifier trained to predict contrast enhancement phase from CT images. AJR Am J Roentgenol 211:1184–1193