Computers in Biology and Medicine 132 (2021) 104312

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: http://www.elsevier.com/locate/compbiomed

Improved motor outcome prediction in Parkinson's disease applying deep learning to DaTscan SPECT images

Matthew P. Adams [a], Arman Rahmim [b], Jing Tang [a,c,*]

[a] Department of Electrical and Computer Engineering, Oakland University, Rochester, MI, USA
[b] Departments of Radiology and Physics, University of British Columbia, Vancouver, BC, Canada
[c] Department of Bioengineering, Oakland University, Rochester, MI, USA

ARTICLE INFO

Keywords:
Parkinson's disease
Motor outcome prediction
DAT SPECT
Convolutional neural network

ABSTRACT

Purpose: Dopamine transporter (DAT) SPECT imaging is routinely used in the diagnosis of Parkinson's disease (PD). Our previous efforts demonstrated the use of DAT SPECT images in a data-driven manner by improving prediction of PD clinical assessment outcome using radiomic features. In this work, we develop a convolutional neural network (CNN) based technique to predict clinical motor function evaluation scores directly from longitudinal DAT SPECT images and non-imaging clinical measures.

Procedures: Data of 252 subjects from the Parkinson's Progression Markers Initiative (PPMI) database were used in this work. The motor part (III) score of the unified Parkinson's disease rating scale (UPDRS) at year 4 was selected as outcome, and the DAT SPECT images and UPDRS_III scores acquired at year 0 and year 1 were used as input data. The specified inputs and outputs were used to develop a CNN based regression method for prediction. Ten-fold cross-validation was used to test the trained network and the absolute difference between predicted and actual scores was used as the performance metric. Prediction using inputs with and without DAT images was evaluated.

Results: Using only UPDRS_III scores at year 0 and year 1, the prediction yielded an average difference of 7.6 ± 6.1 between the predicted and actual year 4 motor scores (range [5, 77]). The average difference was reduced to 6.0 ± 4.8 when longitudinal DAT SPECT images were included, which was determined to be statistically significant via a two-sample *t*-test, and demonstrates the benefit of including images.

Conclusions: This study shows that adding DAT SPECT images to UPDRS_III scores as inputs to deep-learning based prediction significantly improves the outcome. Without requiring segmentation and feature extraction, the CNN based prediction method allows easier and more universial application.

1. Introduction

Parkinson's disease (PD) is a common neurodegenerative disorder with a variety of symptoms leading to motor function issues such as bradykinesia, resting tremor, rigidity, and postural instability in some cases as the disease progresses further [1]. While the exact cause is unknown, the disease is associated with a loss of dopaminergic neurons in the substantia nigra [2–4]. There is a significant need to find reliable biomarkers of disease progression due to the lack of disease modifying therapies [5]. Because of the heterogeneous nature of the PD symptoms and variability in their progression, there is also much interest in predicting disease outcome to help adapt clinical trials for the proper patient groups.

Clinical trials with neuroimaging have demonstrated the challenges in diagnosing early-stage PD [6–8]. While dopamine transporter (DAT) SPECT imaging (clinically referred to as the DaTscan) has been used extensively for the diagnosis of PD, it also provides valuable information in early disease patients with inconclusive parkinsonian symptoms [9,10]. Being able to use images to predict a patient's outcome with reasonable accuracy would make it easier to determine the long-term efficacy of a treatment by comparing actual to predicted outcomes.

Clinical applications of DAT SPECT images focuses on visual review, while quantitative analysis may help with some of the challenges in early stage detection and improve progression tracking. Radiomics based studies have made progress in a wide variety of clinical applications to improve diagnosis, prognosis, treatment response assessment,

* Corresponding author. Department of Electrical and Computer Engineering, Oakland University, Rochester, MI, USA.
E-mail address: jingtang@gmail.com (J. Tang).

https://doi.org/10.1016/j.compbiomed.2021.104312
Received 14 December 2020; Received in revised form 26 February 2021; Accepted 3 March 2021
Available online 6 March 2021

and progression tracking [11–13]. Similar progress has also been demonstrated for PD, with advances made in the use of DAT SPECT radiomics and other biomarkers to predict cognitive decline and improve diagnosis [14–17]. We previously studied using radiomic features from DAT SPECT images along with non-imaging clinical features to predict future motor function, and also identified the specific imaging and clinical features that were related to the prediction performance [18–23]. These works demonstrated the interest in and feasibility of moving away from visual inspection of images towards data-driven techniques.

While research in using radiomic features to make predictions has demonstrated success, the approach comes with its own challenges. Image features need to be extracted through segmentation and registration. More steps are then needed to identify the features contributing to the prediction performance. Convolutional neural networks (CNNs), drawing inspiration from biological systems such as brain structures, have shown improved performance in image classification and detection problems [24,25]. They could help address the challenges encountered in feature extraction and selection for PD outcome prediction.

CNNs are able to identify relevant features by updating network parameters based on the training data [26]. Supervised learning is used to update the parameters of the receptive fields so that the CNN can find relationships between the input and output without explicitly defining image features [24]. After training, the neural network can then generalize to real-world examples that have not been learned on. This has led deep learning to see its applications to medical images for disease classification [27–29]. A CNN was trained to automatically diagnose PD using DAT SPECT images, resulting in performance similar to clinical standard without feature extraction [30]. The same group who did the aforementioned study also applied deep learning to PET images to predict cognitive decline in Alzheimer's disease patients, demonstrating better performance than a feature-based method [31].

In this study, we develop a CNN based technique to predict a patient's year 4 motor function score directly from the DAT SPECT images, rather than extracted radiomic features, together with non-imaging clinical measures. We believe this CNN based approach is promising because we have demonstrated in previous studies that DAT SPECT image features contributed to cognitive and motor function scores prediction [18,20,21]. We have also shown that a CNN based approach was able to achieve comparable performance to a feature based approach in a task of categorizing PD patients based on their motor function at year 4, using the baseline DAT SPECT images and clinical measures [32]. Taking advantage of the learning and generalization capability of CNNs, the proposed technique is expected to make the patient motor outcome prediction from longitudinal imaging and clinical measures.

2. Materials and methods

2.1. Patient data

Longitudinal data were extracted from the PPMI database (www.ppmi-info.org/data) (Parkinson Progression Marker Initiative, 2011). The unified Parkinson's disease rating scale (UPDRS) – part III (motor) score in year 4 was used as the outcome. Predictors included the DAT SPECT images and UPDRS_III score from both the baseline (year 0) and year 1. These measures were chosen as inputs because our earlier studies showed that they were the most relevant to prediction performance. Other clinical features such as gender, age, MoCA, and disease duration at times of diagnosis and appearance of symptoms were irrelevant [18,21]. A UPDRS_III score was only considered valid if the patient was not on medication at the time of evaluation. Any patient that had all of the aforementioned data available from the PPMI database was included in this study, resulting in 252 PD subjects (baseline age of 62.4 ± 9.7 between 35 and 86; 168 male, 84 female) with year 4 outcome UPDRS_III distribution of 28.7 ± 11.7, ranging between 5 and 77.

All images collected in the PPMI database followed a standard acquisition protocol, to account for different SPECT systems used throughout the sites participating in the study. Imaging was done 4 ± 0.5 h after a dosage of 111–185 MBq radiotracer injection. Thyroid protection was provided by pre-treating subjects with an iodine solution before radiotracer injection. The data, in the form of a 128 × 128 matrix, was acquired over a total scan duration of about 30–45 min by sampling every 3°, 120 projections, 20% symmetric photopeak windows centered on 159 keV and 122 keV. The raw SPECT projection data was reconstructed using the iterative OSEM (ordered-subset expectation maximization) method on a HERMES system (Hermes Medical Solutions, Stockholm, Sweden). PMOD (PMOD Technologies, Zurich, Switzerland) was then used to apply Chang's attenuation correction to the reconstructed data. PMOD was also used for spatial normalization into Montreal Neurological Institute (MNI) space based on a multisite European database of healthy control patients [33]. This resulted in a size of 91 × 109 × 91 voxels for the final images, with voxel size of 2 × 2 × 2 mm^3. As an example, DAT SPECT images of two subjects (a male and a female) taken at years 0 and 1 are shown in Fig. 1. Patient (a) saw his UPDRS_III score progress from 41 to 48 from baseline to year 4 while patient (b)'s score went from 13 to 22 over the same duration. There seem to be image characteristics that could be related to the scores. We expect the method developed in this work to make use of these characteristics.

2.2. Data processing

The DAT SPECT images were processed before being used for prediction. The images, originally of size 91 × 109 × 91, were zero padded to the size of 109 × 109 × 109. The images were also normalized to fall

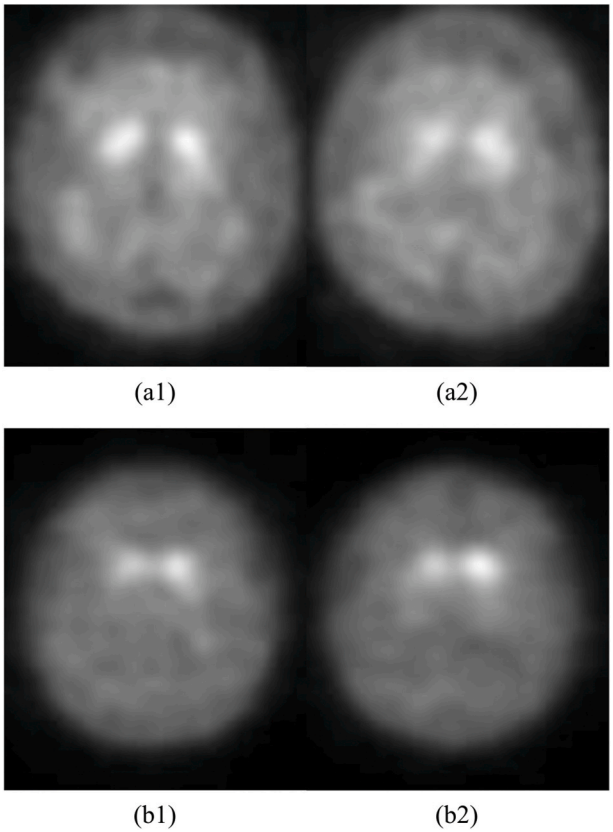

Fig. 1. Example longitudinal DAT SPECT images of two patients (a, b) collected at year 0 (1) and year 1 (2).

within the range of [0, 1] using

$$I_{NORM} = \frac{I_{ORIGINAL} - I_{MIN}}{I_{MAX} - I_{MIN}},$$

where I_{NORM} represents the normalized voxel intensity, and $I_{ORIRINAL}$, I_{MAX}, I_{MIN} are the original, maximum, and minimum intensities of a given image before the normalization. This process was done to keep the feature distribution more consistent after being computed by the CNN, which could lead to faster and more stable learning convergence [34].

We augmented our dataset by left/right flipping all the images to effectively double the size of the dataset. In deep learning applications where the dataset size is limited, augmenting the data has been shown to reduce overfitting by providing a wider variety of samples for training [35]. We chose to only flip *both* year 0 and year 1 images together, rather than allowing for flipping only year 0 or only year 1 images. Even though flipping only year 0 or year 1 images would allow greater augmentation, we would like to ensure that any features that progress from year 0 to year 1 remain in the same relative location. The rationale for using left/right flipping as the method for data augmentation will be provided in more detail in Discussion.

The UPDRS_III score is prone to noteworthy variability between evaluations, even among the same patient. This is largely due to the variability in the patient's symptoms coupled with the subjective nature of the evaluation. The patient receives a score from 0 to 4 for each motor function task based on the rater's assessment of their performance, but the different scores do not have a quantifiable measurement associated with them [36]. This leads to a combination of the patient's performance and the evaluator's assessment to be the key drivers of the patient's score. Other works mitigated this challenge by averaging each score of interest with any available scores 6 months before or after [18]. Our study adopted this approach as well.

2.3. CNN based prediction

The CNN structure we use is shown in Fig. 2. The DAT images form the input layer and the year 4 score is the output. This structure was built from the CNN used in another study [30], since it demonstrated success in classifying patients as having PD or being normal. Assuming that it could be similar features or characteristics that contribute to classification and prediction, we expect that this CNN would serve as a suitable starting point for our task. The hidden layers consist of three 3-D convolutional layers with the filter sizes of $7 \times 7 \times 7$, $5 \times 5 \times 5$, and $3 \times 3 \times 3$ and strides of three, one, and one voxels, and 16, 32, and 64 filter banks, respectively. Rectified linear unit (ReLU) activations are applied after each convolutional layer. Each convolutional layer is followed by a max-pooling layer with size $3 \times 3 \times 3$ and stride of two voxels. The outputs of this series of layers are then flattened into a vector. We added a second parallel set of convolutional layers to the original CNN design so that there is one set of independent operations for each of the year 0 and year 1 images. After the convolutional layers, the year 0 and year 1 UPDRS_III scores are added to the flattened vector from both the year 0 and year 1 series of convolutional layers as scalar values. This flattened layer feeds into three sequential fully-connected layers of 256, 64, and 16 neurons, which are then connected to the output for the year 4 motor function score.

The network was implemented in Tensorflow, a popular framework for machine learning [37]. The loss backpropagation and Adam optimization algorithms were used to update the weight of every neuron connection during training, with cross-entropy used as the cost function [38]. The learning rate was empirically initialized at 0.001 with a total of 50 training iterations. Mean average error was used as the training cost function. The fully connected neuron layers include 0.5 dropout used only for training. The training and testing procedure was repeated to conduct 10-fold cross-validation [39]. The full prediction procedure with 10-fold cross-validation is shown in Fig. 3.

2.4. Evaluation

After completing the 10-fold cross-validation, each predicted year 4 UPDRS_III score was compared to the patient's actual score. The comparison was made based on the absolute difference between the predicted and actual scores. To evaluate whether the CNN can properly identify information from DAT SPECT images that contributes to the prediction accuracy, the cross-validation was repeated both with and without the image data (by setting all voxel values to 0 for the latter). A two sample *t*-test was performed to determine whether there is a statistically significant improvement when the image data is included.

In addition to conducting the experiments with both year 0 and year 1 data as the input, we also performed the training and testing with only input data from year 0 or from year 1. When only using the year 0 data

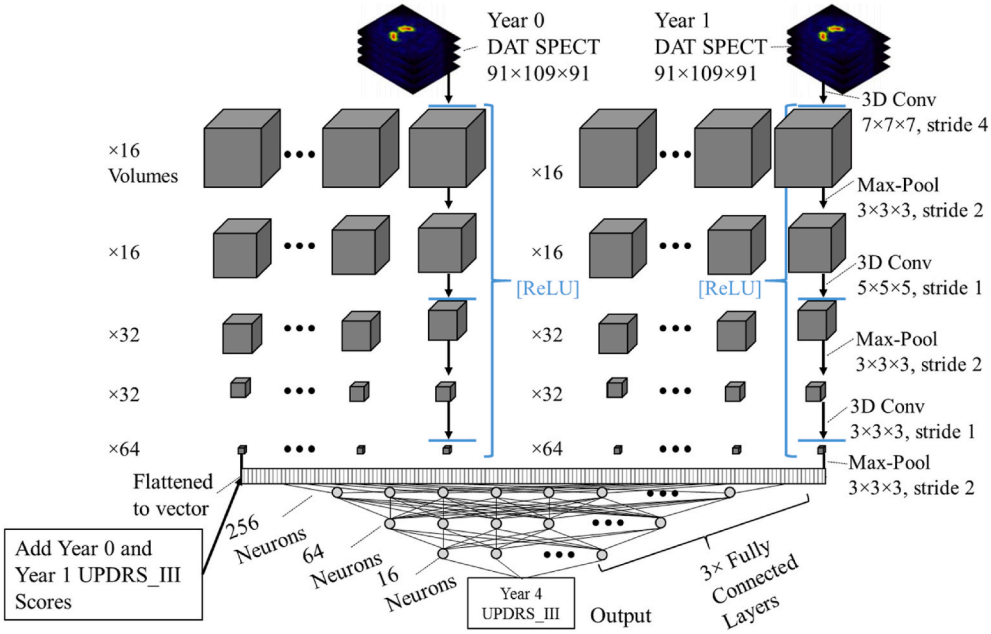

Fig. 2. CNN structure used for prediction of UPDRS_III at year 4, based on longitudinal DAT SPECT images and UPDRS_III scores.

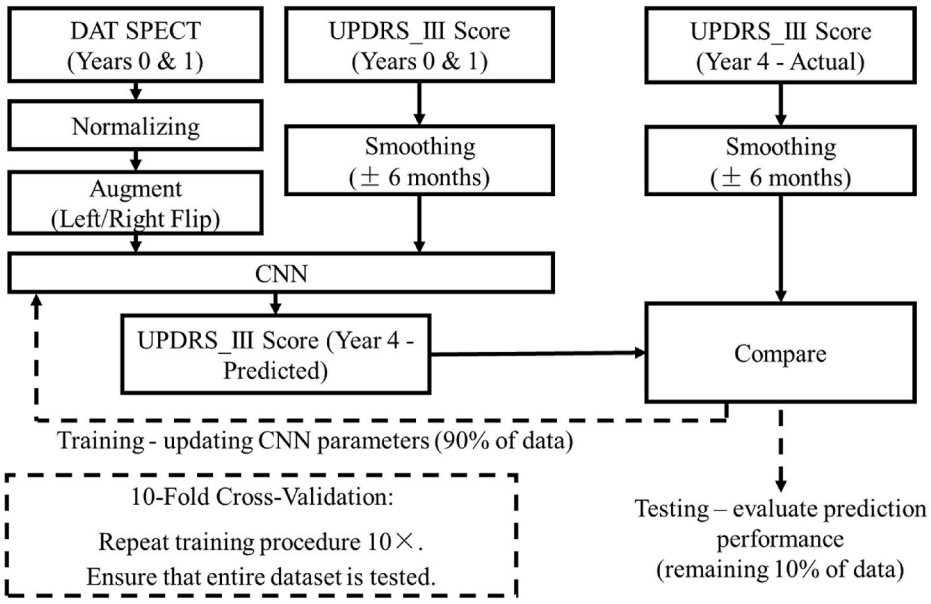

Fig. 3. Training and testing procedure to predict UPDRS_III at year 4 from longitudinal DAT SPECT images and UPDRS_III scores.

the year 1 data is set to all zeros, and vice-versa. This was done to determine whether the year 0 or year 1 data contributed more meaningfully to the prediction performance.

3. Results

Fig. 4 depicts our key findings with the plots showing the predicted versus actual year 4 UPDRS_III scores for both (a) without and (b) with the DAT SPECT images as the input. The UPDRS_III scores at year 0 and year 1 are used in both cases. An ideal plot would be that all the data points fall on the identity line, where the prediction of every subject is equal to his/her actual score. The results shown in (b) apparently follow the identity line more closely than those in (a). The visually appreciable advancement is reflected in the average absolute error among the 252 subjects being 6.0 ± 4.8 in (b) as opposed to 7.6 ± 6.1 in (a). This is also consistent with the result of the two-sample *t*-test showing statistically significant improvement (p-value < 0.001) by adding DAT SPECT images as input to the CNN architecture.

We present in Fig. 5 the results when data from only year 0 (baseline) were used in the outcome prediction for both (a) without and (b) with the DAT SPECT images as the input. The performance in both cases are generally poorer compared with those from also including year 1 data in the input, showing the result cluster either further away from the identity line or more scattered. When adding the year 0 DAT SPECT images to the year 0 UPDRS_III as the input, the prediction results in an average absolute error of (b) 9.3 ± 7.8, as opposed to (a) 12.1 ± 8.9 when images are not included. The two-sample *t*-test shows statistically significant improvement (p-value < 0.001) in prediction when the year 0 DAT SPECT images are used with the year 0 motor function scores. The comparison between the results in (a) and (b) echoes our findings in the last paragraph, i.e., the added DAT SPECT images improve the prediction performance. In addition, the reduction of absolute errors with the year 1 information included shows that the longitudinal information from year 1 is certainly needed for optimal motor outcome prediction.

We plot the results in Fig. 6 when the predictions were performed with data only from year 1 for both (a) without and (b) with the DAT

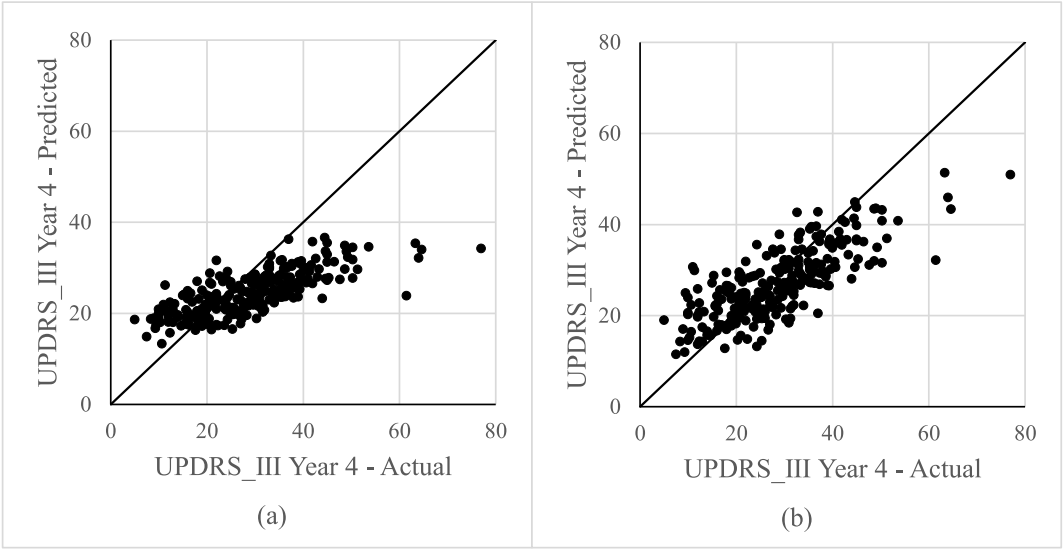

Fig. 4. Predicted versus actual UPDRS_III outcome results at year 4, when using only UPDRS_III scores (a) and when using both UPDRS_III and DAT SPECT images (b). Data from both years 0 and 1 were included.

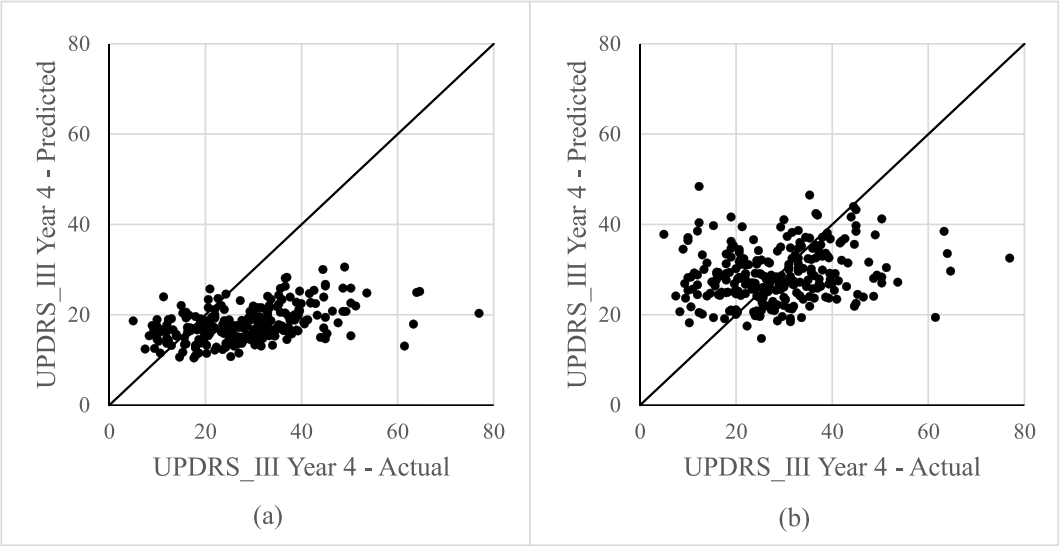

Fig. 5. Predicted versus actual UPDRS_III outcome results at year 4, when using only year 0 UPDRS_III scores as input (a) and when using both UPDRS_III and DAT SPECT images from year 0 as input (b).

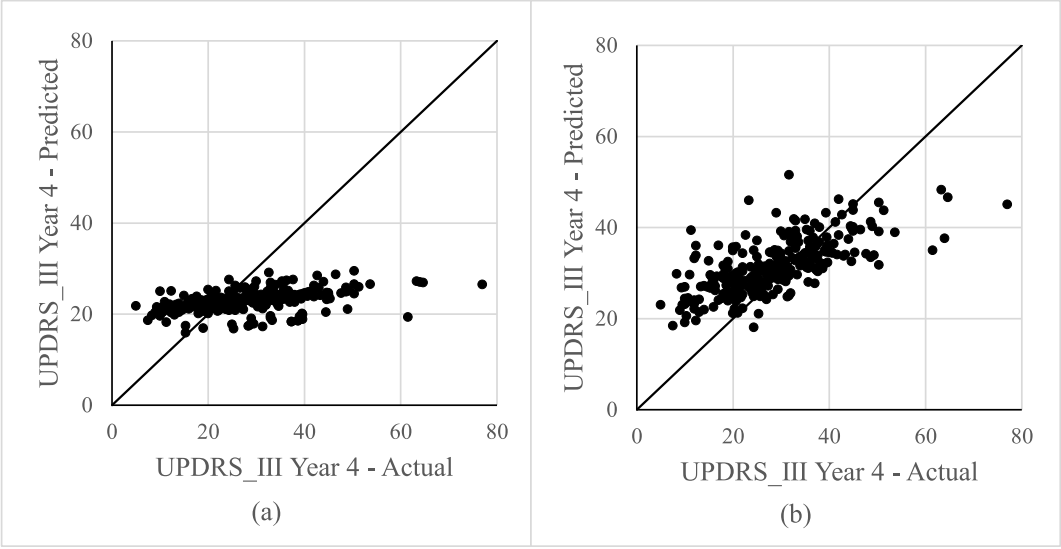

Fig. 6. Predicted versus actual UPDRS_III outcome results at year 4, when using only year 1 UPDRS_III scores as input (a) and when using both UPDRS_III and DAT SPECT images from year 1 as input (b).

SPECT images as the input. The prediction with both the UPDRS_III scores and the DAT SPECT images as the input results in an average absolute error of (b) 7.1 ± 5.8 as opposed to (a) 9.5 ± 7.8 when the DAT SPECT images are not included. With only year 1 input data in prediction, the two-sample *t*-test shows statistically significant improvement (p-value < 0.001) when DAT SPECT images are used alongside motor function scores. While the results are not as good as when the longitudinal data were included, they are better than those when the predictions were conducted with data from year 0. This could be due to the fact that the year 1 data is chronologically closer to the prediction milestone at year 4. However, the longitudinal data from year 0 is still needed for improved prediction performance. This is demonstrated by the two-sample *t*-test showing statistically significant improvement when longitudinal DAT SPECT images and motor function scores are used compared to using only year 0 (p-value < 0.001) or only year 1 (p-value $= 0.02$) data as the input.

4. Discussion

We studied the convergence of the CNN model for the training and testing data (year 0 and year 1) when considering both prediction cases with and without DAT SPECT images, and present the results in Fig. 7. It can be seen in both cases that the training loss (measured in mean average error) converges steadily, whereas the testing loss converges with larger variability. The variability in the testing process, mainly the sharp increase observed in the loss, is likely due to the fact that this study only had 252 subjects available. It is noted that complex deep-learning tasks typically require thousands of subjects for optimal performance and generalization. Ideally, training and testing losses would both track downwards. A small dataset, however, tends to cause CNNs to overfit to the training data while not generalizing as well to the testing data. This manifests as a divergence in loss between the training and testing data. The fact that the testing loss still trends consistently downwards is encouraging, especially given that there was a limited amount of patient data to work with.

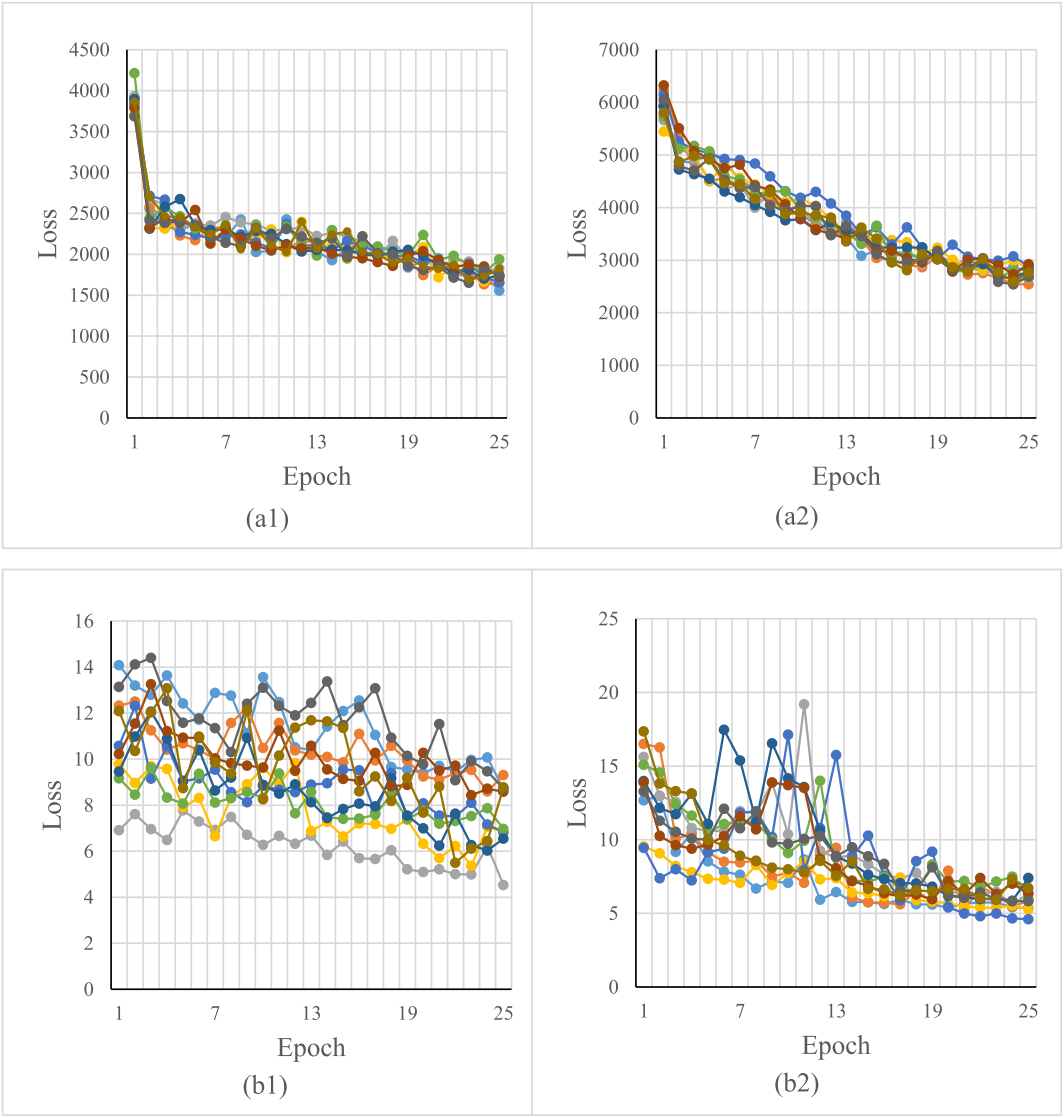

Fig. 7. CNN convergence of training (a) and testing (b) datasets, for prediction cases both without (1) and with (2) DAT SPECT images. Each color signifies a cross-validation fold, and loss is measured in mean average error.

We use loss convergence in optimizing the CNN structure to improve upon our previous work as a baseline [40]. We started off with good prediction performance coupled with poor convergence on the testing data, implying that the model was overfitting. While performance is the main goal, convergence is an important consideration as well since generalizing this method beyond the limited data set used in this study is desirable. By observing the convergence data while modifying the CNN parameters, we found that using a network structure with too many parameters was contributing to the original over-fitting. After reducing the number of neurons in the first, second, and third fully-connected layers from 1024, 256, 64 to 256, 64, 16 respectively, and increasing the stride distance on the first convolutional layer from three to four, the results remained similar while moving to a more generalizable result. This implies that the main reason that the original model did not generalize well was that there were too many parameters relative to the size of the dataset. This issue was resolved by reducing the complexity of the CNN to better match the sample size.

In our method, we left/right flipped the images to generate additional samples to train the CNN model. The goal of data augmentation is to provide training data which are realistically representative of the population but differ from the available data. Flipping the images increases the size of the dataset due to brain symmetry. This operation was shown to have a positive impact in a previous study which used a similarly structured CNN for a related goal with similar data [30]. We also observed that left/right flipping reduced the prediction error in our experiments. Although some scaling and rotation operations were tested, they did not demonstrate improvement. Using only the effective augmentation helps to avoid additional computational cost, which is an important consideration for real-world applications. Therefore, we decided to use only the left/right flipping augmentation in our study.

As mentioned in Introduction, being able to better predict patient outcome can assist in developing more effective clinical trials, while identifying biomarkers that better track disease progression aids in assessing the effectiveness of proposed therapies. This work shows the potential in tackling these challenges through the demonstrated prediction improvements with the DAT SPECT images. In our previous work, we investigated using radiomic features to make predictions about disease progression in PD patients [18,19,23]. These features were confirmed to improve the predictions over using clinical features alone, but there are apparent challenges with these radiomic feature involved approaches. The features relevant to the prediction need to be extracted and identified through extra procedures and thorough experiments. The approach detailed in this and another related work [41] bypass these challenges by using the entire DAT SPECT image directly with CNNs. It

should also be noted that, while the method outlined in this work reduces the steps required to make predictions, the performance is still comparable to that of a previous study on the same problem [18]. This indicates that CNNs effectively interpret the image data without additional operations, using any helpful information from the images to allow prediction without the necessity of identifying which features are relevant. On the other hand, when images are directly used for prediction, what information or features in the images contribute to the performance is not clear. This calls for additional investigation, as finding what radiomic features add to prediction performance and how they help could lead to better understanding of the disease mechanisms. In this front, methods have been shown to identify what features are recognized by the CNNs [42–46]. In future work, we would like to contribute by identifying the features related to PD prediction that can be recognized by a CNN. This would lead to opportunities to identify new biomarkers for the development of disease modifying therapies.

It is very likely that the UPDRS_III score may not be enough on its own to properly track PD progression. Including non-motor metrics would be helpful for assessment of disease progression in clinical trials, such as the Beck Depression Inventory (BDI) and Mini-mental State Exam (MMSE) [47]. We would also like to work on other metrics related to PD prediction including cognitive assessment (MoCA) and time to initiation of symptomatic therapy (TIST) to see if they can be used to make further improvements in motor function prediction. We will build on this work to investigate whether further improvement can be achieved from using these measures in addition to the imaging data and UPDRS_III scores from baseline and year 1. It would also be interesting to determine whether other metrics than motion function can be predicted using the methods outlined in this study.

5. Conclusions

In this work, we designed a CNN that predicts motor function at year 4 using longitudinal (year 0 and year 1) DAT SPECT images and UPDRS_III data. We tracked loss convergence data for the CNN-based approach to ensure that the deep learning method generalizes to test data as well as possible. Our results demonstrated that the addition of DAT SPECT images to using only the UPDRS_III scores as input significantly improved prediction of the UPDRS_III score at year 4. We also found that the predictions made with only year 0 or year 1 data were not as well performed when compared with the case when longitudinal data were included. This deep learning based analysis of DAT SPECT images meets the original goal of showing the potential to improve motor function prediction of PD, while bypassing the need to conduct image segmentation and feature extraction, thus allowing for easier and more universal application.

Declaration of competing interest

There are no conflicts of interest.

Acknowledgements

The project was supported in part by the Michael J. Fox Foundation (Research Grant 2016, ID: 9036.01), including use of data available from the PPMI—a public-private partnership—funded by The Michael J. Fox Foundation for Parkinson's Research and funding partners (listed at www.ppmi-info.org/fundingpartners). This work was also supported in part by the National Science Foundation (grant ECCS-1454552). We gratefully acknowledge the support of NVIDIA Corporation with the donation of the Titan Xp GPU used for this research.

References

[1] A. Kouli, et al., Stoker T.B., Greenland J.C., Parkinson's Disease: Etiology, Neuropathology, and Pathogenesis. Parkinson's Disease: Pathogenesis and Clinical Aspects [Internet], Codon Publications, Brisbane (AU), 2018.
[2] L.V. Kalia, A.E. Lang, Parkinson's disease, Lancet 386 (9996) (2015) 896–912.
[3] G. Chinaglia, et al., Mesostriatal and mesolimbic dopamine uptake binding sites are reduced in Parkinson's disease and progressive supranuclear palsy: a quantitative autoradiographic study using [3H]mazindol, Neuroscience 49 (2) (1992) 317–327.
[4] E.S. Garnett, et al., A rostrocaudal gradient for aromatic acid decarboxylase in the human striatum, Can. J. Neurol. Sci. 14 (3 Suppl) (1987) 444–447.
[5] K. Marek, et al., Biomarkers for Parkison's [corrected] disease: tools to assess Parkinson's disease onset and progression, Ann. Neurol. 64 (Suppl 2) (2008) S111–S121.
[6] Parkinson Study Group, A randomized controlled trial comparing pramipexole with levodopa in early Parkinson's disease: design and methods of the CALM-PD study, Clin. Neuropharmacol. 23 (1) (2000) 34–44.
[7] A.L. Whone, et al., Slower progression of Parkinson's disease with ropinirole versus levodopa: the REAL-PET study, Ann. Neurol. 54 (1) (2003) 93–101.
[8] S. Fahn, Parkinson Study Group, Does levodopa slow or hasten the rate of progression of Parkinson's disease? J. Neurol. 252 (Suppl 4) (2005) IV37–IV42.
[9] S. Thobois, et al., Contributions of PET and SPECT to the understanding of the pathophysiology of Parkinson's disease, Neurophysiol. Clin. 31 (5) (2001) 321–340.
[10] A.M. Catafau, et al., DaTSCAN Clinically Uncertain Parkinsonian Syndromes Study Group, Impact of dopamine transporter SPECT using I-ioflupane on diagnosis and management of patients with clinically uncertain parkinsonian syndromes, Mov. Disord. 19 (10) (2004) 1175–1182.
[11] M. Scrivener, et al., Radiomics applied to lung cancer: a review, Transl. Cancer Res. 5 (4) (2016) 398–409.
[12] Y. Fan, et al., Application of radiomics in central nervous system diseases: a systematic literature review, Clin. Neurol. Neurosurg. 187 (2019) 105565.
[13] J.E. Park, et al., A systematic review reporting quality of radiomics research in neuro-oncology: toward clinical utility and quality improvement using high-dimensional imaging features, BMC Canc. 20 (1) (2020) 29.
[14] F. Yoshii, et al., Combined use of dopamine transporter imaging (DAT-SPECT) and 123I-metaiodobenzylguanidine (MIBG) myocardial scintigraphy for diagnosing Parkinson's disease, J. Neurol. Sci. 375 (2017) 80–85.
[15] S. Lorio, et al., The combination of DAT-SPECT, structural and diffusion MRI predicts clinical progression in Parkinson's disease, Front. Aging Neurosci. 11 (2019) 57.
[16] D. Arnaldi, et al., Prediction of cognitive worsening in de novo Parkinson's disease: clinical use of biomarkers, Mov. Disord. 32 (12) (2017) 1738–1747.
[17] C. Caspell-Garcia, et al., Parkinson's Progression Markers Initiative (PPMI), Multiple modality biomarker prediction of cognitive impairment in prospectively followed de novo Parkinson disease, PloS One 12 (5) (2017) e0175674.
[18] A. Rahmim, et al., Improved prediction of outcome in Parkinson's disease using radiomics analysis of longitudinal DAT SPECT images, Neuroimage: Clinical 16 (2017) 539–544.
[19] J. Tang, et al., Artificial neural network based outcome prediction in DAT SPECT imaging of Parkinson's disease, J. Nucl. Med. 58 (Suppl 1) (2017) 292.
[20] M.R. Salmanpour, et al., Optimized machine learning methods for prediction of cognitive outcome in Parkinson's disease, Comput. Biol. Med. 111 (2019) 103347.
[21] M.R. Salmanpour, et al., Machine learning methods for optimal prediction of motor outcome in Parkinson's disease, Phys. Med. 69 (2020) 233–240.
[22] A. Rahmim, Application of texture analysis to DAT SPECT imaging: relationship to clinical assessments, Neuroimage: Clinical 12 (2016) e1–e9.
[23] J. Tang, et al., Artificial neural network-based prediction of outcome in Parkinson's disease patients using DaTscan SPECT imaging features, Mol. Imag. Biol. 21 (6) (2019) 1165–1173.
[24] Y. LeCun, et al., Deep learning, Nature 521 (2015) 436–444.
[25] A. Khan, et al., A survey of the recent architectures of deep convolutional neural networks, Artif. Intell. Rev. 53 (2020) 5455–5516.
[26] N. Aloysius, M. Geetha, A Review on Deep Convolutional Neural Networks, 2017 International Conference on Communication and Signal Processing (ICCSP), 2017, pp. 588–592.
[27] E.S. Kumar, C.S. Bindu, Medical image analysis using deep learning: a systematic literature review. Emerging Technologies in Computer Engineering: Microservices in Big Data Analytics 985, Springer, Singapore, 2019, pp. 81–97.
[28] A. Maier, et al., A gentle introduction to deep learning in medical image processingEine sanfte Einführung in Tiefes Lernen in der Medizinischen Bildverarbeitung, Z. Med. Phys. 29 (2) (2019) 86–101.
[29] W. Shen, et al., Multi-scale Convolutional Neural Networks for Lung Nodule Classification. Information Processing in Medical Imaging 9123, Springer, Cham, 2015.
[30] H. Choi, et al., Refining diagnosis of Parkinson's disease with deep learning-based interpretation of dopamine transporter imaging, Neuroimage: Clinical 16 (2017) 586–594.
[31] H. Choi, et al., Predicting cognitive decline with deep learning of brain metabolism and amyloid imaging, Behav. Brain Res. 344 (2018) 103–109.
[32] M.P. Adams, et al., Prediction of Outcome in Parkinson's Disease Patients from DAT SPECT Images Using a Convolutional Neural Network, in: 2018 IEEE Nuclear Science Symposium and Medical Imaging Conference Proceedings (NSS/MIC), Sydney, NSW, Australia, 2018, pp. 1–4.
[33] A. Varrone, et al., European multicentre database of healthy controls for [123I]FP-CIT SPECT (ENC-DAT): age-related effects, gender differences and evaluation of

different methods of analysis, Eur. J. Nucl. Med. Mol. Imag. 40 (2) (2013) 213–227.
[34] C. Parmar, et al., Data analysis strategies in medical imaging, Clin. Canc. Res. 24 (15) (2018) 3492–3499.
[35] C. Shorten, T.M. Khoshgoftaar, A survey on image data augmentation for deep learning, J. Big Data 6 (60) (2019).
[36] Movement Disorder Society Task Force on Rating Scales for Parkinson's Disease, The unified Parkinson's disease rating scale (UPDRS): status and recommendations, Mov Disord 18 (7) (2003) 738–750.
[37] M. Abadi, et al., TensorFlow: A System for Large-Scale Machine Learning. 12th USENIX Symposium on Operating Systems Design and Implementation, USENIX Association, 2016, pp. 265–283.
[38] D. Kingma, J. Ba, ADAM: A Method for Stochastic Optimization, 3rd International Conference for Learning Representations, San Diego, 2015.
[39] M. Stone, Cross-validatory choice and assessment of statistical predictions, J. Roy. Stat. Soc. 36 (2) (1974) 111–147.
[40] M.P. Adams, J. Tang, Convolutional Neural Network Based Motor Score Prediction Using DAT SPECT Imaging of Parkinson's Disease Suppl 1, 60, J. Nucl. Med., 2019.
[41] K.H. Leung, et al., Using Deep-Learning to Predict Outcome of Patients with Parkinson's Disease, in: 2018 IEEE Nuclear Science Symposium and Medical Imaging Conference Proceedings (NSS/MIC), Sydney, NSW, Australia, 2018, pp. 1–4.
[42] M.D. Zeiler, R. Fergus, Visualizing and Understanding Convolutional Networks. Computer Vision – ECCV 2014 8689, Springer, Cham, 2014.
[43] Z. Qin, et al., How convolutional neural networks see the world — A survey of convolutional neural network visualization methods, Mathematical Foundations of Computing 1 (2) (2018) 149–180.
[44] F. Grün, et al., A Taxonomy and Library for Visualizing Learned Features in Convolutional Neural Networks, arXiv:1606.07757, 2016.
[45] A. Jamaludin, et al., SpineNet: Automated classification and evidence visualization in spinal MRIs, Med. Image Anal. 41 (2017) 63–73.
[46] F. Ciompi, et al., Towards automatic pulmonary nodule management in lung cancer screening with deep learning, Sci. Rep. 7 (46479) (2017).
[47] P. Huang, et al., The Parkinson Study Group, Using global statistical tests in long-term Parkinson's disease clinical trials, Mov. Disord. 24 (12) (2009) 1732–1739.