3D Convolutional Neural Networks for Classification of Alzheimer's and Parkinson's Disease with T1-Weighted Brain MRI

Nikhil J. Dhinagar*[a], Sophia I. Thomopoulos [a], Conor Owens-Walton [a],
Dimitris Stripelis [b], Jose Luis Ambite [b], Greg Ver Steeg [b],
Daniel Weintraub [c], Philip Cook [d], Corey McMillan [c], Paul M. Thompson [a]
[a]Imaging Genetics Center, Mark & Mary Stevens Institute for Neuroimaging & Informatics, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA
[b]Information Sciences Institute, University of Southern California, Los Angeles, CA, USA
[c]University of Pennsylvania School of Medicine, University of Pennsylvania, Philadelphia, PA, USA
[d]Penn Image Computing and Science Laboratory, Department of Radiology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA

ABSTRACT

Parkinson's disease (PD) and Alzheimer's disease (AD) are progressive neurodegenerative disorders that affect millions of people worldwide. In this work, we propose a deep learning approach to classify these diseases based on 3D T1-weighted brain MRI. We analyzed several datasets including the Parkinson's Progression Markers Initiative (PPMI), an independent dataset from the University of Pennsylvania School of Medicine (UPenn), the Alzheimer's Disease Neuroimaging Initiative (ADNI), and the Open Access Series of Imaging Studies (OASIS) dataset. PPMI and ADNI were partitioned to train (70%), validate (20%), and test (10%) a 3D convolutional neural network (CNN) for PD and AD classification. The UPenn and OASIS datasets were used as independent test sets to evaluate the model performance during inference. We also implemented a random forest classifier as a baseline model by extracting key radiomics features from the same T1-weighted MRI scans. The proposed 3D CNN model was trained from scratch for the classification tasks. For AD classification, the 3D CNN model achieved an ROC-AUC of 0.878 on the ADNI test set and an average ROC-AUC of 0.789 on the OASIS dataset. For PD classification, the proposed 3D CNN model achieved an ROC-AUC of 0.667 on the PPMI test set and an average ROC-AUC of 0.743 on the UPenn dataset. We also found that model performance was largely maintained when using only 25% of the training dataset. The 3D CNN outperformed the random forest classifier for both the PD and AD tasks. The 3D CNN also generalized better on unseen MRI data from different imaging centers. Our results show that the proposed 3D CNN model was less prone to overfitting for AD than for PD classification. This approach shows promise for screening of PD and AD patients using only T1-weighted brain MRI, which is relatively widely available. This model with additional validation could also be used to help differentiate between challenging cases of AD and PD when they present with similarly subtle motor and non-motor symptoms.

Keywords: Parkinson's disease, Alzheimer's disease, deep learning, MRI, CNN, image analysis, brain, classification

1. INTRODUCTION

Alzheimer's disease (AD) and Parkinson's disease (PD) are the two most common neurodegenerative diseases in the world. According to the World Health Organization (WHO), around 50 million people have dementia worldwide, and AD likely contributes to 60-70% of these cases.[1] A further study showed that 6.1 million people were affected by PD globally in 2016.[2] Structural magnetic resonance imaging (MRI) offers the potential to screen individuals for diagnostic biomarkers of PD and AD. Standard volumetric T1-weighted brain MRI is widely used and accessible world-wide, and has the advantage of not subjecting the patient to ionizing radiation[3] compared to other modalities, such as PET.

*Corresponding author: Nikhil J. Dhinagar, email: dhinagar@usc.edu

Recently, a multi-center study was conducted by the PD working group of the Enhancing Neuroimaging Genetics through Meta-Analysis (ENIGMA) consortium on a cohort of 2,357 PD patients and 1,182 healthy controls using T1-weighted brain MRI scans from 19 different international centers.[4] The authors analyzed cortical thickness, regional surface area, and volumes for subcortical regions of the brain revealing a distributed pattern of atrophy that advances as the disease progresses. Similar work in AD has stemmed from large cohort studies such as the Alzheimer's Disease Neuroimaging Initiative (ADNI), which has published multiple works demonstrating significant changes in brain metrics in the disorder.[5] While important, such case-control comparisons only identify distinctive features at the group level, and do not offer a means to classify disease status in new individuals. More than 80% of PD patients develop cognitive deficits; this is relatively common in prodromal and *de novo* PD. Therefore, differentiating early AD from early is a key clinical issue.

Machine learning and deep learning models have been developed for various neuroimaging applications, including image reconstruction and enhancement, segmentation and labeling of regions of interest, and disease classification and subtyping. When designing models for these tasks, several factors must be considered, including data availability, the size and efficiency of the neural network architecture, and its performance on diverse datasets other than those used to train the algorithm. Many studies have shown promise in applying convolutional neural networks (CNNs) to brain MRI scans for regression tasks such as brain age prediction, a commonly used benchmarking task that attempts to predict a person's age from their brain scan.[6,7] Other works have proposed methods for classification of neurodegenerative diseases. Islam and Zhang proposed a 2D model based on the Inception-V4 architecture for classification of AD.[8] Sivaranjini and Sujatha[9] proposed a CNN based on the AlexNet architecture for PD classification. Many of the methods proposed previously however make predictions in a slice-wise manner. Yagis et al. address potential data leakage issues in some of the earlier neuroimaging works, several of which performed slice-based predictions.[10] Also, some of the prior 2D methods may not account for the spatial relationships of all the slices in the 3D MRI volume. West et al. show that 3D CNNs can outperform their 2D counterparts for some neuro-imaging based classification tasks.[3]

In a recent innovation, Lu et al.[11] pooled over 85,000 brain MRI scans to build a CNN that classified a person's sex from their brain scan, and they fine-tuned the network for AD classification using transfer learning. In our own prior work,[12,13] we used a recurrent neural network (RNN) to aggregate information from an ordered series of 2D slice-based CNNs to predict a person's age from their 3D MRI scan (see also Peng et al.[14]), and we later extended the same method using attention modules to perform AD classification.[6] Despite promising performance, we did not test the method on additional cohorts beyond the one used to train the algorithm, nor did we test the method on multiple diseases, which perhaps represents a more clinically relevant application.

Contributions: In this paper, we propose a 3D CNN that extracts features from 3D T1-weighted brain MRI scans and classifies PD and AD per subject. The key contribution of this work includes:

- A 3D CNN architecture for classification of PD and AD using T1-weighted volumetric brain MRI.
- We tested a random forest-based classifier with well-known radiomics features as a baseline model. The 3D CNN outperforms the baseline and generalizes better.
- We evaluated our models on multiple datasets, including external datasets from different centers to independently test our proposed model.
- We show the model performs the classification task reasonably well when trained on a subset of the training data.

2. DATA

2.1 Datasets

The datasets used in this work are summarized in **Table 1** and discussed in this text. We analyzed 3D volumetric brain MRI data from the Parkinson's Progression Markers Initiative (PPMI) dataset (N = 471; 341 individuals with PD/130 healthy age-matched controls; 308M/163F; age range: 30.6-84.9 years) and an independent dataset from the University of Pennsylvania (UPenn) as the out-of-distribution test dataset (N = 117; 104 PD/13 controls; 75M/42F; age: 50-84) for

classification of PD. The PPMI collects baseline scans for untreated patients and the UPenn is a convenience cohort of treated PD patients. For classification of AD we used non-accelerated 3D T1-weighted scans from the Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset (n = 531, 243 people diagnosed with probable AD and 288 healthy controls (HC); 268M/263F; age: 55.7–91.4 years) and the Open Access Series of Imaging Studies (OASIS) dataset (n = 234; 46 with AD and 188 HC; 78M/156F; age: 33-96) based on OASIS1.

Table 1. Summary of the datasets we analyzed in this paper (Parkinson's Progression Markers Initiative (PPMI), University of Pennsylvania (UPenn), Alzheimer's Disease Neuroimaging Initiative (ADNI), Open Access Series of Imaging Studies (OASIS))

Dataset	Total MRI (n)	Healthy Controls	PD/AD Subjects	Male/Female	Age Range
PPMI	471	130	341 (PD)	308M/163F	30.6-84.9 years
UPenn	117	13	104 (PD)	75M/42F	50.0-84.0 years
ADNI	531	243	288 (AD)	268M/263F	55.7–91.4 years
OASIS	234	188	46 (AD)	78M/156F	33.0-96.0 years

2.2 Image Pre-processing

The T1-weighted brain MRI volumes were pre-processed[15] using a sequence of steps, including nonparametric intensity normalization (N4 bias field correction),[16] 'skull stripping' for brain extraction,[17] linear registration to a template with 6 degrees of freedom and isometric voxel resampling to 2 mm.[18] The pre-processed images are of size 91x109x91. The T1-weighted MR images were scaled to take values between 0 and 1 using the min-max scaling approach for the 3D CNN model.

2.2 Data Preparation

We created non-overlapping data partitions for training, validation and testing of the proposed model. 10% of PPMI (PD) and 10% of ADNI (AD) were used for testing for the classification tasks. We created balanced subsets of the UPenn (PD) and OASIS (AD) datasets for independent testing. We used the rest of the ADNI and PPMI dataset for training (includes 20% of these datasets as a validation set for the 3D CNN).

3. METHODS

3.1 Random Forest Classifier

We implemented a machine learning pipeline as a baseline for AD and PD classification, as seen in **Figure 1**. The model uses the T1-weighted MRI volume and a whole-brain mask as the input. The whole brain mask is created to focus

feature extraction on the region within the brain volume. 93 radiomics features including textural features based on first-order statistical co-occurrence matrices were extracted using PyRadiomics.[19] The features were selected based on recursive feature elimination, min-max scaled and used to train a random forest classifier.

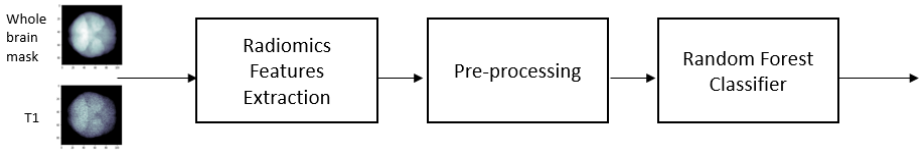

Figure 1. Machine Learning Pipline for Neurodegenerative Disease Prediction

3.2 3D CNN Model Architecture

The architecture of the 3D CNN model[20] is illustrated in **Figure 2**. The input to this model is 91x109x91. The CNN consists of 4 modules consisting of 3 layers: convolutional (conv), max-pooling (maxpool) and batch normalization (BN) for feature extraction. The convolutional layers have 64, 64, 128, and 256 filters respectively. The feature extractor block is connected by a global average pooling layer to a fully connected (FC) layer with 512 neurons, a dropout layer and a fully connected layer with a sigmoid activation function for the binary classification problem. Binary cross-entropy loss is used as the loss function for the CNN.

Figure 2. Architecture of the 3D Convolutional Neural Network (CNN)

3.3 3D CNN Model Training and Hyperparameter Optimization

All layers of the CNN were trained on the PPMI and ADNI training data for PD/AD classification. The ROC-AUC and the loss were used to ensure convergence during each of the training epochs. The hyperparameters of the CNN were optimized based on a random search approach. The key hyperparameters that are tuned included the optimizer (Adam,[21] Adam with weight decay, stochastic gradient descent), learning rate (2e-3 to 1e-5), learning rate scheduling, dropout (0 to 0.5), use of bias initializer for class imbalance, early stopping, training epochs, and batch size.

The hyperparameters used for classification of AD included training for 50 epochs, dropout of 0.2, batch size of 4, an Adam optimizer with learning rate of 2e-4, and weight decay of 1e-4. The hyperparameters used for classification of PD included training for 50 epochs with early stopping if there is no improvement in validation loss for 30 epochs, dropout of 0.2, batch size of 4, an Adam optimizer with learning rate of 2e-4, and weight decay of 1e-4. We used small random image rotations with angles ε {-20, -10, -5, 5, 10, 20} degrees to augment the data. We initialize a bias in the final output layer equal to the logarithm of the number of samples in class 0 divided by the number of samples in class 1.[22, 23] This is to alleviate the imbalance in the PPMI dataset. The model with best validation performance was used for testing on the test sets for the two classification tasks.

3.4 Model Testing

To evaluate the models, we used standard performance metrics including the receiver-operator characteristic curve-area under the curve (ROC-AUC), the precision-recall area under the curve (PR-AUC), accuracy, precision, recall and F1-score. The Youden index[24] was used to optimize the threshold from the ROC curve. The performance metrics were calculated as an average on all the balanced subsets of the UPenn and OASIS1 datasets.

4. RESULTS

Below we show results of the proposed 3D CNN for classification of PD and AD with T1-weighted MRI data.

4.1 Alzheimer's Disease Classification

The training, validation ROC, and classification loss curves are shown in **Figure 3** for the 3D CNN. The validation curves closely follow the training performance as seen in the figure. The performance of the 3D CNN and random forest classifier on the ADNI test set and the OASIS dataset are summarized in terms of the different metrics in **Table 2**. The ROC curves for the 3D CNN, calculated from these test sets, are shown in **Figure 4**. The curves in different colors in the plot on the right in **Figure 4** and **Figure 7** represent the ROC curves for the balanced subsets from the corresponding test set.

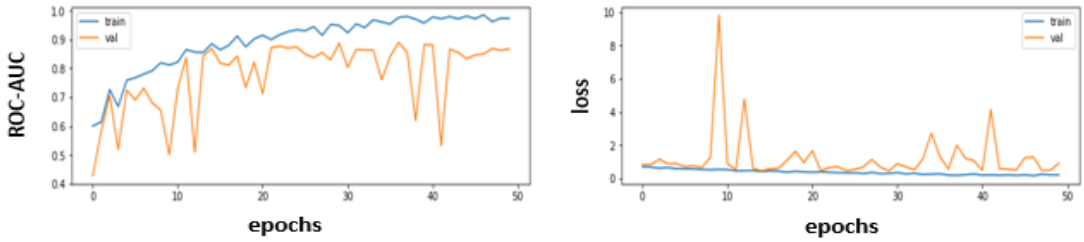

Figure 3. Training and validation curves for ROC-AUC (left) and loss (right) for AD classification

The random forest classifier achieved an ROC-AUC of 0.730 on the ADNI test set and 0.558 on the OASIS data. The model yielded a 0.646 PR-AUC, 0.722 precision, 0.591 recall and 0.650 F1 score on the ADNI test set. When tested on the OASIS dataset, the model achieved 0.556 average PR-AUC with a standard deviation of 0.05, 0.575 average precision with a standard deviation of 0.04, 0.603 average recall with a standard deviation of 0.1, and 0.580 average F1 score with a standard deviation of 0.04. The 3D CNN achieved an ROC-AUC of 0.878 on the ADNI test set and 0.789 OASIS test set. The model yields a 0.816 PR-AUC, 0.808 precision, 0.808 recall and 0.808 F1 score on the ADNI test set. When tested on the OASIS dataset the model achieved a 0.789 average PR-AUC with a standard deviation of 0.03, 0.724 average precision with a standard deviation of 0.05, 0.793 average recall with a standard deviation of 0.05, and 0.755 average F1 score with a standard deviation of 0.02. Results show that our 3D CNN outperforms the baseline random forest for AD classification and tends to generalize better on an independent dataset.

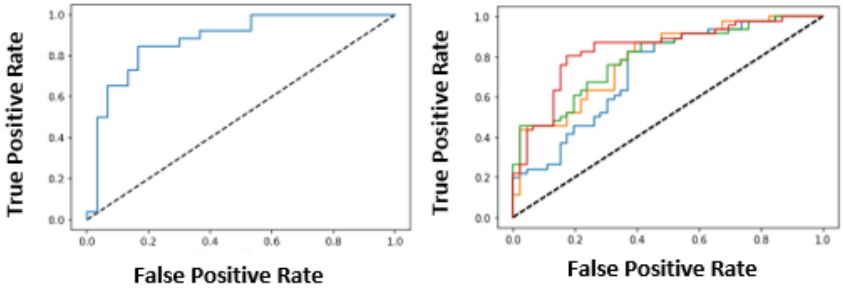

Figure 4. ROC curve on the ADNI test set (left) and ROC curves for each of the balanced subsets from the OASIS dataset (right)

In **Table 3,** we summarize the effect of training data size on the test performance of the model. The ADNI and OASIS datasets are used to evaluate the model to this end. Different fractions of the ADNI dataset were used for training, ranging from just 10% to 100% of the data. We also evaluated the effect of using data augmentation during training. **Figure 5** captures the model performance trend for ADNI and OASIS with and without data augmentation.

Table 2. Performance of Random Forest and 3D CNN for AD classification

Metric	ADNI		OASIS (Average of balanced subsets with the standard deviation)	
	Random Forest	3D CNN	Random Forest	3D CNN
ROC- AUC	0.730	**0.878**	0.558 (0.038)	**0.789 (0.038)**
PR AUC	0.646	**0.816**	0.556 (0.048)	**0.795 (0.042)**
Accuracy	0.741	**0.821**	0.571 (0.018)	**0.742 (0.036)**
Precision	0.722	**0.808**	0.575 (0.041)	**0.724 (0.055)**
Recall	0.591	**0.808**	0.603 (0.105)	**0.793 (0.039)**
F1-score	0.650	**0.808**	0.580 (0.037)	**0.755 (0.029)**

Table 3. Effect of Training Data size on Model Performance

ADNI Training Data	10% (n =37)	25% (n=92)	50% (n=185)	75% (n=278)
ADNI test ROC AUC	0.460	0.696	0.701	0.803
ADNI test ROC AUC [w/o data augmentation]	0.496	0.695	0.739	0.796
OASIS Average ROC AUC (SD)	0.612 (0.040)	0.763 (0.040)	0.628 (0.040)	0.735 (0.050)
OASIS Average ROC AUC (SD) [w/o data augmentation]	0.647 (0.043)	0.603 (0.042)	0.788 (0.027)	0.789 (0.047)

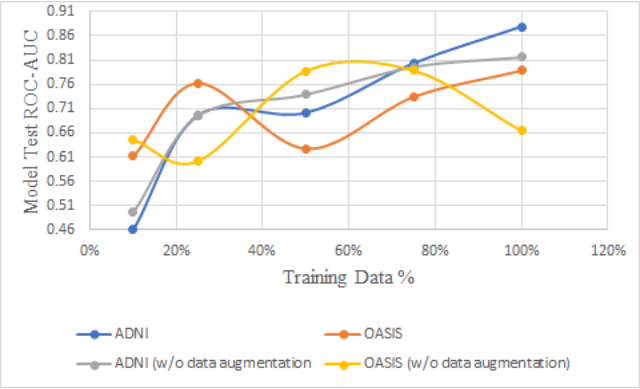

Figure 5. Model test ROC-AUC (ADNI test, OASIS) versus the Training Dataset size (ADNI), where on the y-axis we show performance for models trained on 10% (n=37), 25% (n=92), 50% (n=185), 75% (n=278), and 100% (n=371) of the available training data. The performance trend shows that the model benefits from using additional training data and may continue to improve as more training datasets become available. Data augmentation helps the model reduce overfitting during the training stage.

4.2 Parkinson's Disease Classification

The training, validation ROC and classification loss curves are shown in **Figure 6** for the 3D CNN for PD classification. The performance of the 3D CNN and random forest classifier during inference on the PPMI test set and the UPenn dataset are summarized in terms of the different metrics in **Table 4**. **Figure 7** shows the ROC curves for the 3D CNN on the PPMI test set, and over each balanced subset of the UPenn dataset.

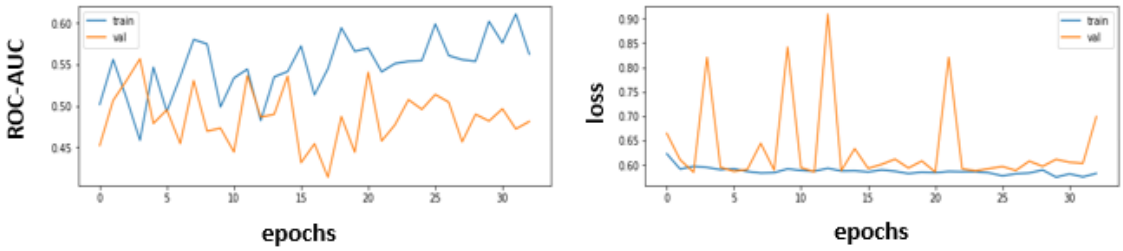

Figure 6. Training and validation curves for ROC-AUC (left) and loss (right) for PD classification

The random forest classifier achieved an ROC-AUC of 0.524, 0.815 PR-AUC, 0.864 precision, 0.487 recall, and 0.623 F1 score on the PPMI test set. The random forest when tested on the UPenn test set achieved an average ROC-AUC of 0.534 with a standard deviation of 0.07, 0.556 average PR-AUC with a standard deviation of 0.05, 0.575 average precision with a standard deviation of 0.04, 0.603 average recall with a standard deviation of 0.1 and 0.580 average F1 score with a standard deviation of 0.04. The 3D CNN achieved an ROC-AUC of 0.667, 0.812 PR-AUC, 0.839 precision, 0.706 recall and 0.776 F1 score on the PPMI test set. The model when tested on the UPenn test set achieved an average ROC-AUC of 0.743 with a standard deviation of 0.11, 0.777 average PR-AUC with a standard deviation of 0.1, 0.816 average precision with a standard deviation of 0.08, 0.560 average recall with a standard deviation of 0.21 and 0.631 average F1 score with a standard deviation of 0.18. Results show that our 3D CNN outperforms the baseline random forest for PD classification and generalizes better when tested on an independent dataset.

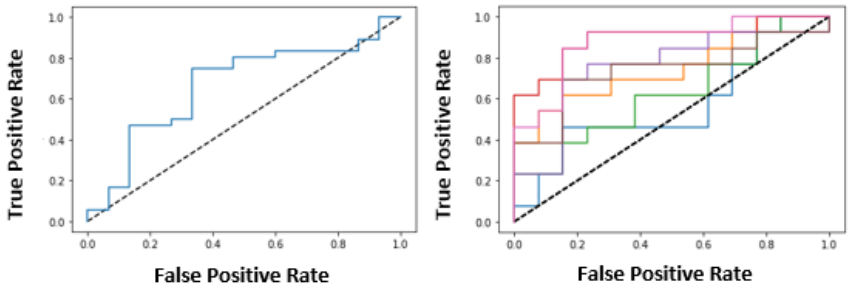

Figure 7. ROC curve on the PPMI test set (left) and ROC curves for each of the balanced subsets from the UPenn dataset (right)

Table 4. Performance of Random Forest and 3D CNN for PD classification

Metric	PPMI		UPenn (Average of balanced subsets with the standard deviation)	
	Random Forest	3D CNN	Random Forest	3D CNN
ROC AUC	0.524	**0.667**	0.534 (0.066)	**0.743 (0.111)**
PR AUC	0.815	**0.812**	0.520 (0.067)	**0.777 (0.102)**
Accuracy	0.521	**0.706**	0.567 (0.066)	**0.709 (0.082)**
Precision	0.864	**0.839**	0.563 (0.071)	**0.816 (0.085)**
Recall	0.487	**0.722**	0.673 (0.248)	**0.560 (0.208)**
F1-score	0.623	**0.776**	0.581 (0.148)	**0.631 (0.177)**

5. DISCUSSION AND FUTURE WORK

In this paper, we present a 3D CNN that outperforms a baseline random forest method for the classification of AD and PD from T1-weighted volumetric brain MRI. We presented results comparing the 3D CNN's performance with the baseline method, for both tasks, in section 4. The 3D CNN achieved an ROC-AUC of 0.878 and 0.667 for the AD (10% of ADNI) and PD (10% of PPMI) classification. The 3D CNN yields an average ROC-AUC of 0.789 and 0.784 for the AD (OASIS) and PD (UPenn) classification. The standard deviation on the UPenn test (0.11) for the 3D CNN versus on the OASIS test set (0.03), suggests that the model tuned for the AD task is relatively more stable than on the PD task. This can also be observed via the ROC curves on the right of **Figures 4** and **7**. The random forest classifier achieved an ROC-AUC of 0.730 and 0.524 for the AD (10% of ADNI) and PD (10% of PPMI) classification. The random forest yielded an average ROC-AUC of 0.558 and 0.534 for the AD (OASIS) and PD (UPenn) classifications. The 3D CNN also generalized better than the random forest classifier on the independent test sets.

In **Table 3**, we show how the size of the training dataset affects the model's performance during inference. As expected, the model performance degrades with a smaller training dataset. Even so, the model performs reasonably well with only 25% of the ADNI training data, i.e., 92 subjects. **Figure 5** indicates that data augmentation could be beneficial to the model's performance by reducing overfitting during training.

Empirically, the 3D CNN was more prone to overfitting on the PPMI data for PD classification than ADNI for AD classification. Our PPMI dataset was imbalanced for the two classes, i.e., the class ratio was 0.72(PD):0.28(HC), whereas ADNI was relatively well balanced, i.e., 0.46(AD):0.54(HC). A larger dataset with more balance between the classes might improve performance and reduce overfitting during training for PD classification. The 3D CNN proposed here is a relatively efficient architecture (~1.3 million parameters) compared to other typical 3D deep neural networks. We plan to perform additional experiments with 2D variants as this might also help to reduce overfitting issues for low data problems such as the tasks defined here.

6. CONCLUSIONS

The experiments and results presented here highlight the potential of the 3D CNN model to classify neurodegenerative diseases based on T1-weighted brain MRI. The 3D CNN's performance was robust when trained only on a subset of the complete dataset. The 3D CNNs outperformed the baseline random forest classifier. We showed some differences in

training and test performance for the proposed model in PD and AD classification tasks. When trained on more data and with additional validation the proposed pipeline could help in non-invasively screening individuals for PD and AD from structural MRI data. Even classifiers with moderate performance may be useful screening tools, prior to more invasive procedures, such as PET scans and CSF assays, which have been used to define biological AD and to help diagnose PD. Future work will also assess the added value of additional imaging protocols, such as quantitative MRI,[25] which may boost classification performance. We plan to test the added value of MRI data harmonization approaches based on generative adversarial networks, such as CycleGANs[26] or CALAMITI, which can make MRI data from different scanners or protocols more comparable.

Disclosures

Dr. McMillan receives research funding outside the submitted work from Biogen, Inc and provides consulting services for personal fees from Invicro and Axon Advisors on behalf of Translational Bioinformatics, LLC. He also receives an honorarium as Associate Editor of NeuroImage: Clinical. Dr. Thompson receives research funding outside the submitted work from Biogen, Inc. and has received consulting fees from Kairos Venture Capital. Dr. Weintraub has received research funding or support from Michael J. Fox Foundation for Parkinson's Research, Alzheimer's Therapeutic Research Initiative (ATRI), Alzheimer's Disease Cooperative Study (ADCS), the International Parkinson and Movement Disorder Society (IPMDS) the National Institute on Aging (NIA) and the U.S. Department of Veterans Affairs; honoraria for consultancy from Acadia, Aptinyx, CHDI Foundation, Clintrex LLC (Otsuka), Eisai, Great Lake Neurotechnologies, Janssen, Sage, Scion, Signant Health, Sunovion and Vanda; and license fee payments from the University of Pennsylvania for the QUIP and QUIP-RS.

Acknowledgements

This work was supported by in part by the U.S. National Institutes of Health, under NIH grants U19 AG062418, R01 NS107513, P01 AG066597 (Imaging Core), U01 AG068057, by DARPA under Agreement HR00112090104, and by a Zenith Award from the U.S. Alzheimer's Association.

REFERENCES

[1] World Health Organization., "Dementia," 2020, <https://www.who.int/news-room/fact-sheets/detail/dementia> (24 July 2021).

[2] Collaborators, G. 2016 P. D., "Global, regional, and national burden of Parkinson's disease, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016 GBD," Lancet Neurol. **17**(11), 939–953 (2018).

[3] West, C., Soltaninejad, S. and Cheng, I., [Assessing the capability of deep-learning models in Parkinson's disease diagnosis], Springer International Publishing (2020).

[4] Laansma, M., Bright, J., Al-Bachari, S., Anderson, T., Ard, T., Assogna, F., Baquero, K., Berendse, H., Blair, J., Cendes, F., Dalrymple-Alford, J., de Bie, R., Debove, I., Dirkx, M., Druzgal, J., Emsley, H., Garraux, G., Guimarães, R., Gutman, B., et al., "An International Multicenter Analysis of Brain Structure across Clinical Stages of Parkinson's Disease: The ENIGMA-Parkinson's Study" (2020).

[5] Jack, C. R., Barnes, J., Bernstein, M. A., Borowski, B. J., Brewer, J., Clegg, S., Dale, A. M., Carmichael, O., Ching, C., Decarli, C., Desikan, R. S., Fennema-notestine, C., Fjell, A. M., Fletcher, E., Fox, N. C., Gunter, J., Gutman, B. A., Holland, D., Hua, X., et al., "Magnetic resonance imaging in Alzheimer's Disease Neuroimaging Initiative 2," Alzheimer's Dement. **11**(7), 740–756 (2015).

[6] Lam, P., Zhu, A. H., Gari, I. B., Jahanshad, N. and Thompson, P. M., "3D Grid-Attention Networks for Interpretable Age and Alzheimer's Disease Prediction from Structural MRI," arXiv (2020).

[7] Gupta, U., Lam, P. K., Steeg, G. Ver and Thompson, P. M., "Improved brain age estimation with slice-based set networks," Proc. - Int. Symp. Biomed. Imaging **2021-April**, 840–844 (2021).

[8] Islam, J. and Zhang, Y., "A Novel Deep Learning Based Multi-class Classification Method for Alzheimer's Disease Detection Using Brain MRI Data," Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics) **10654 LNAI**, 213–222 (2017).

[9] Sivaranjini, S. and Sujatha, C. M., "Deep learning based diagnosis of Parkinson's disease using convolutional

neural network," Multimed. Tools Appl. **79**(21–22), 15467–15479 (2020).
[10] Yagis, E., De Herrera, A. G. S. and Citi, L., "Generalization Performance of Deep Learning Models in Neurodegenerative Disease Classification," Proc. - 2019 IEEE Int. Conf. Bioinforma. Biomed. BIBM 2019, 1692–1698 (2019).
[11] Lu, B., Li, H., Chang, Z., Li, L., Chen, N., Zhou, H., Li, X., Wang, Y. and Cui, S., "A Practical Alzheimer Disease Classifier via Brain Imaging-Based Deep Learning on 85 , 721 Samples" (2021).
[12] Lam, P. K., Santhalingam, V., Suresh, P., Baboota, R., Zhu, A. H., Thomopoulos, S. I., Jahanshad, N. and Thompson, P. M., "Accurate brain age prediction using recurrent slice-based networks," 4–13 (2020).
[13] Gupta, U., Stripelis, D., Lam, P. K., Thompson, P. M. and Steeg, G. Ver., "Membership Inference Attacks on Deep Regression Models for Neuroimaging," MIDL **1996**, 1–21 (2021).
[14] Peng, H., Gong, W., Beckmann, C. F., Vedaldi, A. and Smith, S. M., "Accurate brain age prediction with lightweight deep neural networks," Med. Image Anal. **68**, 101871 (2021).
[15] Jenkinson, M., Beckmann, C. F., Behrens, T. E. J., Woolrich, M. W. and Smith, S. M., "FSL," Neuroimage **62**, 782–790 (2012).
[16] Tustison, N. J., Cook, P. A. and Gee, J. C., "N4ITK: Improved N3 Bias Correction," IEEE Trans Med Imaging. **29**(6), 1310–1320 (2010).
[17] Isensee, F., Schell, M., Pflueger, I., Brugnara, G., Bonekamp, D., Neuberger, U., Wick, A., Schlemmer, H. P., Heiland, S., Wick, W., Bendszus, M., Maier-Hein, K. H. and Kickingereder, P., "Automated brain extraction of multisequence MRI using artificial neural networks," Hum. Brain Mapp. **40**(17), 4952–4964 (2019).
[18] Tustison, N. J., Cook, P. A., Klein, A., Song, G., Das, S. R., Duda, J. T., Kandel, B. M., van Strien, N., Stone, J. R., Gee, J. C. and Avants, B. B., "Large-scale evaluation of ANTs and FreeSurfer cortical thickness measurements," Neuroimage **99**, 166–179 (2014).
[19] Griethuysen, J. J. van., "Computational Radiomics System to Decode the Radiographic Phenotype," Arter. Thromb Vasc Biol **118**(24), 6072–6078 (2016).
[20] Zunair, H., Rahman, A., Mohammed, N. and Cohen, J. P., "Uniformizing Techniques to Process CT Scans with 3D CNNs for Tuberculosis Prediction," Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics) **12329 LNCS**, 156–168 (2020).
[21] Kingma, D. P. and Ba, J. L., "Adam: A method for stochastic optimization," 3rd Int. Conf. Learn. Represent. ICLR 2015 - Conf. Track Proc., 1–15 (2015).
[22] Karpathy, A., "A Recipe for Training Neural Networks," 2019, <http://karpathy.github.io/2019/04/25/recipe/#2-set-up-the-end-to-end-trainingevaluation-skeleton--get-dumb-baselines> (24 July 2021).
[23] Tensorflow., "Classification on imbalanced data," Tensorflow, <https://www.tensorflow.org/tutorials/structured_data/imbalanced_data> (21 July 2021).
[24] Youden, W. J., "Index for rating diagnostic tests," Cancer **3**(1), 32–35 (1950).
[25] Lambert, C., "Mapping subcortical nuclei using qMRI: Applications for neurodegenerative disorders," Organ. Hum. Brain Mapp. (2021).
[26] Liu, M., Maiti, P., Thomopoulos, S., Zhu, A., Chai, Y., Kim, H. and Jahanshad, N., "Style Transfer Using Generative Adversarial Networks for Multi-Site MRI Harmonization," bioRxiv Prepr. (2021).