Differentiating Alzheimer's Disease from Cognitively Normal Individuals Using Convolutional Neural Networks: A Reproducible Study

Elina Thibeau-Sutre[1], Camille Brianceau[2], and Ninon Burgos[2]

[1] Department of Applied Mathematics, Technical Medical Centre, University of Twente, Enschede, The Netherlands
e.thibeau-sutre@utwente.nl
[2] Sorbonne Université, Institut du Cerveau - Paris Brain Institute - ICM, CNRS, Inria, Inserm, AP-HP, Hôpital de la Pitié Salpêtrière, F-75013, Paris, France
ninon.burgos@cnrs.fr

Abstract. Alzheimer's disease (AD) is a progressive neurodegenerative disorder that affects millions of individuals worldwide. Early and accurate diagnosis is crucial for effective intervention and patient care. This study aims to develop a reproducible deep learning model based on convolutional neural networks (CNNs) to differentiate between AD patients and cognitively normal participants using brain MRI scans.
The chosen CNN model is not able to appropriately differentiate AD patients from cognitively normal (CN) participants. The attribution maps associated to the trained network highlighted regions known to be affected by the disease (medial temporal regions).
This reproducible study questions the potential of convolutional neural networks in differentiating Alzheimer's disease patients from cognitively normal individuals based on brain MRI scans and a possible clinical application. The open-source code used in this study is made available to facilitate further research and ensure transparency and reproducibility in the field of neuroimaging-based AD diagnosis.

Keywords: Alzheimer's disease · Deep Learning · Magnetic Resonance Imaging.

1 Introduction

Alzheimer's disease (AD) affects over 20 million people worldwide. Neuroimaging provides useful information to identify AD [1], such as the atrophy due to gray matter loss with anatomical magnetic resonance imaging (MRI). A major interest is then to analyze those markers to identify AD at an early stage. Machine learning and deep learning methods have the potential to assist in identifying patients with AD by learning discriminative patterns from neuroimaging data [2].

As the most widely used architecture of deep learning, convolutional neural networks (CNN) have attracted huge attention thanks to their great success in image classification [3]. Contrary to conventional machine learning, deep learning allows the automatic abstraction of low-to-high level latent feature representations. Thus, one can hypothesize that deep learning depends less on image preprocessing and requires less prior on other complex procedures, such as feature selection, resulting in a more objective and less bias-prone process [4].

The purpose of this paper is to explain the results of a deep learning network trained to differentiate Alzheimer's disease patients from cognitively normal participants. The source code for the experiments and models described in this paper will be made available on GitHub and is attached to this submission during the review process.

2 Materials

In this study, we utilized data from the Open Access Series of Imaging Studies dataset[3]. This cohort includes 416 participants categorized into two distinct groups: CN (Cognitively Normal) or AD (Alzheimer's Disease) [5]. This cohort covers a large age range of 18 to 96. To facilitate our analysis, we further subdivided the CN group into two distinct subgroups based on age. The CN_{old} subgroup consists of participants with a minimum age of 62 and the CN_{young} subgroup includes participants who are strictly younger than 62.

The CN_{old} subjects have an average age of 78.8 and a mean MMSE score of 28.8, while the CN_{young} subjects have an average age of 28.8. As for the AD subjects, they have an average age of 79.0 and a mean MMSE score of 26.5. Clinical scores were not available for the younger participants involved in this study, but one could hypothesize that the mean MMSE score of the CN_{young} subjects is its maximal value (30).

3 Methods

3.1 Preprocessing of T1-weighted MRI

The OASIS dataset underwent curation and conversion into the Brain Imaging Data Structure (BIDS) format [6] using the Clinica software (v0.7.6) [7,8]. For preprocessing the T1-weighted MR images, we applied the `t1-linear` pipeline from Clinica [7], which acts as a wrapper for the ANTs software (v2.4.4) [9]. Bias field correction was applied using the N4ITK method [10]. To facilitate cross-subject comparisons, we performed an affine registration of the images to the MNI space using ANTs [11]. Subsequently, the registered images were rescaled to standardize the intensity values within the minimum and maximum ranges. To remove background noise, we cropped the images, resulting in a final size of $169 \times 208 \times 179$ with isotropic 1 mm voxels. To ensure the reproducibility of our results, we set the random seed of ANTs to 42.

[3] Access the OASIS dataset here: `https://oasis-brains.org`

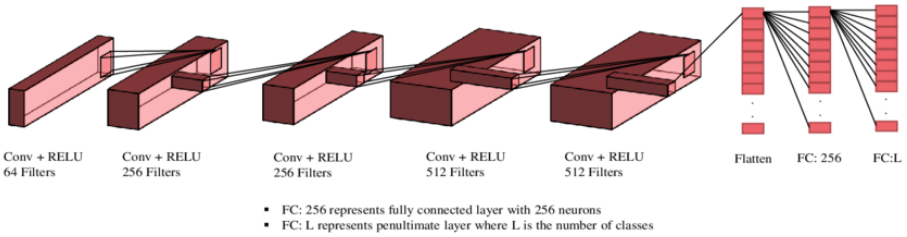

Fig. 1: Architecture of the 3D subject-level CNN. For each convolutional block, we only display the convolutional and max pooling layers. Conv: convolutional layer;RELU: Rectified Linear Unit; FC: fully connected layer.

3.2 Deep learning network

Architecture In order to fully leverage the spatial information present in the MR images, we opted for a 3D CNN architecture, as illustrated in Figure 1. This architecture consisted of 5 convolutional blocks and 3 fully connected (FC) layers. Each convolutional block is a sequence of one convolutional layer, one batch normalization layer, one ReLU activation layer, and one max-pooling layer.

Training The network was trained to differentiate AD from CN participants by optimizing the cross-entropy loss criterion. The weights of the network were optimized using Adam optimizer ($\beta 1 = 0.9$, $\beta 2 = 0.999$, $\epsilon = 10^{-8}$, no weight decay) with a learning rate of 3.31e-3, during 100 epochs. A batch size of 8 is used. At the end of each epoch the loss of the network is computed on the validation set. The best network corresponds to the one which obtained the lowest validation loss value during the training process.

3.3 Evaluation strategy

Explainability In this study, we decided to focus on using the gradient back-propagation method for generating attribution maps, as it is both widely utilized and conceptually straightforward. These attribution maps essentially capture the gradients of an output node concerning an image. In our specific case, this output node corresponds to the CN group. Essentially, the pixel intensities within an attribution map indicate the changes required to transform a given image into a sample that resembles the CN group.

To create a group attribution map for AD patients, we took the mean value of the 10 attribution maps derived from our dataset's AD patients. However, it's important to note that this voxel-based approach often produces somewhat noisy outputs. To address this, we applied a Gaussian filter with a standard deviation of $\sigma = 2$ to the group attribution map, which helped enhance the visualization of relevant regions for our analysis.

4 Results

The performance of the network is lower when considering only the old population compared to using the whole CN group. Indeed 8 old CN participants on 10 are classified as AD patients (see Table 1).

Table 1: Confusion matrix of the CNN.

	AD	CN
AD	9	1
CN (old)	8	2
CN (young)	0	10

Both clinical scores are correlated with the probability of the diagnosis. However the strongest correlation is with age (correlation coefficient = 0.87). The sex is not correlated with the diagnosis.

In Figure 2, one can see that the network prominently highlights the medial temporal lobe, a region known to be atrophied in Alzheimer's disease. Nevertheless, on the central slices (75 & 95), the network's attention extends beyond the brain, including areas adjacent to the cerebellum.

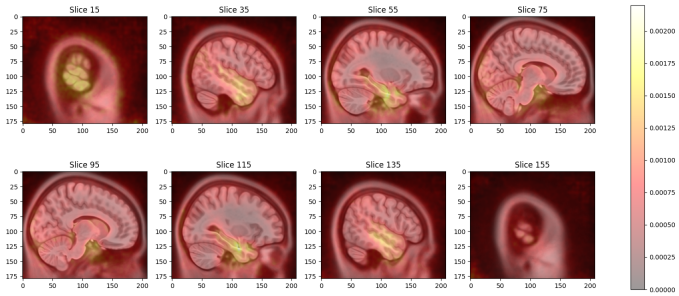

Fig. 2: Group attribution map of AD patients.

5 Conclusion

While our network's outputs exhibit some notable associations with traditional clinical scores like MMSE and CDR, and the attribution map effectively spotlights regions known to be atrophied in AD, a particularly problematic discovery

is the robust link we've uncovered with age, a well-established contributor to brain atrophy and related changes.

Our upcoming research endeavors will explore why the attribution map includes regions beyond the typically implicated medial temporal areas. Furthermore, we'll focus on understanding how these additional regions might be linked to age detection, further enriching our comprehension of the multifaceted landscape of AD diagnosis.

References

1. Ewers, M., Sperling, R.A., Klunk, W.E., Weiner, M.W., Hampel, H.: Neuroimaging markers for the prediction and early diagnosis of Alzheimer's disease dementia. Trends in Neurosciences **34**(8) (2011) 430–442
2. Wen, J., Thibeau-Sutre, E., Diaz-Melo, M., Samper-González, J., Routier, A., Bottani, S., Dormont, D., Durrleman, S., Burgos, N., Colliot, O.: Convolutional neural networks for classification of Alzheimer's disease: Overview and reproducible evaluation. Medical Image Analysis **63** (2020) 101694
3. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet Classification with Deep Convolutional Neural Networks. In: Advances in neural information processing systems. Volume 25. (2012) 1097–1105
4. LeCun, Y., Bengio, Y., Hinton, G.: Deep learning. Nature **521**(7553) (2015) 436–444
5. Marcus, D.S., Wang, T.H., Parker, J., Csernansky, J.G., Morris, J.C., Buckner, R.L.: Open Access Series of Imaging Studies (OASIS): Cross-sectional MRI Data in Young, Middle Aged, Nondemented, and Demented Older Adults. Journal of Cognitive Neuroscience **19**(9) (2007) 1498–1507
6. Gorgolewski, K.J., Auer, T., Calhoun, V.D., Craddock, R.C., Das, S., Duff, E.P., Flandin, G., Ghosh, S.S., Glatard, T., Halchenko, Y.O., Handwerker, D.A., Hanke, M., Keator, D., Li, X., Michael, Z., Maumet, C., Nichols, B.N., Nichols, T.E., Pellman, J., Poline, J.B., Rokem, A., Schaefer, G., Sochat, V., Triplett, W., Turner, J.A., Varoquaux, G., Poldrack, R.A.: The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. Scientific Data **3** (2016) 160044
7. Routier, A., Burgos, N., Díaz, M., Bacci, M., Bottani, S., El-Rifai, O., Fontanella, S., Gori, P., Guillon, J., Guyot, A., Hassanaly, R., Jacquemont, T., Lu, P., Marcoux, A., Moreau, T., Samper-González, J., Teichmann, M., Thibeau-Sutre, E., Vaillant, G., Wen, J., Wild, A., Habert, M.O., Durrleman, S., Colliot, O.: Clinica: An Open-Source Software Platform for Reproducible Clinical Neuroscience Studies. Frontiers in Neuroinformatics **15** (2021) 39
8. Samper-González, J., Burgos, N., Bottani, S., Fontanella, S., Lu, P., Marcoux, A., Routier, A., Guillon, J., Bacci, M., Wen, J., Bertrand, A., Bertin, H., Habert, M.O., Durrleman, S., Evgeniou, T., Colliot, O.: Reproducible evaluation of classification methods in Alzheimer's disease: Framework and application to MRI and PET data. NeuroImage **183** (2018) 504–521
9. Avants, B.B., Tustison, N.J., Stauffer, M., Song, G., Wu, B., Gee, J.C.: The Insight ToolKit image registration framework. Frontiers in Neuroinformatics **8** (2014) 44
10. Tustison, N.J., Avants, B.B., Cook, P.A., Zheng, Y., Egan, A., Yushkevich, P.A., Gee, J.C.: N4ITK: improved N3 bias correction. IEEE Trans. Med. Imaging **29**(6) (2010) 1310–1320

11. Avants, B.B., Epstein, C.L., Grossman, M., Gee, J.C.: Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. Medical Image Analysis **12**(1) (2008) 26–41