Expert Systems With Applications 195 (2022) 116622

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

A deep ensemble hippocampal CNN model for brain age estimation applied to Alzheimer's diagnosis

Katia Maria Poloni, Ricardo José Ferrari *, for the Alzheimer's Disease Neuroimaging Initiative[1]

Department of Computing, Federal University of São Carlos, Rod. Washington Luis, Km 235, 13565-905, São Carlos, SP, Brazil

ARTICLE INFO

Keywords:
Brain-age estimation
Age biomarker
Alzheimer's disease
Mild cognitive impairment
Deep Learning
Convolutional Neural Networks

ABSTRACT

Age-associated diseases rise as life expectancy increases. The brain presents age-related structural changes across life, with different extends between subjects and groups. During the development of neurodegenerative diseases, these changes are more intense and accentuated. As Alzheimer's disease (AD) develops, the brain reflects accelerated aging with minor extends associated with mild cognitive impairment (MCI), i.e., the prodromal stage of AD. Therefore, it is crucial to understand a healthy brain aging process to predict a cognitive decline. This study produced an efficient age estimation framework using only the hippocampal regions that explores the associations of the brain age prediction error of age-matched cognitively normal (CN) subjects with AD and MCI subjects. For this, we have developed two convolutional neural networks. The first achieved very competitive state-of-the-art metrics, i.e., mean absolute error (MAE) of 3.31 and root mean square error (RMSE) of 4.65. The second has also achieved competitive metrics, but more importantly, we founded a statistically significant analysis of our delta estimation error between the compared groups. Further, we correlated our results with clinical measurements, e.g., Mini-Mental State Examination (MMSE) score, and obtained a significant negative correlation. In addition, we compared our results with other published studies. Therefore, our findings suggest that our delta could become a biomarker to support AD and MCI diagnosis.

1. Introduction

The human brain exhibits a biologically complex process of age-related changes across life (Cole & Franke, 2017a) characterized by region-specifics and non-linear patterns of coordinated and sequenced events during development (Cherubini, Caligiuri, Péran, Sabatini, Cosentino, & Amato, 2016) and with a general decline in cognitive performance, causing generalized atrophy with aging (Cole et al., 2017; Resnick, Pham, Kraut, Zonderman, & Davatzikos, 2003). Thus, age advancement is associated with an increased prevalence of brain diseases, mostly neurodegenerative, as Alzheimer's disease (AD), Parkinson's disease, and amyotrophic lateral sclerosis (Cole et al., 2017).

The onset of age-associated diseases varies in a wide age range; consequently, the brain aging effects are distinct among subjects. Schizophrenia, for instance (Schnack, Van Haren, Nieuwenhuis, Hulshoff Pol, Cahn, & Kahn, 2016), affects much younger subjects than Alzheimer's (Feng, Lipton, Yang, Small, & Provenzano, 2020). Therefore, it is crucial to advance our understanding of healthy brain aging, as this may help us identify biomarkers to predict cognitive decline related to neurodegenerative diseases (Cole et al., 2017).

Brain age prediction from neuroimaging data and using machine learning and computer vision techniques has been widely studied and proposed in several different methods (Cole & Franke, 2017a; Dinsdale et al., 2021; Feng et al., 2020; Huang et al., 2017; Ito et al., 2018; Jiang et al., 2020; Pardakhti & Sajedi, 2019, 2020; Peng, Gong, Beckmann, Vedaldi, & Smith, 2021; Ueda et al., 2019) and has increasingly provided insights on the effects of age-associated brain changes and how diseases affect the aging brain (Cole et al., 2017). The age predicted from these models is considered the actual biological brain age due to being estimated from whole-brain imaging data (Cole & Franke, 2017a). Therefore, the delta generated between the predicted age (brain age) and the actual age can be used as a biomarker for the early identification and support for the diagnosis of age-related brain disorders. A positive delta, for instance, implies that the subject's brain looks older than their actual age, which could indicate accelerated brain aging (Franke, Ziegler, Klöppel, Gaser, Initiative,

* Corresponding author.
E-mail addresses: katiampoloni@gmail.com, kpoloni@estudante.ufscar.br (K.M. Poloni), rferrari@ufscar.br (R.J. Ferrari).
URL: http://www.bipgroup.dc.ufscar.br (R.J. Ferrari).

[1] Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf.

https://doi.org/10.1016/j.eswa.2022.116622
Received 26 October 2021; Received in revised form 22 December 2021; Accepted 27 January 2022
Available online 9 February 2022

et al., 2010). Neurodegenerative diseases, such as AD, cause more intense and accentuated structural changes than cognitively normal (CN) aging (Guadalupe et al., 2014; Woolard & Heckers, 2012).

One of the main challenges in the development of a brain age estimation CNN model, with application in the prediction of Alzheimer's diseases, is the reduced number of MR images available for an elderly population (age over 75 years). The reason is the difficulty to acquire the images of debilitate patients, since they are required to remain still during the MRI scan. For the purpose of training prediction CNN models, the reduced number of images limits the stratification of the population into narrower age groups, which is essential to provide accurate age predictions.

Magnetic resonance imaging (MRI) has been successfully used to capture different tissue, structure, and function information of the human body (Dinsdale et al., 2021). Due to their high contrast of soft tissues and good spatial resolution, structural MRI allows visualizing details and subtle changes in brain tissues (Alzheimer's Association, 2020; Johnson, Fox, Sperling, & Klunk, 2012; Luo & Tang, 2017) and has been used in several studies on age prediction (Cole & Franke, 2017a; Dinsdale et al., 2021; Feng et al., 2020; Huang et al., 2017; Ito et al., 2018; Jiang et al., 2020; Pardakhti & Sajedi, 2019, 2020; Peng et al., 2021; Ueda et al., 2019). In these studies, the most commonly used pulse sequence is the T1-weighted as it is the most informative about the brain structure, especially the depiction of the main anatomical structures and tissues (i.e., grey matter (GM), white matter (WM), and cerebrospinal fluid (CSF)) (Dinsdale et al., 2021; Miller et al., 2016).

Due to the increase in life expectancy and, consequently, age-related diseases, there are efforts worldwide seeking neurocognitive treatments and early diagnosis (Organization et al., 2019), since most of these diseases have limited treatment options, creating a financial and social burden on society (Cole et al., 2017). Alzheimer's disease is the most common form of dementia that accounts for up to 80% of all cases (Alzheimer's Association, 2020). It is an irreversible, cureless, and progressive disease that predominantly affects the elderly population and has become one of the most significant health problems globally (Alzheimer's Association, 2020; Zhang, Liu, An, Gao, & Shen, 2017).

The clinical diagnosis of AD is a challenging evaluation process that follows the clinical criteria defined by the National Institute of Neurological and Communicative Disorders and Stroke and the Alzheimer's Disease and Related Disorders Association (NINDS-ADRDA) (Beach, Monsell, Phillips, & Kukull, 2012) and requires the elimination of other potential causes of dementia (Hyman et al., 2012; McKhann et al., 2011). The accuracy of the clinical diagnosis of AD compared to the autopsy, for instance, reported imprecision rates between 12% and 23% (Beach et al., 2012; Klatka, Schiffer, Powers, & Kazee, 1996; Lim et al., 1999), considering patients diagnosed with AD who did not have enough pathologies at autopsy to explain the presence of dementia (Gaugler, Ascher-Svanum, Roth, Fafowora, Siderowf, & Beach, 2013). This is probably due to clinical symptoms such as memory loss appearing subtly in an AD antecedent phase, known as mild cognitive impairment (MCI), which makes difficult the early clinical diagnosis. Researchers believe this problem can be overcome by administering future treatments developed to slow or halt AD progression and preserve brain function at the onset of brain disease disorders (Ardekani, Hadid, Blessing, & Bachman, 2019). During AD development, the brain shows measurable atrophies that indicate AD signs but with different intensities, being the hippocampus, a brain structure located in the medial temporal lobe, the first brain structure to experience such changes (Guadalupe et al., 2014; Shi, Liu, Zhou, Yu, & Jiang, 2009). The hippocampus, predominantly composed of the GM tissue, plays a decisive role in forming and retaining episodic memory (Ardekani et al., 2019).

Computer-aided systems from medical imaging data have advanced significantly since the emergence of new machine learning techniques. The brain age prediction has been proposed with different techniques as relevance vector machines (RVM) (Franke et al., 2010; Fujimoto et al., 2017; Madan & Kensinger, 2018), support vector machines (SVM) (Pardakhti & Sajedi, 2017; Su, Wang, Shen, & Hu, 2011), Gaussian process regression (Cole & Franke, 2017a), and deep learning, mostly with convolutional neural networks (CNNs) (Cole & Franke, 2017a; Dinsdale et al., 2021; Huang et al., 2017; Ito et al., 2018; Peng et al., 2021; Ueda et al., 2019). These studies usually estimate the brain age within an extensive and unbalanced age range (≈between 20 and 80 years), which contributes to increasing the number of images available for training the algorithms, especially in deep learning. However, if the data is not balanced with the age range, it could lead to a biased analysis if applied to compare with specific diseases. Pardakhti and Sajedi (2020) investigated the effects of Alzheimer's on the brain; nevertheless, they used a model trained with subjects aged between 20 and 80 years and do not provide information about the distribution of the age used for the AD comparison. Thus, a proper analysis of the results is unfeasible since there may be a bias due to older age instead of developing the disease. Furthermore, they used only metrics based on absolute values in the paper, which are insufficient to infer an overestimated age for brains with AD.

Deep learning models have shown quite promising results (Cole & Franke, 2017a) for the brain age prediction task since they can learn and represent features from the images with smaller loss function values (Baumgartner et al., 2019; Peng et al., 2021). However, deeper networks still face several challenges in medical imaging, e.g., small training data availability, the need for more GPU memory for processing 3D data, and the unavailability of 3D pre-trained models. The researchers have been mitigating these issues by downsampling the input (Korolev, Safiullin, Belyaev, & Dodonova, 2017), using small patches (Kamnitsas et al., 2017; Liu, Zhang, Adeli, & Shen, 2018), or image portions (Dinsdale et al., 2021; Pardakhti & Sajedi, 2020) and 2D slices (Huang et al., 2017; Lin et al., 2018). However, it is difficult to use these constraints without performance/information loss.

In addition to CNNs, ensemble learning has been shown to be useful in medical imaging analysis since it can overcome the challenge of training classifiers in situations of limited data availability and the common 3D nature of medical imaging data (Dong, Yu, Cao, Shi, & Qianli, 2020; Logan et al., 2021). A handful number of studies have been published in the literature using ensemble learning on Alzheimer's classification. Giovannetti et al. (2021) use temporal, multi-frequency, and spatial data from MEG recordings and MRI scans in the form of functional connectivity (FC) maps that were incorporated into deep features by using transfer learning. In their model, the authors use an ensemble learning architecture to cooperatively combine the decision of multiple predictive modules on the basis of different FC mapping. Ahmed, Kim, Lee, and Jung (2020) proposed an ensemble of ROI-based CNN classifiers for staging the Alzheimer disease spectrum (preclinical AD, mild cognitive impairment due to AD, and dementia due to AD and normal controls) using magnetic resonance imaging. The authors used patches extracted from the three MRI orthogonal views of selected cerebral regions to learn CNNs. Despite the simplicity and efficiency of processing 2D slices, subtle changes in the analyzed brain structures may go undetected. Although not directly related to Alzheimer's diagnosis, He, Shao, Zhong, and Zhao (2020) used ensemble transfer CNNs driven by multi-channel signals for fault diagnosis of rotating machinery cross working conditions. In their work, the authors train a series of CNNs, modified with stochastic pooling and Leaky rectified linear unit (LReLU), using multichannel signals. Then, the learned parameters of each CNN are transferred to initialize the corresponding target CNN, which is then fine-tuned using a few target training samples. Finally, an ensemble learning strategy is designed to fuse each individual target CNN to obtain the final result.

Other deep learning particularities for the task include the use of a CNN architecture based on modified versions of the VGG-13 (Simonyan & Zisserman, 2014) with three predominant changes: (i) addition of

batch normalization layers (Ioffe & Szegedy, 2015), (ii) conversion of all 2D operations to 3D and (iii) replacement of the softmax activation function by a linear function, considering it is a regression problem (Huang et al., 2017; Ito et al., 2018; Jiang et al., 2020; Peng et al., 2021). Concerning the other hyperparameters, the mean absolute error (MAE) was mainly used as a loss function, as well as the L2 regularization (weight decay) and optimization using SGD (Wijnhoven & de With, 2010) with momentum. Additionally, they used a learning rate decaying as a constant throughout the epochs or by reaching the plateau, and data augmentation transformations, with reports of increased training speed. Moreover, the models needed to be fully trained due to the unavailability of pre-trained weights for the 3D data. Therefore, transfer learning is barely mentioned.

Based on the exposure, the main research goal of this study is to explore the associations of brain age delta (i.e., the mean error (ME), which we will refer to as a delta ($\Delta_{\text{brain\_age}}$ = Estimated age – Chronological age)) of age-matched CN subjects with AD and MCI subjects. To reach our goal, we combined the recent CNN improvements with two datasets containing CN, MCI, and AD subjects and designed an experiment for brain age prediction using only the hippocampal regions, i.e., a hippocampal age prediction. The hippocampal atrophy directly relates to normal aging for many factors, e.g., the GM volume decreases during adulthood (Good, Johnsrude, Ashburner, Henson, Friston, & Frackowiak, 2001; Resnick et al., 2003) and progressively decreases in AD due to increased neuronal cell death (Guadalupe et al., 2014; Woolard & Heckers, 2012). Further, there is some criticism about condensing the whole-brain information into a single number (Cole et al., 2017), making regional analyses more attractive. For instance, Li, Liu, Wang, Wang, Xu, and Qiu (2017) have reported significant findings on this task using only a regression model with the two hippocampus volumes as attributes.

In this study we designed two training experiments. The first uses subjects between 20 and 70 years and is focused on the brain age prediction task. The second also aims at the age prediction, although we used subjects with ages larger than 70 years since our goal is to perform an age-matched comparison between the CN age predictions with AD and MCI. We performed the two sets of experiments for the two hippocampal regions, i.e., left and right, and ensemble each model prediction to create the final prediction. The main contributions of this study can be summarized as:

- the development of a hippocampal age estimation method using an efficient 3D CNN architecture with an end-to-end framework to process new images within less than seven minutes.
- introduction of modifications to the existent EfficientNet architecture: (i) all 2D network operations were changed to 3D, (ii) the input image resolution was changed to fit our data and, (iii) the final output layer was changed to a linear activation function to predict age as the output scalar.
- propose of a transfer learning strategy that uses the CNN weights from the full-training of our first set of experiments (patients with ages between 20 and 70 years), since there are no 3D pre-trained models available, to fine-tuning with a small population (patients older than 70 years) the target CNN in our second set of experiments to overcome the small number of images available with cognitive diseases.
- design of several data augmentation combined with an oversample over the age bins to obtain an even distribution and therefore to achieve network convergence and unbiased results. All results were evaluated with quantitative and qualitative assessments using statistical tests to support our findings and indirectly compared to other works.
- statistical analysis using Pearson pairwise correlation to verify the correlation between MMSE score (clinical value) and the predicted age delta, to corroborate our findings with medical scores.
- make available to the community the pre-trained hippocampal-based weights of four 3D models, left and right hippocampus aged between 20–70 years and larger than 70 years.

This paper is organized as follows: Section 2 describes the methods used in our proposed framework. Section 3 presents the results, and Section 4 presents the discussions. Finally, Section 5 concludes the paper.

2. Material and methods

Fig. 1 illustrates the general framework of the proposed method. As previously mentioned, we designed two training experiments for brain age prediction using only the hippocampal regions; the first experiment used images of patients with ages between 20 and 70 years, and the second, patients older than 70 years. We divided the experiments by age to reach our primary research goal, which is to explore the associations of brain age delta of age-matched CN subjects with AD and MCI subjects. The first training has two goals: (i) serve as a baseline to compare our findings with existing models, and (ii) provide pre-trained weights for the second, since we have fewer images for larger ages. The second uses the first model's pre-trained weights, allowing us to perform a training experiment with elderly CN subjects. Then, we performed an age-matched and unbiased comparison between the CN age predictions with AD and MCI and assessed the age prediction differences between the three groups using the ANOVA statistical test to verify statistically significant differences between each pair of diagnosis groups. Both training experiments used the same CNN architecture and left and right ensemble hippocampal design, as illustrated.

2.1. Dataset

We used MR images from the Neuroimage Analysis Center (NAC) (Halle et al., 2017), the Information Extraction from Images (IXI[2]), and the Alzheimer's Disease Neuroimaging Initiative (ADNI) (Jack et al., 2017).

The NAC is a research center affiliated with the Surgical Planning Laboratory and Harvard University that contains 149 3-D triangular meshes labeled by an expert. All meshes were spatially aligned to a T1-weighted (T1-w) reference image from a healthy 42-year-old male and had 1 mm isotropic resolution and a matrix size of $256 \times 256 \times 256$ voxels.

The IXI database is part of the Brain Development project at Imperial College London that contains about 600 MR images of CN individuals, weighted T1-w, T2-w, PD, Magnetic Resonance Angiography (MRA) Diffusion Tensor Imaging (DTI) images in 15 directions.

The ADNI initiative was launched in 2003 by a group of research institutions, private pharmaceutical companies, and non-governmental organizations, with the help of researchers worldwide. The initiative has been looking for ways to determine AD progression by developing clinical biomarkers for detecting the disease in its early stages. This database contains MR images, functional RM, Positron Emission Tomography, DTI, genetic and biochemical information.

For this study, we used the T1-w NAC reference image and two hippocampal 3-D triangular meshes for the preprocessing stage (Section 2.2) to mitigate problems inherent to image acquisitions, standardize all study images for processing, and define the region of interest (ROIs) in our analysis. From the IXI, we used 563 MR T1-w images, and from the ADNI, 842 MR T1-w images, including longitudinal data. The used ADNI MR images were acquired by the sequence Magnetization Prepared Rapid Gradient Echo (MPRAGE[3]) and 1.5T and 3T

[2] https://brain-development.org/ixi-dataset/.

[3] *"These MPRAGE files are considered the best in the quality ratings and have undergone gradwarping, intensity correction, and have been scaled for gradient drift using the phantom data"*. - http://adni.loni.usc.edu/methods/mri-tool/mri-analysis/

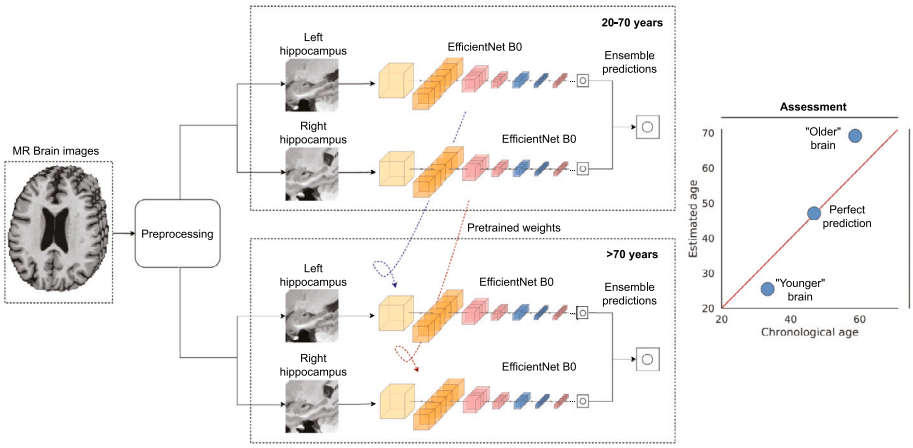

Fig. 1. Overview of the brain age prediction framework.

Table 1
Number of CN subjects used in each step of the model design.

	Age range	Train		Validation		Test	
		ADNI/IXI	Total	ADNI/IXI	Total	ADNI/IXI	Total
Unique subjects	20–70	127/449	576	15/29	44	15/28	43
	>70	167/57	224	151/0	151	151/0	151
Longitudinal data	20–70	**238**/449	687	15/29	44	15/28	43
	>70	**272**/57	329	151/0	151	151/0	151

Table 2
Cohort demographics of the dataset used for the evaluation of our regression model.

Diagnosis	# Subjects	Age (70-85)	Gender (M/F)	MMSE
CN	302	75.79 ± 4.14	151/151	29.56 ± 0.50 (29-30)
MCI	251	76.80 ± 4.30	80/171	27.04 ± 0.82 (26-28)
AD	209	77.12 ± 4.43	94/115	22.15 ± 2.76 (4-25)

MMSE stands for Mini-Mental State Examination.

scanners from three leading manufacturers (Philips, General Electric, and Siemens) from over 50 locations in the U.S. and Canada.[4] Detailed information about the images used in each experiment is shown in Table 1.

As it can be noticed, the IXI images were mainly used for the first experiment since it contains young subjects, and the ADNI images for the second because it contains images from elderly subjects. Since our goal is to explore the associations of brain age delta of age-matched CN subjects with AD and MCI subjects, the second experiment contains only CN subjects from the ADNI dataset in the validation and test splits. Table 2 presents the demographic information of the subjects used in this study, including the already referenced CN subjects plus age-matched MCI and mild-AD subjects.

2.2. Preprocessing

In this study, we first preprocessed all study images (i.e., ADNI and IXI datasets) with the Non-Local Means (NLM) technique (Buades, Coll, & Morel, 2005) for noise reduction, followed by the N4-ITK technique (Tustison et al., 2010) for bias field correction. Then, using the T1-w template image from the NAC dataset as a reference, we performed image intensity standardization using the histogram matching algorithm proposed in Nyúl, Udupa, and Zhang (2000), followed by affine spatial alignment using the Nifty-Reg image registration tool (Ourselin, Stefanescu, & Pennec, 2002). Furthermore, to define the hippocampal regions, we used the left and right hippocampi meshes from the NAC dataset as a reference and obtained hippocampal regions of the size of 64 × 64 × 64. For this, we performed a deformable registration using the study images as a reference and the NAC as fixed and obtained the transformation matrix. Then, we applied it to the meshes and discarded the deformed images generated. Lastly, we positioned the deformed meshes on the preprocessed images and defined our regions of interest by cropping a 64 × 64 × 64 patch around the hippocampal meshes gravity center.

2.3. Convolutional neural network

The EfficientNets are a family of CNNs (2D) models created by the Google research brain team[5] that have achieved better accuracy and efficiency based on ImageNet previously proposed CNNs (Tan & Le, 2019). This family of networks is based on a base architecture, B0, created from a search for architectures that simultaneously optimize accuracy and floating-point operations per second (FLOPS). Its architecture is mainly composed of blocks from the mobile inverted bottleneck convolution (MBConv) (Sandler, Howard, Zhu, Zhmoginov, & Chen, 2018), with the addition of compression and excitation optimization (Hu, Shen, & Sun, 2018). Using the base architecture, scaling applications change the depth (#layers), width (#channels), and resolution of the network simultaneously.

Because the architecture of EfficientNets has shown better accuracy and efficiency than other networks proposed in the literature and also because its base architecture, B0, has a smaller number of parameters

[4] http://adni.loni.usc.edu/about/centers-cores/study-sites/.

[5] https://research.google/teams/brain/.

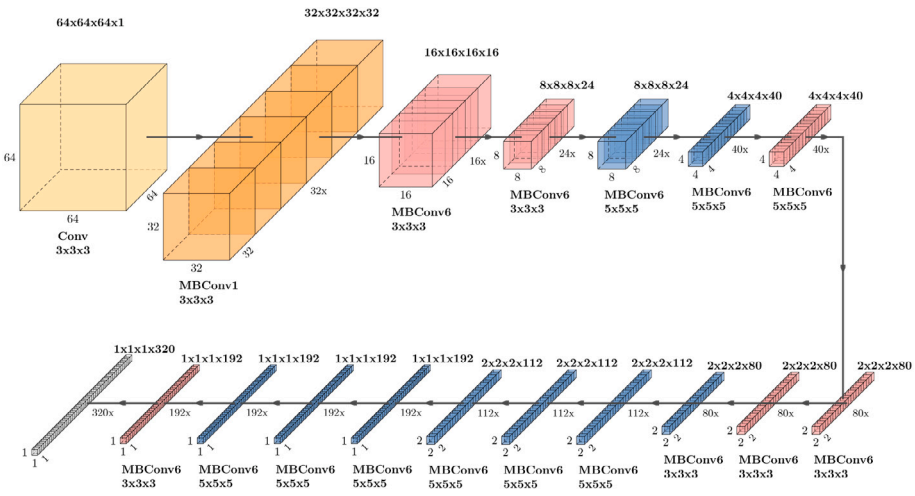

Fig. 2. Efficient Net B0 architecture used in this study.

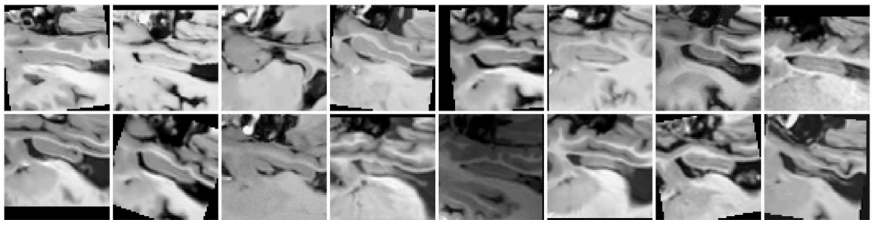

Fig. 3. Example of sixteen random batch images.

(5.3M), we decided to use this architecture with the hippocampal ROIs as input to produce a single scalar regression output representing the predicted age of the subject. To identify the best set of hyperparameters, we performed multiple experiments as described in the following subsections.

Architecture

The architecture used in this research consists of 16 blocks. We adapted the architecture for 3D by changing all 2D operations for 3D and the output for regression by changing the final output layer to a linear activation function to predict age as the output scalar. The input size corresponds to the hippocampal ROI dimensions, i.e., 64×64×64×1, generating the architecture presented in Fig. 2, with a total of 4 million parameters.

We performed all experiments separately for the left and right hippocampal regions and combined each scalar prediction to create the final network. Therefore, the final age prediction consists of the average value of the two network predictions.

Data augmentation

The ADNI dataset provides longitudinal follow-up sessions from the same subjects, which we considered a natural augmentation. At the training phase, we used the ADNI longitudinal data combined with up to twenty augmentation transformations and combinations, consisting of two affine operations for each axis (x, y, and z) with random choices of the parameters between a defined interval, i.e., a translation between ±10 pixels and rotation between ±20 degrees, resulting in six transformations. We also applied additive Gaussian noise with the sigma value randomly choose between 5 and 30 and random bias field multiplicative noise with a maximum magnitude of polynomial coefficients between −0.3 and 0.3. Further, we combined the transformations to generate a total of twenty different transformations, e.g., a random rotation along the x-axis followed by a random translation along the z-axis. Fig. 3 shows an example of sixteen random batch images.

We also designed the augmentations to obtain a uniform age span; therefore, we generated more augmentations for subjects from age ranges with fewer subjects and vice-versa. For this, we stratify the populations into age bins with a three years interval. Figs. 4(a) and 4(b) show the age distribution comparison of the population for both experiments, before and after the data augmentation. For the first training experiment, ages between 20 and 70, we obtained 5098 training images, and for the second, ages greater than 70 years, we obtained 1759 training images.

We performed all transformations with the Torchio (Pérez-García, Sparks, & Ourselin, 2020)[6] python library and applied it on the fly (online data augmentation) to save memory and include more data variations since the transformation parameters will generate different augmentations in each epoch.

2.4. Hyperparameter settings

We trained two different networks for the left and right hippocampus and combined the outputs by calculating the mean result to create the final prediction and form an ensemble network. Our results showed that the model ensemble had increased the stability of the network predictions with variance reduction due to the use of more complementary data, i.e., two different ROIs from the same brain.

We calculated the mean absolute error (MAE), the root mean square error (RMSE), and the $\Delta_{\text{brain\_age}}$ for each hippocampal network and ensemble. Further, for the first training, we calculated the Pearson correlation coefficient (r).

To find the best set of hyperparameters, we evaluated the validation datasets with a stratified hold-out procedure using the MAE as a loss function. To stratify the subject ages, we performed a discretization considering a three years interval. Fig. 5 shows the age distribution

[6] https://pypi.org/project/torchio/.

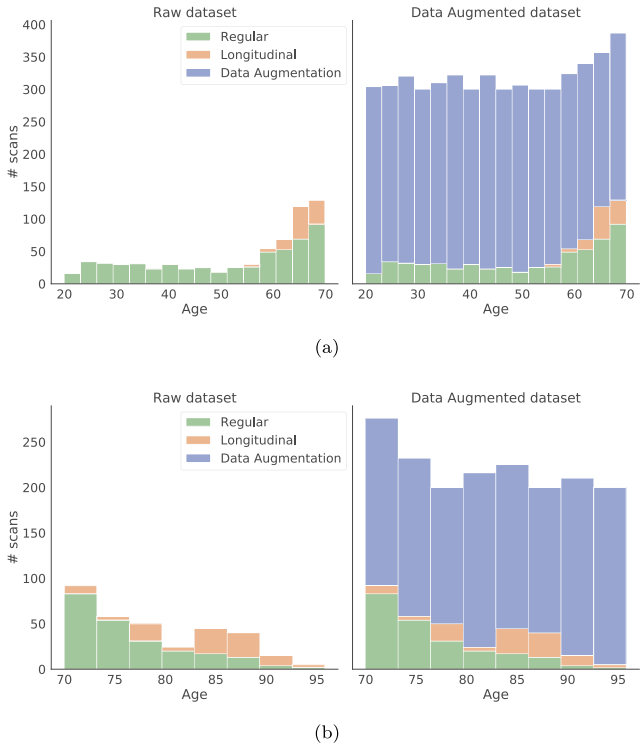

Fig. 4. The age distribution of the study population of the raw dataset and the data augmented dataset. (a) For the experiment aged between 20 and 70, and (b) For the experiment larger than 70 years.

for the validation and test groups for both experiments. From the histograms, we can observe that the test and validation data have similar distributions.

We trained all models using a cyclical learning rate (CLR) strategy (Smith, 2017) with exponential decay. In summary, the CLR varies the learning rate (LR) between a lower and an upper bound (base_lr and max_lr) using a decaying policy. In practice, these periodic changes in LR values help to avoid saddle points or local minima, consequently accelerating the training process (Smith, 2017). To derive the optimal bounds for CLR initialization, we run the model for 50 epochs with a linear increase of the LR between $1e^{-10}$ and $1e^{+1}$ and access the LR over the epochs to find its "optimal range", as suggested in Smith (2017).

Further, we tested different optimization and regularization techniques using each of the respective found CLR bounds. For the optimization, we tested the SGD, SGD with momentum (Qian, 1999), Adam (Kingma & Ba, 2014), and RMSprop (Tieleman & Hinton, 2012) algorithms. To avoid overfitting, we tested the following regularization techniques: weight decay (L2), dropout, and data augmentation. The first set of experiments were assessed for 150 epochs and the second for 100 epochs.

Furthermore, we implemented all experiments using PyTorch (v. 1.7.1) and trained the models using parallel batch sized on two NVIDIA 1080-TI GPUs. The hyper-parameters used to train the neural networks are shown in Table 3. Source code will be available at GitHub[7] after publication.

2.5. Statistical analysis

As mentioned, the main research goal of this work is to explore the associations of the $\Delta_{\text{brain\_age}}$ of age-matched CN subjects with AD and MCI subjects. Therefore, we tested if the $\Delta_{\text{brain\_age}}$ scores obtained in our second experiment presented statically significant differences than those of the other two groups: MCI and AD. To this end, we performed the analysis of variance (ANOVA) (Lars & Svante, 1989) statistical test and assessed the pairwise relationship between the groups with the Tukey's honestly significant difference (HSD) (Abdi & Williams, 2010) post-hoc test. The ANOVA null hypothesis states that all means are equal, against the alternative hypothesis stating that at least one mean differs, considering a significance interval $\alpha_t = 0.05$. However, since we are testing three groups with ANOVA, we cannot identify which ones differ. Therefore, we used the HSD test to examine the significant pairwise mean differences between the groups, with $\alpha_t = 0.05$.

Further, we assessed the relationship between the $\Delta_{\text{brain\_age}}$ scores and the MMSE cognitive measure scores with a Pearson's pairwise correlation to measure the strength and direction of the association between the two continuous variables.

3. Results

Our experiments used the training images described in Section 2.1 with the augmentations described in Section 2.3. We present both training experiment results for the left and right hippocampal regions and the ensemble prediction results in the following sections.

3.1. Experiment 1: patients aged between 20 and 70 years

We started the model analysis by investigating the influence of the regularization techniques, e.g., data augmentation, dropout rates, and weight decay. For this, we plotted the model's loss curves in each scenario and assessed the patterns between training and validation. Fig. 6 shows the loss curves contrasting the training stage with and without data augmentation for 150 epochs. We can notice that the model being trained without data augmentation struggled with local minima or saddle points and did not improve over the epochs. In contrast, the model with augmentation has significantly improved the training results.

Further, we evaluate the dropout rates and weight decay influence using data augmentation; however, the models struggle with local minima with weight decay and high percents of dropout. We also tested different optimizers, and the RMSprop, the same optimizer used in the EfficientNet (Tan & Le, 2019) paper, achieved the lowest loss in our experiments. Therefore, we present and discuss the results using the hyperparameters mentioned in Table 3. We used data augmentation, dropout of 0.2 and 0.3, a batch size of 128, and the RMSprop optimizer. We executed this first experiment for 150 epochs and selected the model that achieved the lowest validation loss among the epochs.

Table 4 shows the results of the first training. As we can notice, the left and right hippocampus results are very similar, while the ensemble has improved our results. We achieved a test MAE of 3.31, RMSE of 4.65, r of 0.95, and a $\Delta_{\text{brain\_age}}$ of −0.68 for the ensemble model.

Fig. 7 shows the correlation plots of the estimated ages with the chronological ages for each hippocampus region and the ensemble prediction. From the plots, we can observe that both networks achieve different individual age predictions, despite the correlation (r) results for the left and right hippocampus being the same, 0.93. The differences can be noticed by assessing the marginal histogram plots. The estimated age bins and the chronological bins are similarly distributed for the left hippocampus but shown a larger difference between the bins corresponding to the second largest age bin, i.e., approximately between 55 and 63 years. The same pattern can be found for the right hippocampus, but with a larger difference between the bins corresponding to the smaller ages, i.e., approximately between 20 and 27 years. These differences between the model's predictions can be complementary and help to improve the results. When combining the model's predictions (ensemble model), we obtained a higher correlation value, $r = 0.95$, and a much similar histogram distribution between

[7] https://github.com/kapoloni/age_prediction.

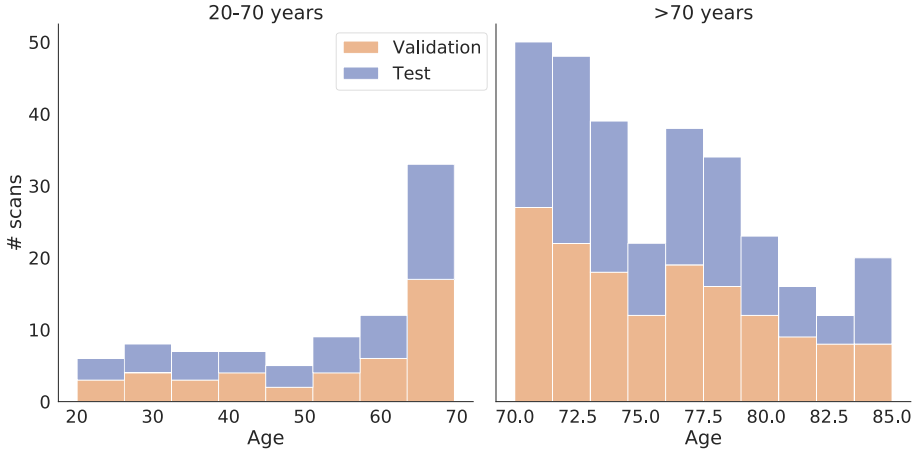

Fig. 5. The age distribution of the study population for the validation and test images.

Table 3
Hyper-parameters used in the experiments.

Age	Hyperparameters					Max of
	Side	CLR (exp_range)	Dropout rate	Batch size	Optimizer	
20-70	Left	$[10e^{-5.2}, 10e^{-3.4}]$	0.3	128	RMSprop	150 epochs
	Right	$[10e^{-5.2}, 10e^{-3.6}]$	0.2	128	RMSprop	150 epochs
>70	Left	$[10e^{-4.7}, 10e^{-3.3}]$	0.2	128	RMSprop	100 epochs
	Right	$[10e^{-4.7}, 10e^{-3.3}]$	0.2	128	RMSprop	100 epochs

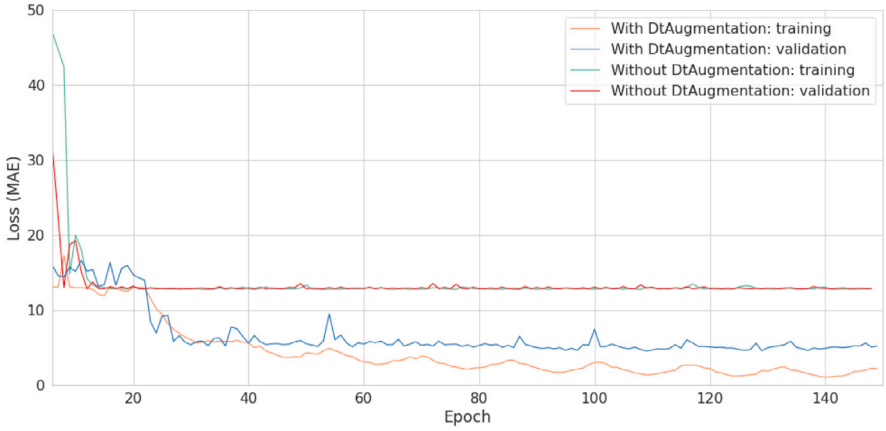

Fig. 6. Model loss curves with and without data augmentation.

Table 4
Brain age prediction results for the first experiment.

Side	Validation				Test			
	MAE	RMSE	Δ_{brain_age}	r	MAE	RMSE	Δ_{brain_age}	r
Left	4.8	6.29	0.49	0.92	4.24	6.03	0.17	0.93
Right	4.24	5.81	−0.04	0.92	4.71	6.11	−1.54	0.93
Ensemble	3.87	5.01	0.23	0.95	**3.31**	**4.65**	**−0.68**	**0.95**

the estimated age and chronological ages. Moreover, we performed a Pearson's pairwise correlation and obtained a p-value < 0.001 for all three models, implying that the estimated ages and the chronological ages correlation presented a statically significant linear relationship.

Furthermore, we compared the input images of age-matched subjects with low and high $\Delta_{\text{brain\_age}}$ prediction as demonstrated in Fig. 8 - the slice images are axial, coronal, and sagittal. Both subject results can be identified when examining the correlation plots in Fig. 7. Subjects 1 and 2 were approximately 43 years old, but subject 1 was predicted to be 32.56 years old, while subject 2 was predicted to be 44.22. Thus, subject 1 has a $\Delta_{\text{brain\_age}}$ of −10.41, meaning the model has underestimated the age by at least ten years, and subject 2 has a $\Delta_{\text{brain\_age}}$ of 1.04, meaning the model has overestimated the age by approximately one year. In addition, from the slice images, we observe the subject 1 has a younger appearance, i.e., presents less atrophy than subject 2, which can be evidenced by the low presence of the CSF tissue (darker color) on subject 1 image slices. This creates difficulties for the model prediction and, therefore, lowers its accuracy.

Comparison with existent methods

Table 5 shows the study results presenting the current state-of-the-art in structural MR image brain age prediction.

Although they are not directly comparable, we have achieved competitive results, even using only the hippocampal regions.

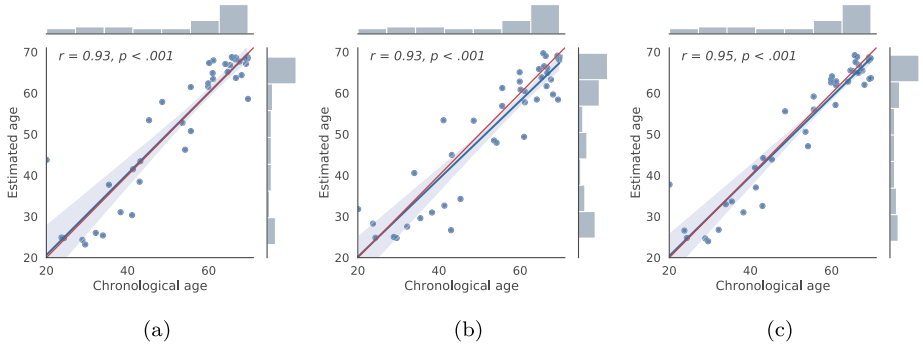

Fig. 7. Correlation plots of the estimated ages with the chronological ages from the CNN model for the left hippocampus (a), the right hippocampus (b), and the ensemble model (c).

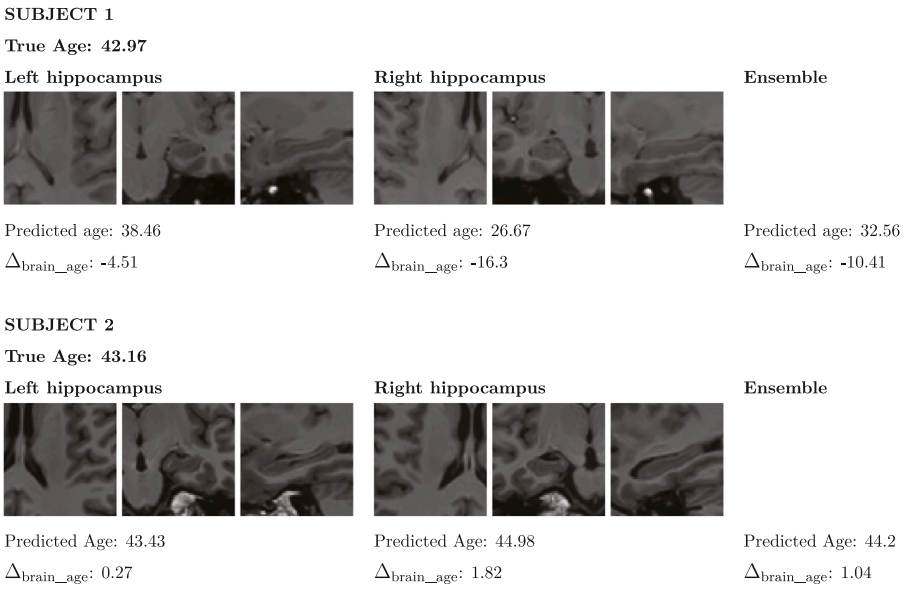

Fig. 8. MRI Images from two subjects both with true age ≈ 43.

3.2. Experiment 2: patients aged greater than 70 years

The second set of experiments used the pre-trained best models of the first training experiments. Since the weights are not random, we performed the experiment for only 100 epochs, and we selected the model that achieved the lowest validation loss among the epochs. We also assessed the influence of regularization and optimizer techniques. The best set of hyperparameters are presented in Table 3. We used data augmentation, dropout of 0.2, a batch size of 128, and the RMSprop optimizer.

Table 6 shows the validation and test results of the second training. The left and right hippocampus results are very similar, and the ensemble prediction has improved the results. We achieved a test MAE of 3.66, RMSE of 4.58, and a $\Delta_{\text{brain\_age}}$ of 1.24. As expected, the results are pretty similar to the first experiment; since we continued using images from CN subjects, we have only changed the age range.

Table 7 presented the results of the second training evaluated with images belonging to MCI and AD subjects. Comparing with the previous test results (CN subject images), we can notice an overestimation of the $\Delta_{\text{brain\_age}}$ for the MCI and AD groups, following a pattern of CN<MCI<AD scores. Furthermore, we can also notice the same pattern for the MAE and RMSE metrics.

Statistical analysis

To verify if the presented pattern is statistically significant, we performed the ANOVA and Tukey HSD tests. Table 8 presents the results of these tests, which include the mean and standard deviation of the $\Delta_{\text{brain\_age}}$ values, the ANOVA p-value, and the conclusions after running the Tukey's HSD post-hoc test.

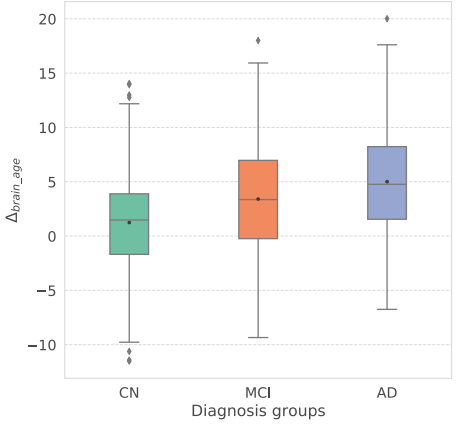

Fig. 9. Boxplots of $\Delta_{\text{brain\_age}}$ values by diagnosis groups.

The values have a noticeable trend of decaying as the severity of diagnosis: the values are lower in CN than MCI and MCI than AD. This trend is consistent since it shows an increase in biological brain aging

Table 5
Comparison methods.

	#Images [age range]	CNN	Batch size	Loss function	Optimization	Regularization	Learning rate policy	Results		
Huang et al. (2017)	1099 [20-80]	2D brain	16	MAE	SGD + *Momentum*	L2 translation DTA: rotation crop	Step decay (one epoch)	MAE: 4		
Cole and Franke (2017a)	2001 [18-90]	3D brain	28	MAE	SGD + *Momentum*	L2 translation DTA: rotation	Step decay (one epoch)	MAE: 4.16 RMSE: 5.31 r: 0.96		
Ueda et al. (2019)	1101 [20-80]	3D brain	16	MAE	SGD + *Momentum*	L2 translation DTA: rotation crop	Step decay (one epoch)	MAE: 3.67 RMSE: 4.71 r: 0.97		
Feng et al. (2020)	2694 [18-97]	3D brain	5	MAE	Adam	L2 DTA: longitudinal	Range	MAE: 4.21 r: 0.96		
Pardakhti and Sajedi (2020)	562 [20-86]	3D brain	8	MSE	SGD + *Momentum*	L2	Constant	MAE: 5.149 RMSE: 13.5		
Peng et al. (2021)	14503 [44-80]	3D brain	8	Kullback-Liebler	SGD + *Momentum*	L2 *Dropout* DTA: translation rotation (saggital axis)	Step decay (30 epochs)	MAE: 2.14		
Dinsdale et al. (2021)	12802 [44-80]	3D brain	16	MSE	RMSprop	*Early stopping*	Step decay (30 epochs)		Fem	Masc
								MAE	2.86	3.09
								RMSE	13.12	15.13
								r	0.87	0.86
Proposed method	774 [20-70]	3D hippocampus	128	MAE	RMSprop	*Dropout* translation rotation DTA: noise bias field compositions	CLR	MAE: 3.64 RMSE: 5.32 r: 0.94		

DTA stands for Data augmentation.

Table 6
Brain age prediction results for the second experiment.

Side	Validation			Test		
	MAE	RMSE	Δ_{brain_age}	MAE	RMSE	Δ_{brain_age}
Left	3.74	4.7	0.36	4.05	5.12	0.82
Right	3.85	4.92	1.29	3.96	4.98	1.65
Ensemble	3.52	4.39	0.82	**3.66**	**4.58**	**1.24**

Table 7
Brain age prediction results for the MCI and AD groups.

Side	MCI			AD		
	MAE	RMSE	Δ_{brain_age}	MAE	RMSE	Δ_{brain_age}
Left	4.79	6.04	2.21	5.48	6.71	3.88
Right	6.38	7.56	4.6	6.82	8.22	6.13
Ensemble	**5.19**	**4.45**	**3.4**	**5.8**	**7.05**	**5.01**

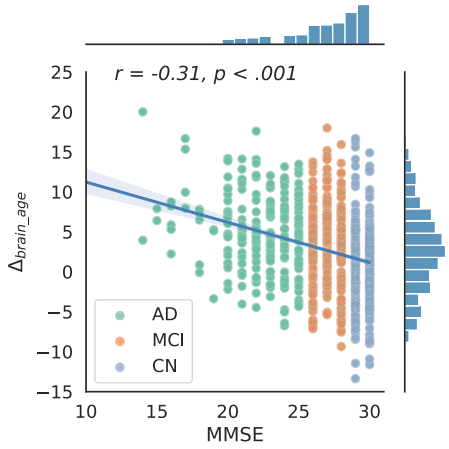

Fig. 10. Correlation plot of the Δ_{brain_age} and MMSE cognitive score.

as Alzheimer's severity increases; during Alzheimer's development, the brain suffers from more intense and accentuated structural changes. Therefore, we expected an overestimation from AD and MCI subject's brains. Fig. 9 shows the box plot of the Δ_{brain_age} values for each diagnosis group.

At last, we performed a Pearson's pairwise correlation to measure the strength and direction of association between the MMSE score (clinical score) and the obtained Δ_{brain_age} values, as illustrated in Fig. 10. We found a statistically significant correlation between the variables (p-value < 0.001). As expected, we obtained a negative correlation coefficient, $r = -0.31$, since both scores have an inverse relationship, i.e., smaller MMSE scores and larger values of Δ_{brain_age} indicate more disease severity. Although the correlation coefficient value has not been high in this comparison, we can denote a close relationship between pathological brain aging and prospective worsening of cognitive functioning.

In order to understand the diagnostic group's differences and correlated them with the age differences, we created averaged images from the three diagnosis groups and two age ranges, 70 to 77 and 78 to 85. Fig. 11 shows the average images from the three groups and the difference between a group pairwise comparison. Assessing only the average images of each group, we notice subtle differences, with an increase of the CSF presence as AD severity increases. Evaluating the pairwise comparisons, we noticed a more considerable difference.

Table 8
Mean and standard deviation of Δ_{brain_age} values by diagnosis group and results of ANOVA ($\alpha = 0.05$) and post-hoc comparisons survived at Tukey HSD ($\alpha_t < 0.05$).

	CN	MCI	AD	ANOVA p-value	post-hoc
Δ_{brain_age}	1.24 ± 4.42	3.4 ± 5.32	5.01 ± 4.97	< 0.001	CN<MCI, CN<AD, MCI<AD

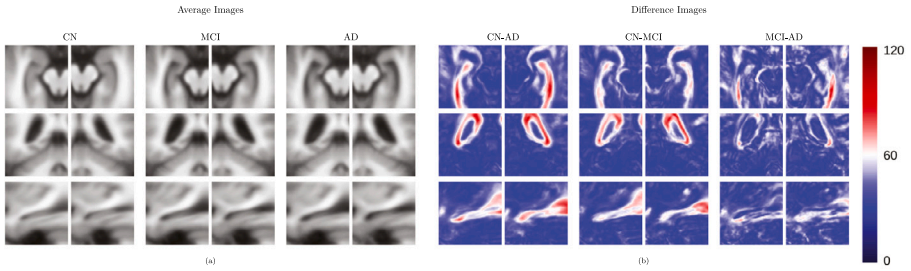

Fig. 11. Difference images between the CN and AD subjects, CN and MCI subjects, and MCI and AD subjects for the **left** and **right** hippocampus for the axial, coronal, and sagittal views, respectively.

First, the CN-AD images showed more significant differences in the hippocampal areas, evidenced by the red color. Then, the CN-MCI images showed similar differences, but to a minor extent. Last, the MCI-AD evidenced a lot more subtle differences.

Fig. 12 shows the average images from the three diagnosis groups separated by age range and the differences between the ages and diagnosis. From these images, we can visually notice the existent atrophy that develops with aging (CN images) and increases with neurodegeneration (MCI and AD images). In addition, this atrophy is enlarged within age groups with more disease severity.

When comparing these images with the box plots presented in Fig. 9, we noticed that our score works better to differentiate CN and AD, then CN and MCI, and last MCI and AD. These visual trends presented are expected and knew in the literature (Franke et al., 2010; Gaser et al., 2013). More importantly, they are consistent with our results since we have found a minor Δ_{brain_age} for the CN than MCI than AD.

4. Discussions

Using an efficient 3D convolutional neural network architecture, we estimated the hippocampal age from raw T1-weighted MRI brain scans of healthy adults accurately. We performed several data augmentation and compositions to obtain a large and evenly distribution of the age bins so that to improve our results and obtain unbiased results. We divided our work into two steps to provided two separable and complementary analyses to allow the network to perform well on both experiments. Furthermore, we trained the networks relatively fast with a quick inference time. For the first set of experiments, the left and right hippocampus training took 3 h and 20 min each, and the inference time for the model ensemble took only 0.12 s on GPU and 0.165s on CPU. For the second experiment, the training took 56 min for each hippocampus, and the model ensemble inference was the same as in the first experiment. The full preprocessing time took approximately 6 min.

Our 3D ensemble model achieved an MAE of 3.31 and an RMSE of 4.65 using the brains aged between 20 and 70 years. Our Pearson's pairwise correlation test showed a significant correlation between the chronological and estimated ages and an r coefficient of 0.95. These results are competitive with the results achieved for other methods reported in the literature (Table 5 is provided for reference), even when using only the hippocampal region. To verify if our Δ_{brain_age} have biological backgrounds or are being caused for model prediction failure, we compared Subjects 1 and 2 in Fig. 8. Both subjects have the same chronological age but very different levels of atrophy and predicted ages. Visually we can notice differences in their anatomical structures, explained by the presence or absence of the CSF. These differences can be explained due to the extensive range of "healthy" variations existent in the human brain as aging, with anatomical changes within a subject and across a population (Dinsdale et al., 2021).

For the second experiment, we used the pre-trained weights from the first models. We performed an age-matched and unbiased comparison between the CN age predictions with AD and MCI. Our 3D ensemble model achieved an MAE of 3.66, an RMSE of 4.58, and a Δ_{brain_age} of 1.24 (CN subjects). Although from a different age range, the results are still comparable or even better than existing ones (Table 5). For the MCI subjects' images, we achieved an MAE of 5.19, an RMSE of 4.45, and a Δ_{brain_age} of 3.4. For the AD subjects, we achieved an MAE of 5.8, an RMSE of 7.05, and a Δ_{brain_age} of 5.01. In Table 8, we reported the results of the ANOVA followed by the Tukey HSD statistical test. With the presented results, we have shown a trend for the Δ_{brain_age} and the other metrics that is consistent with cognitive decline advancement: the values are lower in CN than MCI and MCI than AD. Then, we have also evaluated the strength and direction of the MMSE score and our Δ_{brain_age} and founded a significant and negative correlation. As already reported in other studies (Franke et al., 2010; Gaser et al., 2013), higher delta scores are closely related to measures of clinical disease severity in AD patients. These results strongly support the relationship between profoundly accelerated brain aging and disease severity, most pronounced in subjects being already diagnosed with AD, and prospective worsening of cognitive functioning.

Furthermore, we compared the population average images and subtracted them pairwise, e.g., CN and AD, CN and MCI, and MCI and AD, to illustrate the class differences and provide an interpretative result for our Δ_{brain_age} trend (Fig. 11). We have also compared the average images from subjects between 70–77 years and 78–85 years to roughly examine which brain features are more pronounced with aging and cognitive diagnosis (Fig. 12). We noticed that differences between the age groups on the averaged images are more subtle in CN than MCI and AD, which is also consistent with our results.

5. Conclusions

Several deep learning models are proposed in the literature to predict brain age, with an accurate prediction on the healthy population. There are even studies showing that they could be implanted in clinical practice and used as a clinical biomarker. Among the implementation limitations, one is the end-to-end processing time of a new scan. Cole and Franke (2017a) have created a network architecture that uses an

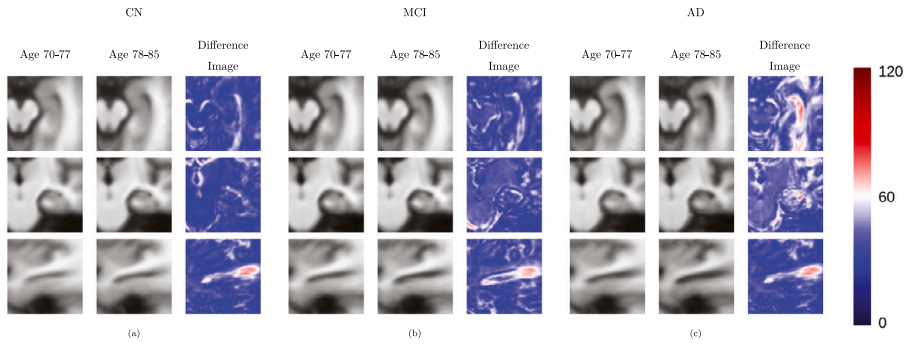

Fig. 12. Difference images between the mean images for diagnosis groups divided by age range between 70–77 and 78–85: the CN group, the MCI, and the AD for the **left** hippocampus.

input image with minimum processing (almost disregarded all preprocessing steps) and obtained an MAE of 4.65 years. The study reported a fast inference time, between 290–940 ms, but failed to mention the image processing time needed before entering the network. We have proposed an ensemble hippocampal age estimation approach that uses an efficient network architecture based on mobile networks (Sandler et al., 2018). We also proposed a two-step training and were able to perform two analyses. Our inference time for both steps was 0.12s, and our end-to-end processing time was less than seven minutes. In the first training, we achieved a very competitive MAE and RMSE and provided a qualitative analysis of two age-matched subjects with different predictions. Our second step provided a statistically significant analysis of the $\Delta_{\text{brain\_age}}$ using three distinct groups (CN, MCI, and AD) with distinct aging effects and stages of neurological diseases. We corroborate our results with clinical measurements, e.g., MMSE score, that could indicate the possibility of using our $\Delta_{\text{brain\_age}}$ as a biomarker to identify and support AD and MCI. As future work, we suggested a per-structure analysis once the hippocampus has already obtained such promising results. Furthermore, due to the longitudinal aspect of the ADNI dataset, we intend to determine if the $\Delta_{\text{brain\_age}}$ could capture such differences presented in the time scans (usually repeated every six months).

CRediT authorship contribution statement

Katia Maria Poloni: Conceptualization, Methodology, Software, Writing – original draft. **Ricardo José Ferrari:** Conceptualization, Supervision, Writing – review & editing, Resources.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI), United States (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

Funding statement

The authors would like to thank the São Paulo Research Foundation (FAPESP), Brazil (grant numbers 2018/08826-9 and 2018/06049-5) and the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001 for the financial support of this research.

References

Abdi, H., & Williams, L. J. (2010). Tukey's honestly significant difference (hsd) test. *Encyclopedia of Research Design*, *3*, 1–5.

Ahmed, S., Kim, B. C., Lee, K. H., & Jung, H. Y. (2020). Ensemble of ROI-based convolutional neural network classifiers for staging the Alzheimer disease spectrum from magnetic resonance imaging. *PLoS One*, *15*, Article e024712.

Alzheimer's Association (2020). Alzheimer's disease facts and figures. *Alzheimer's & Dementia: The Journal of the Alzheimer's Association*, *16*(391).

Ardekani, B. A., Hadid, S. A., Blessing, E., & Bachman, A. H. (2019). Sexual dimorphism and hemispheric asymmetry of hippocampal volumetric integrity in normal aging and alzheimer disease. *American Journal of Neuroradiology*, *40*, 276–282.

Baumgartner, C. F., Tezcan, K. C., Chaitanya, K., Hötker, A. M., Muehlematter, U. J., Schawkat, K., et al. (2019). Phiseg: Capturing uncertainty in medical image segmentation. In *International conference on medical image computing and computer-assisted intervention* (pp. 119–127). Springer.

Beach, T. G., Monsell, S. E., Phillips, L. E., & Kukull, W. (2012). Accuracy of the clinical diagnosis of alzheimer disease at national institute on aging alzheimer disease centers, 2005–2010. *Journal of Neuropathology and Experimental Neurology*, *71*, 266–273.

Buades, A., Coll, B., & Morel, J.-M. (2005). A review of image denoising algorithms, with a new one. *Multiscale Modeling & Simulation*, *4*, 490–530.

Cherubini, A., Caligiuri, M. E., Péran, P., Sabatini, U., Cosentino, C., & Amato, F. (2016). Importance of multimodal MRI in characterizing brain tissue and its potential application for individual age prediction. *Journal of Biomedical and Health Informatics*, *20*, 1232–1239.

Cole, J. H., & Franke, K. (2017a). Predicting age using neuroimaging: innovative brain ageing biomarkers. *Trends in Neurosciences*, *40*, 681–690.

Cole, J. H., Poudel, R. P. K., Tsagkrasoulis, D., Caan, M. W. A., Steves, C., Spector, T. D., et al. (2017). Predicting brain age with deep learning from raw imaging data results in a reliable and heritable biomarker. *NeuroImage*, *163*, 115–124.

Dinsdale, N. K., Bluemke, E., Smith, S. M., Arya, Z., Vidaurre, D., Jenkinson, M., et al. (2021). Learning patterns of the ageing brain in MRI using deep convolutional networks. *NeuroImage*, *224*, Article 117401.

Dong, X., Yu, Z., Cao, W., Shi, Y., & Qianli, M. A. (2020). A survey on ensemble learning. *Frontiers of Computer Science*, *14*, 241–258.

Feng, X., Lipton, Z. C., Yang, J., Small, S. A., & Provenzano, F. A. (2020). Estimating brain age based on a healthy population with deep learning and structural MRI. *Neurobiology of Aging*, *91*, 15–25.

Franke, K., Ziegler, G., Klöppel, S., Gaser, C., Initiative, A. D. N., et al. (2010). Estimating the age of healthy subjects from t1-weighted MRI scans using kernel methods: exploring the influence of various parameters. *Neuroimage*, *50*, 883–892.

Fujimoto, R., Ito, K., Wu, K., Sato, K., Taki, Y., Fukuda, H., et al. (2017). Brain age estimation from t1-weighted images using effective local features. In *International conference of the IEEE engineering in medicine and biology society* (pp. 3028–3031). IEEE.

Gaser, C., Franke, K., Klöppel, S., Koutsouleris, N., Sauer, H., Initiative, A. D. N., et al. (2013). Brainage in mild cognitive impaired patients: predicting the conversion to Alzheimer's disease. *PLoS One*, *8*, e67346.

Gaugler, J. E., Ascher-Svanum, H., Roth, D. L., Fafowora, T., Siderowf, A., & Beach, T. G. (2013). Characteristics of patients misdiagnosed with Alzheimer's disease and their medication use: an analysis of the nacc-uds database. *BMC Geriatrics*, *13*(137).

Giovannetti, A., Gianluca, S., Casti, P., Mencattini, A., Pusil, S., L'/opez, M. E., et al. (2021). Deep-MEG: spatiotemporal CNN features and multiband ensemble classification for predicting the early signs of Alzheimer's disease with magnetoencephalography. *Neural Computing and Applications*, *33*, 14651–14667.

Good, C. D., Johnsrude, I., Ashburner, J., Henson, R. N., Friston, K. J., & Frackowiak, R. S. (2001). Cerebral asymmetry and the effects of sex and handedness on brain structure: a voxel-based morphometric analysis of 465 normal adult human brains. *Neuroimage*, *14*, 685–700.

Guadalupe, T., Zwiers, M. P., Teumer, A., Wittfeld, K., Vasquez, A. A., Hoogman, M., et al. (2014). Measurement and genetics of human subcortical and hippocampal asymmetries in large datasets. *Human Brain Mapping*, *35*, 3277–3289.

Halle, M., Talos, I., Jakab, M., Makris, N., Meier, D., Wald, L., et al. (2017). Multi-modality MRI-based atlas of the brain. http://www.spl.harvard.edu/publications/item/view/2037.

He, Z., Shao, H., Zhong, X., & Zhao, X. (2020). Ensemble transfer CNNs driven by multi-channel signals for fault diagnosis of rotating machinery cross working conditions. *Knowledge-Based Systems*, *207*, Article 106396.

Hu, J., Shen, L., & Sun, G. (2018). Squeeze-and-excitation networks. In *Conference on computer vision and pattern recognition* (pp. 7132–7141). IEEE.

Huang, T., Chen, H., Fujimoto, R., Ito, K., Wu, K., Sato, K., et al. (2017). Age estimation from brain MRI images using deep learning. In *IEEE 14th international symposium on biomedical imaging* (pp. 894–852). Melbourne, Australia.

Hyman, B. T., Phelps, C. H., Beach, T. G., Bigio, E. H., Cairns, N. J., Carrillo, M. C., et al. (2012). National institute on aging–alzheimer's association guidelines for the neuropathologic assessment of alzheimer's disease. *Alzheimer's & Dementia*, *8*, 1–13.

Ioffe, S., & Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. arXiv preprint arXiv, abs/1502.03167.

Ito, K., Fujimoto, R., Huang, T.-W., Chen, H.-T., Wu, K., Sato, K., et al. (2018). Performance evaluation of age estimation from t1-weighted images using brain local features and CNN. In *2018 40th Annual international conference of the IEEE engineering in medicine and biology society* (pp. 694–697). IEEE.

Jack, C. R. J., Bernstein, M. A., Fox, N. C., Thompson, G., Alexander, P., Harvey, D., et al. (2017). The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods. *Journal of Magnetic Resonance Imaging*, *27*, 685–691.

Jiang, H., Lu, N., Chen, K., Yao, L., Li, K., Zhang, J., et al. (2020). Predicting brain age of healthy adults based on structural MRI parcellation using convolutional neural networks. *Frontiers in Neurology*, *10*(1346).

Johnson, K. A., Fox, N. C., Sperling, R. A., & Klunk, W. E. (2012). *Brain imaging in alzheimer disease. Vol. 2*. Cold Spring Harbor Perspectives in Medicine, a006213–1–23.

Kamnitsas, K., Ledig, C., Newcombe, V. F. J., Simpson, J. P., Kane, A. D., Menon, D. K., et al. (2017). Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. *Medical Image Analysis*, *36*, 61–78.

Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. arXiv preprint arXiv, abm/1412.6980.

Klatka, L. A., Schiffer, R. B., Powers, J. M., & Kazee, A. M. (1996). Incorrect diagnosis of Alzheimer's disease: a clinicopathologic study. *Archives of Neurology*, *53*, 35–42.

Korolev, S., Safiullin, A., Belyaev, M., & Dodonova, Y. (2017). Residual and plain convolutional neural networks for 3d brain MRI classification. In *International symposium on biomedical imaging* (pp. 835–838). IEEE.

Lars, S., & Svante, W. (1989). Analysis of variance (anova). *Chemometrics and Intelligent Laboratory Systems*, *6*, 259–272.

Li, Y., Liu, Y., Wang, P., Wang, J., Xu, S., & Qiu, M. (2017). Dependency criterion based brain pathological age estimation of alzheimer's disease patients with MR scans. *Biomedical Engineering*, *16*, 1–20.

Lim, A., Tsuang, D., Kukull, W., Nochlin, D., Leverenz, J., McCormick, W., et al. (1999). Clinico-neuropathological correlation of alzheimer's disease in a community-based case series. *Journal of the American Geriatrics Society*, *47*, 564–569.

Lin, W., Tong, T., Gao, Q., Guo, D., Du, X., Yang, Y., et al. (2018). Convolutional neural networks-based MRI image analysis for the Alzheimer's disease prediction from mild cognitive impairment. *Frontiers in Neuroscience*, *12*(777).

Liu, M., Zhang, J., Adeli, E., & Shen, D. (2018). Landmark-based deep multi-instance learning for brain disease diagnosis. *Medical Image Analysis*, *43*, 157–168.

Logan, R., Williams, B. G., Ferreira da Silva, M., Indani, A., Schcolnicov, N., Ganguly, A., et al. (2021). Deep convolutional neural networks with ensemble learning and generative adversarial networks for Alzheimer's disease image data classification. *Frontiers in Aging Neuroscience*, *17*, Article 720226.

Luo, Y., & Tang, X. (2017). Automated diagnosis of Alzheimer's disease with multi-atlas based whole brain segmentations. In *Medical imaging 2017: biomedical applications in molecular, structural, and functional imaging* (pp. 275–283). Bellingham, Washington: SPIE.

Madan, C. R., & Kensinger, E. A. (2018). Predicting age from cortical structure across the lifespan. *European Journal of Neuroscience*, *47*, 399–416.

McKhann, G. M., Knopman, D. S., Chertkow, H., Hyman, B. T., Jack, C. R., Kawas, C. H., et al. (2011). The diagnosis of dementia due to Alzheimer's disease: recommendations from the national institute on aging-Alzheimer's association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's & Dementia*, *7*, 263–269.

Miller, K. L., Alfaro-Almagro, F., Bangerter, N. K., Thomas, D. L., Yacoub, E., Xu, J., et al. (2016). Multimodal population brain imaging in the UK biobank prospective epidemiological study. *Nature Neuroscience*, *19*, 1523–1536.

Nyúl, L. G., Udupa, J. K., & Zhang, X. (2000). New variants of a method of MRI scale standardization. *IEEE Transactions on Medical Imaging*, *19*, 143–150.

Organization, W. H., et al. (2019). Risk reduction of cognitive decline and dementia: Who guidelines.

Ourselin, S., Stefanescu, R., & Pennec, X. (2002). Robust registration of multi-modal images: Towards real-time clinical applications. In *Medical image computing and computer-assisted intervention* (pp. 140–147). Heidelberg, Berlin: Springer.

Pardakhti, N., & Sajedi, H. (2017). Age prediction based on brain MRI images using feature learning. In *International symposium on intelligent systems and informatics* (pp. 000267–000270). IEEE.

Pardakhti, N., & Sajedi, H. (2019). Brain age estimation using brain MRI and 3D convolutional neural network. In *2019 9th International conference on computer and knowledge engineering* (pp. 386–390). Mashhad, Iran: IEEE.

Pardakhti, N., & Sajedi, H. (2020). Brain age estimation based on 3D MRI images using 3D convolutional neural network. In *Multimedia tools and applications* (pp. 1573–7721).

Peng, H., Gong, W., Beckmann, C. F., Vedaldi, A., & Smith, S. M. (2021). Accurate brain age prediction with lightweight deep neural networks. *Medical Image Analysis*, *68*, Article 101871.

Pérez-García, F., Sparks, R., & Ourselin, S. (2020). TorchIO: a Python library for efficient loading, preprocessing, augmentation and patch-based sampling of medical images in deep learning. arXiv:2003.04696. http://arxiv.org/abs/2003.04696.

Qian, N. (1999). On the momentum term in gradient descent learning algorithms. *Neural Networks*, *12*, 145–151.

Resnick, S. M., Pham, D. L., Kraut, M. A., Zonderman, A. B., & Davatzikos, C. (2003). Longitudinal magnetic resonance imaging studies of older adults: a shrinking brain. *Journal of Neuroscience*, *23*, 3295–3301.

Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L.-C. (2018). Mobilenetv2: Inverted residuals and linear bottlenecks. In *Conference on computer vision and pattern recognition* (pp. 4510–4520). IEEE.

Schnack, H. G., Van Haren, N. E. M., Nieuwenhuis, M., Hulshoff Pol, H. E., Cahn, W., & Kahn, R. S. (2016). Accelerated brain aging in schizophrenia: a longitudinal pattern recognition study. *American Journal of Psychiatry*, *173*, 607–616.

Shi, F., Liu, B., Zhou, Y., Yu, C., & Jiang, T. (2009). Hippocampal volume and asymmetry in mild cognitive impairment and alzheimer's disease: Meta-analyses of MRI studies. *Hippocampus*, *19*, 1055–1064.

Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. CoRR. abs/1409.1556.

Smith, L. N. (2017). Cyclical learning rates for training neural networks. In *Winter conference on applications of computer vision* (pp. 464–472). IEEE.

Su, L., Wang, L., Shen, H., & Hu, D. (2011). Age-related classification and prediction based on MRI: A sparse representation method. *Procedia Environmental Sciences*, *8*, 645–652.

Tan, M., & Le, Q. (2019). Efficientnet: Rethinking model scaling for convolutional neural networks. In *International conference on machine learning* (pp. 6105–6114). PMLR.

Tieleman, T., & Hinton, G. (2012). Lecture 6.5—RmsProp: Divide the gradient by a running average of its recent magnitude. In *COURSERA: neural networks for machine learning*.

Tustison, N. J., Avants, B. B., Cook, P. A., Zheng, Y., Egan, A., Yushkevich, P. A., et al. (2010). N4ITK: improved N3 bias correction. *IEEE Transactions on Medical Imaging*, *29*, 1310–1320.

Ueda, M., Ito, K., Wu, K., Sato, K., Taki, Y., Fukuda, H., et al. (2019). An age estimation method using 3D-CNN from brain MRI images. In *IEEE 16th international symposium on biomedical imaging* (pp. 380–383). Venice, Italy.

Wijnhoven, R. G., & de With, P. (2010). Fast training of object detection using stochastic gradient descent. In *Pattern recognition (ICPR), 2010 20th international conference on* (pp. 424–427).

Woolard, A. A., & Heckers, S. (2012). Anatomical and functional correlates of human hippocampal volume asymmetry. *Psychiatry Research: Neuroimaging*, *201*, 48–53.

Zhang, J., Liu, M., An, L., Gao, Y., & Shen, D. (2017). Alzheimer's disease diagnosis using landmark-based features from longitudinal structural MR images. *IEEE Journal of Biomedical and Health Informatics*, *21*, 1607–1616.