Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 159 (2019) 981–989

Procedia Computer Science

www.elsevier.com/locate/procedia

23rd International Conference on Knowledge-Based and Intelligent Information & Engineering Systems

Analysis of brain NMR images for age estimation with deep learning

Alberto Rossi[a,b,*], Gioele Vannuccini[b], Paolo Andreini[b], Simone Bonechi[b], Giorgia Giacomini[c], Franco Scarselli[b], Monica Bianchini[b]

[a]*Department of Information Engineering, University of Florence, Via di Santa Marta 3, Florence 50139, Italy*
[b]*Department of Information Engineering and Matemathics, University of Siena, Via Roma 56, Siena 53100, Italy*
[c]*Department of Biotechnology, Chemistry and Pharmacy, University of Siena, Via Aldo Moro 2, Siena 53100, Italy*

Abstract

During the last decade, deep learning and Convolutional Neural Networks (CNNs) have produced a devastating impact on computer vision, yielding exceptional results on a variety of problems, including analysis of medical images. Recently, these techniques have been extended to 3D images with the downside of a large increase in the computational load. In particular, state–of–the–art CNNs have been used for brain Nuclear Magnetic Resonance (NMR) imaging, with the aim of estimating the patients' age. In fact, a large discrepancy between the real and the estimated age is a clear alarm for the onset of neurodegenerative diseases, such as some types of early dementia and Alzheimer's disease. In this paper, we propose an effective alternative to 3D convolutions that guarantees a significant reduction of the computational requirements for this kind of analysis. The proposed architectures achieve comparable results with the competitor 3D methods, requiring only a fraction of the training time and GPU memory.

Keywords: Brain Age Estimation; Magnetic Resonance Imaging; Convolutional Neural Networks; Medical Image Analysis

1. Introduction

The brain is the command center of the human nervous system and controls most of the body's activities, also supervising the reception and processing of sensory information. Furthermore, cognitive abilities, language, emotions, creativity and memory are governed by the brain. Unfortunately, like all other parts of the human body, the brain also suffers from aging.

Aging is not uniform between different people and causes changes in brain size, vascularization and cognition. The problems of memory and cognitive impairment that may occur during aging would be more related to the loss of white matter, connective of the different regions of the brain, rather than the simple degeneration at the level of

* Corresponding author. Tel.: +39 3200138748
E-mail address: alberto.rossi@unifi.it

1877-0509
10.1016/j.procs.2019.09.265

the cerebral cortex (grey matter), as revealed by a study of the Massachusetts Institute of Technology [1]. The white matter consists of beams of neuronal axons that make connections between the neurons, allowing the brain regions to communicate with each other. The grey matter, on the other hand, is the place where neurons are found. In older subjects, a correlation between the decline in cognitive performance and the deterioration of the white matter of the frontal cerebellar regions, where the planning and execution functions are located, can be highlighted. Likewise, the deterioration of the white matter in the parietal and temporal lobes was associated with the weakening of memory.

The term *dementia* refers to the loss of cognitive functions, particularly memory, which is so serious as to interfere with everyday life. Alzheimer's disease is the most common form of dementia. In the brains of patients with Alzheimer's disease, the deposition of the amyloid protein and the death of neurons in the cortex is observed. Radiological examinations show, however, also a damage of the white matter, that part of the brain which is instead mainly constituted by myelin. The damage of the white matter seems to be a crucial element in the pathogenesis of Alzheimer's disease and the correlation between the levels of amyloid in the liquor and the lesion extension seems to suggest a direct link between the amyloid pathology and the damage of the cerebral white matter, which produces "premature aging" [2]. This observation underlines the importance of the evaluation of the overall state of the brain (white–grey matter ratio) in a disease that has always been considered primarily linked to the degeneration of neurons, and opens the way to new techniques of early prognosis and to the identification of new therapeutic targets.

The premature aging of the brain is therefore an alarm for the onset of neurodegenerative diseases [3, 4, 5]. Predictive neuroimaging models can be used to learn the brain age in healthy people. In the case of a suspected case of early dementia, a comparative assessment of the brain age can be made in relation to what is estimated for healthy peers. Moreover, if the estimated age based on brain Magnetic Resonance Imaging (MRI) is significantly greater than the actual age of an individual, this may reflect an unusual accumulation of age–related changes in the brain, an effect that can be quantified simply by comparing the actual with the predicted age. This approach has been adopted in several studies that correlate the presence of neurodegenerative pathologies with an increase in the expected brain age [6, 7]. Similar approaches have also been used to demonstrate the protective influence of meditation [8], physical activity and education [9] on brain aging. Thus, it is easy to realize that the accurate prediction of the brain age can have a great clinical relevance [10].

In this paper we propose a new approach for the brain age estimation from 3D NMR images using deep Convolutional Neural Networks (CNNs). In recent years, CNNs have been used in a wide range of domains and especially in computer vision, where they have produced impressive results [11, 12, 13, 14, 15, 16] Also in the analysis of medical images, CNNs have been widely adopted [17, 18, 19] while, for brain age estimation, CNNs have been used in [20, 10, 21]. Because of the three–dimensional structure of the data, to analyze NMR brain images, 2D convolutions are commonly replaced with 3D convolutions. Unfortunately, the use of 3D convolutions introduces a significant increase in the computational load.

To address this problem, we propose to replace 3D with 2D convolutions, thereby substantially reducing memory and computational requirements. A wide range of experiments was conducted, testing two different approaches:

- The 3D image is decomposed based on axial, sagittal and coronal views, which are processed independently by three different CNNs. The encoded representations are then combined by two successive dense layers that produce the age prediction.

- All sections (slices) along the depth of the 3D image are fed into a pre–trained CNN that produces a coded representation of the slices. The CNN output is then concatenated and processed sequentially by a Bidirectional Long Short Term Memory (BLSTM) [22, 23]. Finally, the output of the BLSTM is the input for a dense level that generates the age prediction.

Our best–scoring architecture is compared to a state–of–the–art 3D–CNN, showing its ability to achieve competitive results in terms of accuracy, but requiring significantly less computational resources. In particular, if the hardware configuration is limited to a single GPU, our model exceeds the 3D–CNN model.

The paper is organized as follows. In Section 2, the CNN architectures, used for brain NMR processing, are presented. The employed dataset, the proposed experimental setup and obtained results are described in Section 3. Finally conclusion are drawn in Section 4.

2. Models

This section introduces all the models used in this study. In particular, in Section 2.1, the 3D convolutional model used as a baseline is presented, while the 3Way–Net and the slice by slice approach based on both CNNs and Long–Short–Memories are described in Section 2.2 and Section 2.3, respectively.

2.1. 3D–CNN

In this paper, the 3D–CNN model for predicting age from NMR brain images proposed by [10] is used as the baseline. The 3D–CNN is composed of five identical blocks, each consisting of a 3D convolution with kernel size of $3 \times 3 \times 3$ and stride 1, followed by a ReLu non–linearity. A second identical 3D convolutional layer is connected in cascade to the first, followed by a batch normalization operation to which a ReLu activation function is applied. Finally, a $2 \times 2 \times 2$ max–pooling operation is used to reduce the size of the feature maps. In particular, the first block has eight feature maps, a number that is doubled block by block (with the last one composed by 128 feature maps). Lastly, two fully connected layers, with ReLu activation functions, are used to perform age prediction. The entire network architecture is shown in Figure 1.

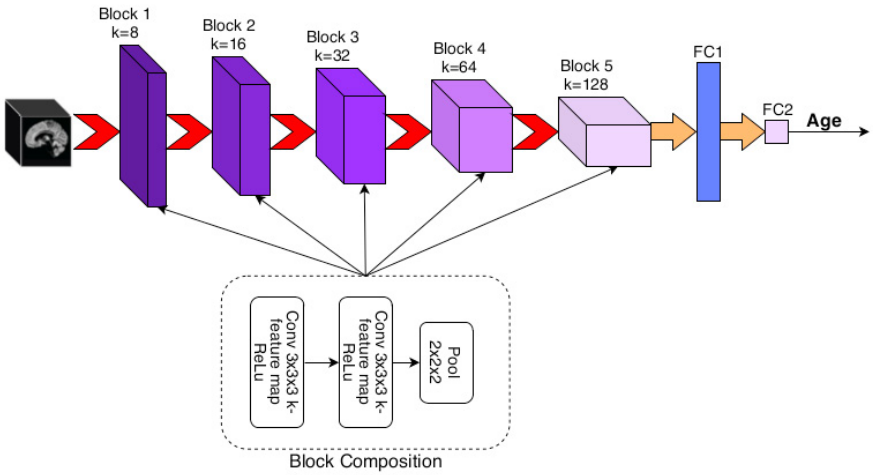

Fig. 1. 3D–CNN architecture proposed in [10].

2.2. 3Way–Net

The logic behind the approach described below is inspired by the idea of obtaining a representation of a three–dimensional object through its orthogonal projections on Cartesian planes. Therefore, from the 3D NMR images, three views — axial, sagittal and coronal[1] — were extracted and processed independently, using three different 2D convolutional networks. Each view includes a sequence of 2D images with different shapes. The sequence of images is treated as a channel from the network. This means that, for the axial plane, we process images composed of $218 \times 182 \times 182$ pixels, for the sagittal plane of $182 \times 182 \times 218$ pixels, and for the coronal plane of size $182 \times 218 \times 182$. To produce the age prediction, the outputs of the three CNNs are concatenated and fed into two fully connected layers, each of which is followed by a ReLu activation function. The proposed architecture, called 3Way–Net, is represented in Figure 2.

To implement the 3Way–Net, two different types of convolutional neural networks were used.

[1] The axial plane (lateral, horizontal) divides the brain into the upper and lower sides, the last including the cerebellum. The sagittal plane (longitudinal, anteroposterior) is a plane parallel to the sagittal suture. It divides the brain into left and right. The coronal (vertical) plane divides the brain into front and back.

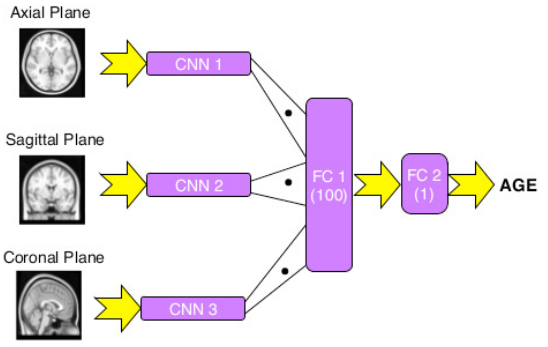

Fig. 2. The architecture of the 3Way–Net.

1. *3Way–CNN* — The first architecture is a simple CNN with four identical blocks. Each block is composed of four 3×3 convolutions with stride 1, with batch normalization and ReLu activation functions. The number of feature maps doubles each time (16, 32, 64, 128), while a max–pooling operation, with kernel size 2×2, is used to reduce their size. A distinct CNN model is implemented for each plane extracted from the 3D NMR image.

2. *3Way–ResNetlike* — The second evaluated model is inspired by the ResNet architecture proposed by [12]. In the ResNet architecture a skip connection across convolutional layers is used to combine different information and to better back–propagate the gradient. The proposed model employs three ResNet blocks with four basic units each. These units consists of two 3×3 convolutions with stride 1, with batch normalization and ReLu activation functions. A skip connection is used to sum the output of an unit with the output of the previous unit. To obtain dimensionality reduction in the first convolution of each block, a stride of 2 is employed. Moreover, in the first skip connection, to match the dimensionality between the input and the output of the unit, a 3×3 convolution with stride 2 is also applied. Finally, the last block is followed by a 2×2 pooling operation. A distinct architecture similar to ResNet is employed for each plane extracted from the NMR image.

2.3. Slice–By–Slice analysis with CNNs and Bidirectional LSTMs

The main idea behind this method is to consider the slices (or sections) of an NMR image as a sequence. The sagittal plane was used as the reference plane. A pre–trained CNN is employed to extract some features from each slice, constituting a sequence which is processed by a bidirectional LSTM (BLSTM). BLSTMs are a particular type of recursive networks based on the Long–Short–Term–Memory architecture [24], which have been proven effective in addressing the problem of long–term dependencies[25, 26]. Conventional LSTMs are able to use only the previous information, i.e. they process temporal (or sequential) data following their natural flow. BLSTMs, on the other hand, analyze the data in both directions (considering the context of each element within the sequence), using two separate hidden layers. Three models have been proposed, based on the use of different CNN architectures, pre–trained on the Image–Net Dataset[27], to act as feature extractors, namely:

- VGG16 [11] for *SbS–R–VGG16*;
- ResNet [12] with 50 layers for *SbS–R–ResNet50*;
- DenseNet [13] for *SbS–R–DenseNet121*.

Three consecutive slices were concatenated to obtain 3–channel images to be fed into the pre–trained networks. The output of the CNNs are the features extracted from this group of three slices. All slices are processed in this way and the extracted features are analyzed as a sequence by a BLSTM with 50 internal units with ReLu activation. Finally, the age prediction task is performed by two fully connected layers, each of which followed by a ReLu activation function. The overall SliceBySlice (SbS–R–CNN) architecture is shown in Figure 3.

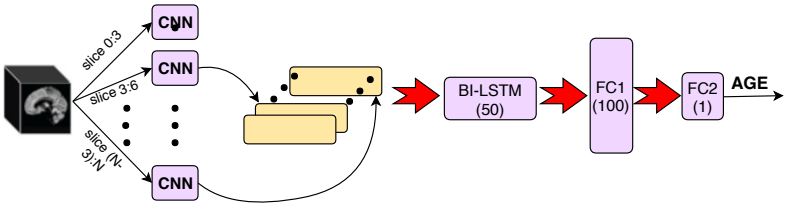

Fig. 3. The architecture of the SbS–R–CNN.

3. Experimental setup and results

The experiments conducted to validate the proposed approaches are described below. In particular, Section 3.1 describes both the dataset and the preprocessing designed to obtain images of the brain without the skullcap and therefore containing only functional information for the problem to be solved (white and grey matter). Subsequently, in Section 3.2, the experimental setup is introduced, while Section 3.3 gathers the obtained results, which demonstrate how our methods guarantee performances similar to 3D methods, with a lower training time and a modest memory occupation.

3.1. Dataset

3.1.1. Information eXtraction of Images

The IXI dataset (Information eXtraction of Images[2] collects 600 images of healthy subjects from three major hospitals in London (the Hammersmith Hospital, the Guy's Hospital and the London Institute of Psychiatry), gathered with three different acquisition tools (Philips 3T, Philips 1.5T and GE). The dataset contains different types of NMR sequences (T1, T2, FLAIR, etc.) with the corresponding metadata (gender, age, etc.). In this paper, we consider only the T1–weighted images as in [28], also reducing the dataset size to 561, due to the absence of the age annotation for the remaining 38 samples. Figure 4 reports the distribution by age of the IXI dataset, being the average age of patients 48.65, with a standard deviation of 16.45, and a range of variation from 20 to 86 years. Each 3D NMR is composed of

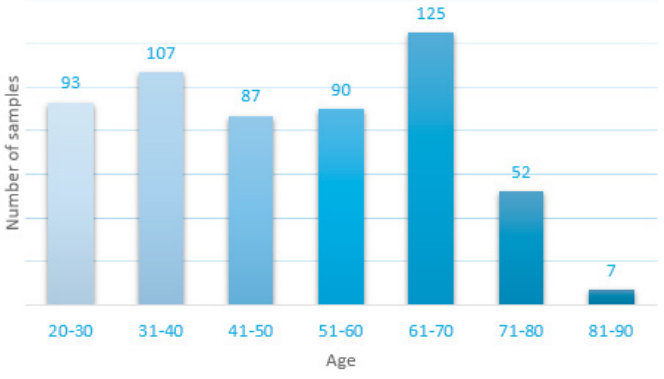

Fig. 4. Distribution by age of the IXI dataset.

a set of images, each of which represents a 1 mm portion of the brain and has a shape of 256×256 pixels. Depending on the acquisition system, each NMR has a different number of slices (130, 140 or 150).

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

3.1.2. Pre–Processing

The samples were first normalized using a co–registration operation, which is a standard pre–processing step in brain image analysis. This operation consists in maximizing the overlapping voxels between the current image and a reference template model, relying on the Normalized Correlation Coefficient (NCC). The overlapping region is defined as:

$$X_0 = \{x_0 : x_0 \in X \cap T(X')\} \tag{1}$$

where X is the template image, X' is the target image, and T is a rigid body transformation. The NCC of $F(X_0)$ and $G(X_0)$, representing the intensity set of the overlapping region in the template and in the target image, respectively, can be defined as:

$$NCC(F,G) = \frac{1}{N_0^2} \frac{\sum_{x_0 \in X_0} (f(x_0) - \bar{f})(g(x_0) - \bar{g})}{\sigma_f \sigma_g} \tag{2}$$

with $\bar{f}$, $\bar{g}$, σ_f and σ_g which represent the mean and the standard deviation of the voxel intensity in $F(X_0)$ and $G(X_0)$, respectively, and $N_0 = |X_0|$. In this work, the MNI152–T1 Weighted [29] template has been used, which is the standard template for T1 NMR. After the pre–processing step, we obtain a set of images of size $182 \times 218 \times 182$.

In order to obtain NMR images without the skull, the Brain Extraction Tool (BET) has been used [30]. This allows us to create two datasets, one containing the entire head (brain and skull) and the other containing only the brain (see Figure 5 for an example of the two types of images). Both datasets consist of 561 images with the same voxel level (between 0 and 1) and the same size. Each data set (head and brain) is divided into a training, a validation and a test set, containing 447, 56 and 58 images, respectively (80% training, 10% validation and 10% test).

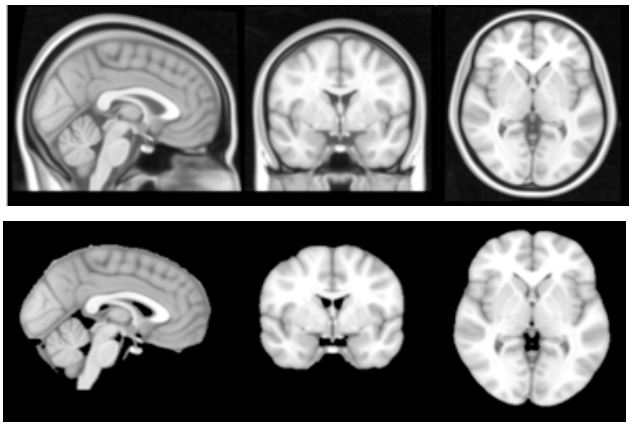

Fig. 5. NMR images of the whole head (top) and of the brain (bottom); brain images are obtained after the skull–stripping process.

3.2. Experimental setup

The analysis conducted in this paper is aimed at comparing different CNN models, computationally cheaper than the algorithm based on 3D convolutions proposed in [10], to perform age prediction from NMR brain images. The comparison mainly focuses on performance, training time, and memory load. In fact, our goal is to find the best model that can be trained on a single Nvidia GTX 1080. Unfortunately, in [10], a dataset that is not publicly available was used in the experiments and, for this reason, a fair comparison required to retrain and test the 3D–CNN on the IXI dataset. To satisfy the data memory constraints of the only available GPU, we were forced to use a batch size of 2

instead of 32 when dealing with 3D convolutions. All models were trained based on the Adam optimizer [31], with an initial learning rate of 0.001, using the Root Mean Square Error (RMSE) loss function (see Eq. 3):

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(\hat{y} - y)^2}{N}} \tag{3}$$

where y and $\hat{y}$ are the target and the predicted value, respectively, and N is the number of samples. The training is stopped if the loss on the validation set does not decrease for at least 30 epochs or when the training reaches a predefined maximum number of epochs (200). The test evaluation has been performed using the Mean Absolute Error (MAE) metric (see Eq. 4), as in [10]:

$$MAE = \frac{\sum_{i=1}^{N}|\hat{y} - y|}{N} \tag{4}$$

We repeated each experiment three times for all the compared models, in order to provide reliable results.

3.3. Experimental results

This section reports the results obtained with our proposed architectures on the IXI dataset, for age estimation based on brain NMR images. In Table 1, our approaches are compared with the state–of–the–art 3D–CNN presented in [10].

Model	MAE Brain (Var)	MAE Head (Var)	Time per epoch (m)	Training time (h)
3D–CNN	6.76 (0.32)	6.89 (0.41)	20	70
3Way–CNN	6.99 (0.37)	7.93 (0.64)	4	4.5
3Way–ResNetlike	6.61 (0.27)	8.29 (0.65)	5	7
SbS–R–VGG16	5.94 (0.32)	6.15 (0.32)	1	1.66
SbS–R–ResNet50	6.86 (0.36)	7.15 (0.18)	1	1.58
SbS–R DenseNet121	6.17 (0.41)	6.32 (0.26)	1	1.5

Table 1. Results.

When the hardware is tied to a single GPU configuration, our best scoring architecture (SbS–R–CNN) significantly outperforms [10], with a 0.82 year improvement in terms of mean absolute error over the 3D–CNN. This suggests that when hardware availability is limited, our method is a viable alternative for processing 3D NMR brain images. Furthermore, the training time is substantially reduced compared to the 3D–CNN. The comparison between the SbS–R–CNNs and the 3Way–Net adopted in this study shows that all SbS–R–CNNs provide better results with reduced training times. Whereas the 447 available training images could not be enough to train a very deep network as the 3Way–Net or the 3D–CNN due to the huge number of parameters. The performances of the different models used as feature extractors in the SbS–R–CNN architectures are also compared. It can be seen that VGG–16 provides better results than ResNet–50 and DenseNet–121. Finally, it is worth noting that even the best methods for assessing the age, starting from brain images, still produce significant errors (of the order of 6 years). In the specific case of the IXI dataset, this is due to the presence of a very significant percentage of samples relating to over 40 people. If, in fact,

the age estimation can be performed with high precision (less than two years) in the case of young people [32], the difficulty of estimation increases exponentially with increasing age, since even in healthy individuals the aging pattern can be significantly differentiated, mainly due to lifestyle.

4. Conclusions

In this paper, we have proposed some new approaches for age estimation, based on brain MRI using deep Convolutional Networks. Because of the structure of the data, MRI is normally performed on the basis of 3D convolutions, which implies a considerable memory load and takes a long time. To this end, some *ad hoc* methods have been examined based on 2D convolutions to optimize both memory consumption and the time required for training. The preliminary experimental results are really promising, showing how the SbS–R–CNN can outperform a state–of–the–art 3D–CNN in the case of a hardware configuration limited to a single GPU, opening the possibility for small health institutions to apply powerful methods for the early diagnosis of neurodegenerative diseases without huge investments. Finally, the proposed approaches can be easily generalized to different applications based on 3D images.

References

[1] Ziegler, D. A., Piguet, O., Salat, D. H., Prince, K., Connally, E., Corkin, S. (2010). "Cognition in healthy aging is related to regional white matter integrity, but not cortical thickness." *Neurobiology of aging*, **31**(11), 1912-1926.

[2] Pietroboni, A. M., Scarioni, M., Carandini, T., Basilico, P., Cadioli, M., Giulietti, G., ... Fenoglio, C. (2018). "CSF -amyloid and white matter damage: a new perspective on Alzheimers disease." *J Neurol Neurosurg Psychiatry*, **89**(4), 352-357.

[3] Schnack, H. G., Van Haren, N. E., 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**(6), 607-616.

[4] Cole, J. H., Underwood, J., Caan, M. W., De Francesco, D., van Zoest, R. A., Leech, R., ... van der Loeff, M. F. S. (2017). "Increased brain-predicted aging in treated HIV disease." *Neurology*, **88**(14), 1349-1357.

[5] Pardoe, H. R., Cole, J. H., Blackmon, K., Thesen, T., Kuzniecky, R., Human Epilepsy Project Investigators. (2017). "Structural brain changes in medically refractory focal epilepsy resemble premature brain aging." *Epilepsy research*, **133**, 28-32.

[6] Cole, J. H., Leech, R., Sharp, D. J., Alzheimer's Disease Neuroimaging Initiative. (2015). "Prediction of brain age suggests accelerated atrophy after traumatic brain injury." *Annals of neurology*, **77**(4), 571-581.

[7] Franke, K., Gaser, C., Manor, B., Novak, V. (2013). "Advanced BrainAGE in older adults with type 2 diabetes mellitus." *Frontiers in aging neuroscience*, **5**, 90.

[8] Luders, E., Cherbuin, N., Gaser, C. (2016). "Estimating brain age using high-resolution pattern recognition: younger brains in long-term meditation practitioners." *Neuroimage*, **134**, 508-513.

[9] Steffener, J., Habeck, C., O'Shea, D., Razlighi, Q., Bherer, L., Stern, Y. (2016). "Differences between chronological and brain age are related to education and self-reported physical activity." *Neurobiology of aging*, **40**, 138-144.

[10] Cole, J. H., Poudel, R. P., Tsagkrasoulis, D., Caan, M. W., Steves, C., Spector, T. D., Montana, G. (2017). "Predicting brain age with deep learning from raw imaging data results in a reliable and heritable biomarker." *NeuroImage*, **163**, 115-124.

[11] Simonyan, K., Zisserman, A. (2014). "Very deep convolutional networks for large-scale image recognition." *arXiv preprint arXiv:1409.1556*.

[12] He, K., Zhang, X., Ren, S., Sun, J. (2016). "Deep residual learning for image recognition." *In Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770-778).

[13] Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K. Q. (2017). "Densely connected convolutional networks." *In Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 4700-4708).

[14] Szegedy, C., Ioffe, S., Vanhoucke, V., Alemi, A. A. (2017). "Inception-v4, inception-resnet and the impact of residual connections on learning." *In Thirty-First AAAI Conference on Artificial Intelligence*.

[15] Long, J., Shelhamer, E., Darrell, T. (2015). "Fully convolutional networks for semantic segmentation." *In Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 3431-3440).

[16] Ren, S., He, K., Girshick, R., Sun, J. (2015). "Faster r-cnn: Towards real-time object detection with region proposal networks." *In Advances in neural information processing systems* (pp. 91-99).

[17] Caruana, R., Lou, Y., Gehrke, J., Koch, P., Sturm, M., Elhadad, N. (2015). "Intelligible models for healthcare: Predicting pneumonia risk and hospital 30-day readmission." *In Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 1721-1730). ACM.

[18] Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., Thrun, S. (2017). "Dermatologist-level classification of skin cancer with deep neural networks." *Nature*, **542**(7639), 115.

[19] Andreini, P., Bonechi, S., Bianchini, M., Mecocci, A., Scarselli, F. (2018). "A Deep Learning Approach to Bacterial Colony Segmentation." *In International Conference on Artificial Neural Networks* (pp. 522-533). Springer, Cham.

[20] Qi, Q., Du, B., Zhuang, M., Huang, Y., Ding, X. (2018, December). "Age Estimation from MR Images via 3D Convolutional Neural Network and Densely Connect." *In International Conference on Neural Information Processing* (pp. 410-419). Springer, Cham.

[21] Huang, T. W., Chen, H. T., Fujimoto, R., Ito, K., Wu, K., Sato, K., ... Aoki, T. (2017, April). "Age estimation from brain MRI images using deep learning." *In 2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)* (pp. 849-852). IEEE.
[22] Schuster, M., Paliwal, K. K. (1997). "Bidirectional recurrent neural networks." *IEEE Transactions on Signal Processing*, **45**(11), 2673-2681.
[23] Baldi, P., Brunak, S., Frasconi, P., Soda, G., Pollastri, G. (1999). "Exploiting the past and the future in protein secondary structure prediction." *Bioinformatics*, **15**(11), 937-946.
[24] Hochreiter, S., Schmidhuber, J. (1997). "Long short-term memory." *Neural computation*, **9**(8), 1735-1780.
[25] Graves, A., Mohamed, A. R., Hinton, G. (2013, May). "Speech recognition with deep recurrent neural networks." *In 2013 IEEE international conference on acoustics, speech and signal processing* (pp. 6645-6649). IEEE.
[26] Bahdanau, D., Cho, K., Bengio, Y. (2014). "Neural machine translation by jointly learning to align and translate." *arXiv preprint arXiv:1409.0473*.
[27] Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., Fei-Fei, L. (2009). "Imagenet: A large-scale hierarchical image database." *In 2009 IEEE conference on computer vision and pattern recognition* (pp. 248-255). IEEE.
[28] Franke, K., Ziegler, G., Klppel, S., Gaser, C., Alzheimer's Disease Neuroimaging Initiative. (2010). "Estimating the age of healthy subjects from T1-weighted MRI scans using kernel methods: exploring the influence of various parameters." *Neuroimage*, **50** (3), 883-892.
[29] Evans, A. C., Janke, A. L., Collins, D. L., Baillet, S. (2012). "Brain templates and atlases." *Neuroimage*, **62**(2), 911-922.
[30] Smith, S. M. (2002). "Fast robust automated brain extraction." *Human brain mapping*, **17**(3), 143-155.
[31] Kingma, D. P., Ba, J. (2014). "Adam: A method for stochastic optimization." *arXiv preprint arXiv:1412.6980*.
[32] Franke, K., Luders, E., May, A., Wilke, M., Gaser, C. (2012). "Brain maturation: predicting individual BrainAGE in children and adolescents using structural MRI." *Neuroimage*, **63**(3), 1305-1312.