2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

Nonlinear registration as an effective preprocessing technique for Deep learning based classification of disease

Daiki Fujibayashi[1], Hiromasa Sakaguchi[1], Ilya Ardakani[1] and Akihiro Okuno[1]

Abstract— **A number of machine learning (ML), and particularly in recent years, deep learning (DL) approaches have been proposed for automatic classification of Alzheimer's disease (AD) using brain structural magnetic resonance imaging (MRI) data. However, the data available are limited in the case of this specific disease. Training a DL model with a large number of feature parameters on a small dataset of MRI scans will likely lead to overfitting. Overfitting reduces the generality and efficiency of the model. In this study, we show that a traditional nonlinear transformation from native space to template space, as a preprocessing stage, is effective in reducing overfitting through the reduction of spatial variations in the input data. To evaluate this effectiveness, we compare two different preprocessing approaches for DL-based AD classification task: (1) affine registration and (2) nonlinear diffeomorphic anatomical registration using exponentiated Lie algebra (DARTEL). The results show that the accuracy of the nonlinear registration based approach is much higher than the affine registration based approach. Furthermore, from the classification results obtained with noisy images, DARTEL is less susceptible to noise than affine registration. In summary, our experimental results suggest that nonlinear transformation is a preferable preprocessing step for training DL-based AD classification models on limited size datasets.**

I. INTRODUCTION

Alzheimer's disease (AD) is the most common form of dementia that has become a major public health concern. The number of people with dementia has been increasing every year around the world, and is predicted to reach 115.4 million in 2050[1]. With the increasing number of patients, automatic diagnostic techniques to accurately detect the early signs of AD have been attracting more attention. One of these techniques is using brain magnetic resonance imaging (MRI) scans of patients. As the disease progresses, structural brain changes occur, such as local volume atrophy[2], occur and MRI scans are commonly used in clinical settings to observe them. Hence, considerable work has been expended to detect AD features from MRI data.

In the past, many machine learning (ML) based techniques such as support vector machines and random forests have been proposed for automatic classification of patients with AD from healthy normal controls (NC) using MRI data[3][4]. These ML-based approaches usually select and use specific features like sets of volumes of regions of interest (ROIs) that have already been reported to have associations with AD pathology. However, the performance of these methods is quite sensitive to how and which ROIs are selected within the images. Given that the pathology of AD is not fully understood and features in images caused by the disease are also not clear, identifying effective ROIs for its classification is a challenging task.

In recent years, deep learning (DL) has produced interesting results in the medical brain imaging field[7][8][9]. DL can capture features in high-dimensional spaces. These DL methods are likely to learn features required for the classification without explicitly defining ROIs in the images. It has been reported that DL-based methods produce more accurate results than conventional ML-based methods. Therefore, it is likely that a learned high-dimensional representation of data can be much more useful in the characterization of AD compared to the traditional ROI-based features.

In general, to train DL models with large number of parameters considerably large datasets are required. For example, a dataset of 1.2 M images was used for DL based classification of two-dimensional (2D) image in[5]. In the case of AD, even one of largest publicly available datasets from the Alzheimer's Disease Neuroimaging Initiative (ADNI) has only approximately 1000 subjects with structural MRI. As the number of dimensions and model parameters increases, the overfitting of training data becomes a major issue[12]. This is a common problem in the field of medical image analysis[13].

Other than using a larger dataset, one possible and effective approach to avoid overfitting is to reduce feature variations irrelevant to the classification. The gray matter (GM) volume in specific regions is reported to exhibit a strong association with AD; however, there is little evidence that other morphometric changes, such as the shape of the cortex, is related to AD. Although setting multiple ROIs is one possible dimensionality reduction method, useful features (other than those contained in the ROIs) might be ignored. In due to high learning capacity, the characterizations unnecessary for classification are also learned, thus making it easier to overfit. Therefore, it is important to reduce the features that are not necessary for classification while leaving the features that are useful. We propose that the "volume map" representation of a 3D voxel image, which can be obtained by a traditional nonlinear registration technique known as large deformation diffeomorphic metric mapping (LDDMM), is an effective

*This work was not supported by any organization

[1]Daiki Fujibayashi is a researcher and Machine Learning Engineer, Splink,inc., Tokyo Chiyoda Kasumigaseki 3-3-2, Japan fujibayashi@splinkns.com

[1]Hiromasa Sakaguchi is a researcher and Machine Learning Engineer, Splink,inc., Tokyo Chiyoda Kasumigaseki 3-3-2, Japan sakaguchi@splinkns.com

[1]Ilya Ardakani is a researcher and Machine Learning Engineer, Splink,inc., Tokyo Chiyoda Kasumigaseki 3-3-2, Japan ardakani@splinkns.com

[1]Akihiro Okuno is a CTO, Splink,inc., Tokyo Chiyoda Kasumigaseki 3-3-2, Japan okuno@splinkns.com

dimensionality reduction preprocessing method used for DL-based classification.

In LDDMM, a 3D voxel image is nonlinearly deformed to a template image by diffeomorphic metric. Through this, the shape information of an individual brain can be reduced, while the spatial information of local volumes is maintained[14]. In this study, we show that LDDMM is useful in reducing overfitting for the DL-based AD classification. Two types of representations of input images are compared in training the DL model: (1) affine transformation of a whole brain image to a standard brain image, which is the one of most often used methods for conventional linear registration, and (2) a GM volume map transformed by Diffeomorphic Anatomical Registration using Exponentiated Lie Algebra (DARTEL), one of the most popular LDDMM algorithms for nonlinear registration. To evaluate the effectiveness of the preprocessing for each of linear transformation and non-linear transformation, we performed the AD classification tasks and assessed based on the impact on two criteria: (1) overfitting owing to the dependency of the number of samples and (2) model robustness to noise.

Overall, our contributions are as follows:

1) We demonstrate that GM volume maps nonlinearly registered to the template brain using DARTEL constitute an effective representation of the data for the DL-based AD classification model, especially with limited sample sizes.
2) For noise-added data, the model trained on the nonlinearly registered data was found to have higher performance than the one trained on the linearly registered data. In other words, the nonlinear transformation was more robust to additive noise.

This study is organized as follows. Related studies are presented in section 2. Details of the data and methods used in our experiments are given in section 3. The results are presented in sections 4. The results are discussed in sections 5. The conclusions are outlined and future work is discussed in section 6.

II. Related work

AD classification research using DL has been conducted extensively. Manhua et al. constructed an original convolutional neural network (CNN) to classify AD. Aiming for high accuracy even with a limited number of MRI scans, the researchers inputted only the hippocampal region as this is one of the regions where brain atrophy occurs in patients with AD[16]. The authors randomly selected 449 participants. The proposed method achieved an accuracy of 88.9% for the classification of AD vs. NC. Mingxia et al. proposed a method known as the LDMIL framework that avoids fixed ROI settings. It extracts significant feature regions from the data and classifies AD vs. NC by DL using both local and global features of brain MRI as input[17]. Liu et al. performed affine transformation preprocessing to the template brain with brain images from 397 subjects from the ADNI database. The preprocessed brain was divided into multiple patches, and learning was performed by combining 3D CNNs and 2D CNNs[15]. Silvia et al. performed a DARTEL transformation on the MRI images of the ADNI dataset and a dataset owned by the University of Wits-Salute-San-Flafaere in Milan, classifying with DL[18]. The ADNI dataset comprised 1409 subjects in total, and the other dataset with 229 subjects were divided randomly for learning and testing. MRI preprocessing was conducted with the DARTEL algorithm, which uses statistical parametric mapping to achieve standardization on the Montreal Neurological Institute (MNI) space[19].

There has been a lot of work on AD/NC classification. However, the effectiveness of main preprocessing technique has not yet been compared and examined. In this study, we compare the diagnostic outcomes of modeling using two major preprocessing methods, nonlinear transformation (DARTEL) and linear transformation (affine).

III. Materials and Methods

A. Dataset

This study uses the ADNI dataset for training and validation for model selection. We randomly selected 600 subjects with diagnostic labels "AD" and "NC" (AD: 300, NC: 300) form ADNI1, ADNI2, and ADNI-GO projects(Table I). We used MRI images and diagnostic labels corresponding to the first diagnosis. Details of the MRI acquisition protocol for the ADNI dataset can be found on the ADNI's official webpage. An independent dataset of 100 subjects (AD: 50, NC: 50) from the Australian Imaging, Biomarker & Lifestyle Flagship Study of Ageing (AIBL) project (TableII), comprising data of more than 1100 people spanning 4–5 years, was used for the testing and comparison of the selected models. The MRI acquisition protocol for the AIBL dataset is adopted from ADNI.

TABLE I
SUMMARY OF ADNI DATA SET(ABBREVIATION:AD=ALZHEIMER'S DISEASE, NC=HEALTHY CONTROL, MMSE=MINI-MENTAL STATE EXAMINATION)

	AD	NC
Number	300	300
Age	75.1 $\pm$ 7.7	74.8 $\pm$ 5.8
Gender(F/M)	133/167	148/152
MMSE	23.16 $\pm$ 3.09	29.04 $\pm$ 1.18

B. Preprocessing

We used the ADNI as a learning dataset (train and validation) and the AIBL as the test dataset. Although the basic structure of the brain is common among individual, the size of the tissue structure of the brain varies between individuals

TABLE II
SUMMARY OF AIBL DATA SET

	AD	NC
Number	50	50
Age	72 $\pm$ 7.9	77.52 $\pm$ 5.8
Gender(F/M)	29/21	30/20

and races. Therefore, when comparing brain structures in several individuals or groups, it is important to handle them in the same coordinate space. We chose two types of transformation algorithms in brain image registration.

In affine registration, the brain images were standardized with the MNI template brain (affine dataset). First, to minimize the effect of magnetic field inhomogeneity on MRI images with different imaging conditions, bias-field correction was performed using N4 algorithm[20]. Subsequently, brain extraction processing was performed for each subject using the mask image of the brain region provided by OASIS[21]. Finally, a rigid body affine transformation was used to perform a convergence calculation that matched the individual brain to the MNI standard brain[22]. The standard brain size used in this experiment was $64 \times 64 \times 64$.

3DT1-DARTEL standardized the brain images[23][24](DARTEL dataset). First, using the learning data group, the standardized parameters used for DARTEL were calculated from the images separated into gray and white matter. Subsequently, using the calculated standardization parameters, anatomical standardization was performed on the learning dataset. A 3DT1-DARTEL template was created by averaging and smoothing the learning data groups that were anatomically standardized with the 3DT1-DARTEL method. A 3D Gaussian filter was used as the smoothing filter, where the full-width-at-half-maximum (FWHM) was set to 8 mm. Using the 3DT1-DARTEL template, anatomical standardization was performed on the learning and test dataset groups. In this experiment, we used the volume map image of GM that had been transformed to the 3DT1-DARTEL template.

C. Gaussian Noise

In this study, we investigate the classification accuracies corresponding to different noise intensities. The purpose of this method was to verify the effects of noise on the accuracy of the model, assuming that noisy images occur in clinical situations. Thus, each model was trained and validated on the dataset with additive noise, and its classification accuracy was compared with that of the trained model without noise.

In this experiment, Gaussian noise was used to artificially reproduce the noisy data that occurs in clinical settings. We used the algorithm implemented in TorchIO[25] version 0.17.47. The formula implemented is as follow. To change the strength of Gaussian noise added to the data, the standard deviation of the distribution was varied from 0 to 200 in incremental steps of 50 .

$$\mathcal{N}(\mu, \sigma^2)\,.$$

D. DL model

In this study, we used Residual Networks (ResNet), which is one of the DL models used extensively in the computer vision field. ResNet is a type of CNN designed to prevent loss of features during the convolution process compared with basic CNNs. It often results in higher identification accuracy than basic CNNs for image classification problems. ResNet's structure contains skip connections, which makes it possible to suppress the loss of features of the training data even in deep CNN models[27]. We used ResNet50 in this study. In order to make up for this small amount of data, we employed transfer learning, a common technique used for training DL models. The weights of our networks were pre-trained on the ImageNet dataset[28].

E. Experiments

In this study, we prepared six datasets where the ratio of AD to NC was kept at 1:1. Each dataset selected randomly from the ADNI dataset. And 80% of each dataset was used for training and 20% was used as validation set for model selection. The selected models were tested with the AIBL dataset. The model was implemented in Pytorch[29], an open source machine learning framework, and the experiment was conducted on a PC equipped with a graphics processing unit NVIDIA K80. We calculated the classification accuracy, sensitivity, and specificity to evaluate the classification performance. Gaussian noise was added to the original images and the same classification experiment was performed. We evaluated the classification accuracy for each intensity of the noise.

IV. Results

Fig.1 shows the accuracies of the models on the AIBL dataset. In this experiment, we found that a higher classification accuracy was obtained by training with the DARTEL dataset. The classification accuracy was 0.66–0.77 when the affine dataset was used, whereas the classification accuracy ranged from 0.82 to 0.87 when the DARTEL dataset was used.

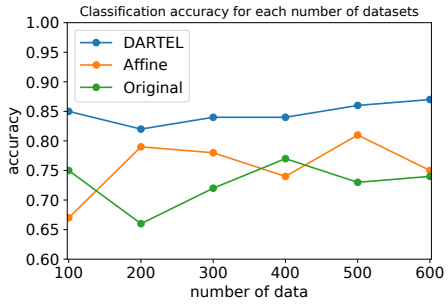
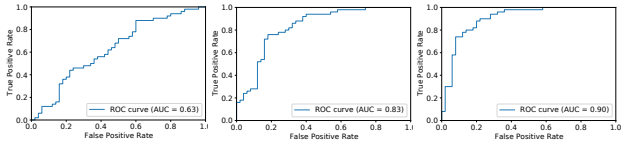


Fig. 1. Transition of classification accuracy when each preprocessing and the number of images are changed.

Fig.2 shows the receiver operating characteristic curve generated from the Softmax output values for the test data, where the DARTEL dataset showed high area under the curve (AUC) values in all experiments. Fig.2 shows sample results from the experiments in which 300 subjects were used for training. The AUC for the original whole brain was 0.63, the AUC for the affine transformed whole brain was 0.83, and the AUC for volume maps was 0.90. The accuracy of the model created using the DARTEL dataset was the highest.

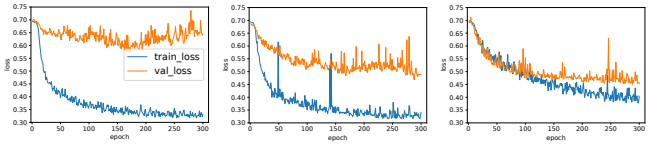
Fig.3 shows the log of the loss function using 300 subjects of training and validation data. When preprocessing the original whole brain images, the affine transformation



(a)Original (b)Affine (c)DARTEL

Fig. 2. Evaluation of the model created using the 300 training datasets for each process.

showed a large difference between validation and training losses. Conversely, when DARTEL was used on whole brain images, the loss value for the validation data converged.



(a)Original (b)Affine (c)DARTEL

Fig. 3. The logs of loss function that the models created by the 300 brain images for each preprocess.

Fig. 4 shows the results of each training log when 4 (a), 4 (b), and 4 (c) are limited to the training data sets of 100, 300, and 600, and DARTEL is used for preprocessing. It can be confirmed that overfitting is reduced by increasing the number of training data. However, comparing the 300 training data with the 600 training data, there was no significant improvement in the value of the loss function.

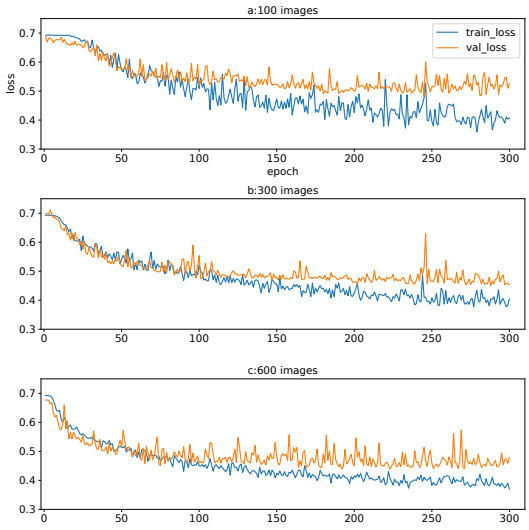
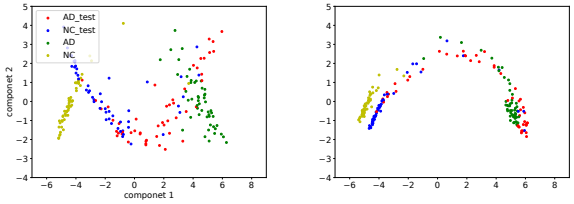


Fig. 4. The logs of loss function

To confirm whether each preprocessing method could extract the features required for classification, we visualized the input of the trained model's Softmax function. The results associated with the extraction of the features just before the final layer on the learning and test data, and the performance of principle component analysis (PCA) are shown in Fig.5. A plot is shown for each patient with the first and second principal components set on an axis after PCA. The distributions of the training and validation data after a linear transformation were different, but the distributions of the training and validation data after a nonlinear transformation were similar.



(a)Affine (b)DARTEL

Fig. 5. PCA with hidden layer outputs for training and test data

Fig.6 shows the results of training and validation of the original noisy images after the respective preprocessing. The data used were the same as in the above experiment, and the results are shown for 300 subjects ' training data. It can be confirmed from Fig.6 that the classification accuracy against the noise intensity decreased from 70% to 56% in the case of affine registration preprocessing methods. However, with DARTEL registration, it was confirmed that the classification accuracy of 69%–75% was maintained even with increased noise intensity.

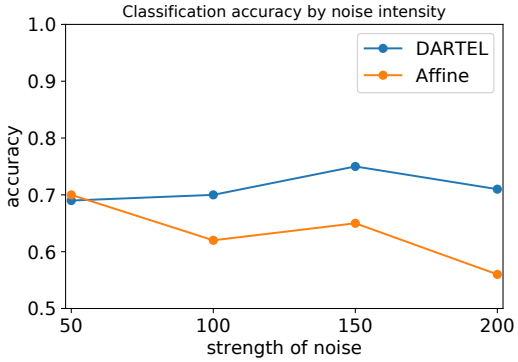


Fig. 6. Classification accuracy

V. Discussion

The purpose of this study was to identify an effective preprocessing method for AD classification of brain MRI images with limited data(MRI of patients with AD). It is known that patients with AD are characterized by GM atrophy[30]. The result shows that the classification accuracy was improved using the DARTEL preprocessing method. This indicates that the GM volume map is likely to be an important feature for AD/NC classification. Owing to GM atrophy, the GM volume of AD is smaller than that in a normal state. Therefore, the volume is an important piece of information in the classification between AD and NC. Thus, with the use of the volume map of GM overfitting was reduced. The reason for this is that the image after the

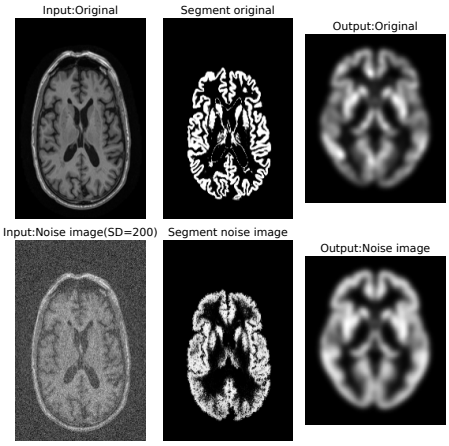


Fig. 7. Images of DARTEL process for each input image

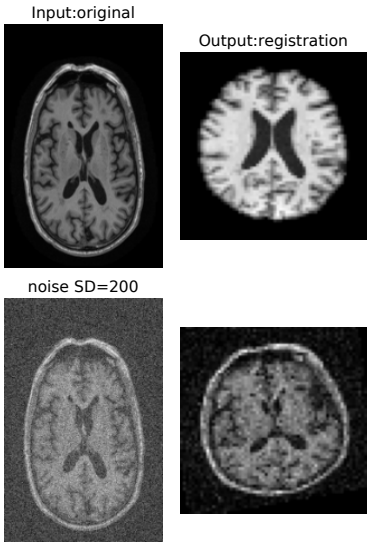


Fig. 8. Results of the Affine transform for each noisy image.

DARTEL algorithm is applied expresses only the volume of the GM and its position information; accordingly, the number of features that can be learned is reduced compared with the image of the whole brain. Given that 3D images of the entire brain contain a large amount of information making the model prone to overfitting, we can expect that learning can be performed more efficiently with data preprocessed using DARTEL. Fig.5 confirms that the nonlinear registration image can detect specific AD/NC features in the training and test data as well.

In particular, in the case of whole brain, overfitting occurred in almost all experiments, and no improvement in the classification accuracy was observed along with the increase in the amount of data. The amount of data used in this experiment were not sufficient for learning generalized features. If the available data increase in the future, the relationship between the classification accuracy, the number of features, and the amount of data required for learning can be clarified.

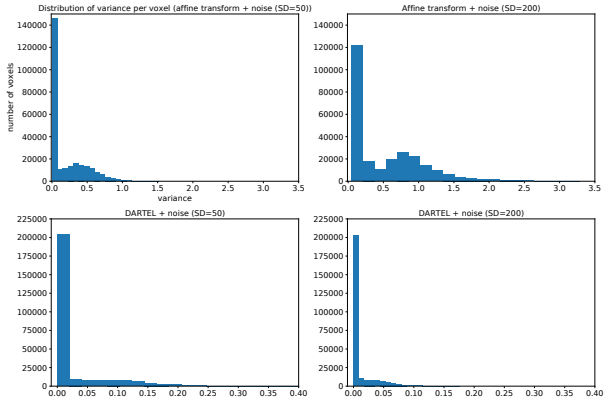


Fig. 9. Accumulate the variance value of each voxel

Given that noisy images are used for analyses in clinical settings, the effects of noise on the accuracy of the model is important for practical applications. When noise was added, the classification accuracy was reduced in affine preprocessing cases. Conversely, in DARTEL preprocessing, it was confirmed that the specific classification accuracy was maintained. To clarify the reason, outputs from each step of the DARTEL process is shown in Fig.7. According to the segmented images in Fig.7, the reason for which the boundaries of the brain region cannot be extracted accurately is attributed to noise. It is considered to be unclear to segment because the boundary between white matter and GM became ambiguous due to noise. In the case of affine processing, overfitting occurs in the same way when noise is added. As the noise intensity increases, it can be confirmed that the contours of the transformed image become dulled and noise remains. We can also confirm in Fig.8 that the standard brain affine transformation was not possible, given that the noise intensity was 200. Other cases of affine conversion failures could be identified for other images. Learning the presence or absence of the skull and noise in the image may have resulted in overfitting.

Given that the model and parameters of DL were fixed in all experiments, we evaluated the intensity of each noise level and the quality of the images during preprocessing. As shown in Fig.9, the voxel-by-voxel variance between the brain images becomes smaller when the noise intensity increased, and DARTEL ' s preprocessing was applied. This indicates that the difference in image features becomes smaller when the noise intensity increases. In the case of affine transformation, registration may fail, as shown in Fig. 8. Thus, it can be considered that the variance for each voxel increased.

VI. Conclusion

We conducted a study to identify effective preprocessing approaches for AD discrimination. Our results showed that using GM volume map and the DARTEL method in preprocessing reduced overfitting and improved the classification accuracy by up to 18%. Therefore, nonlinear registration of

brain MRI data using DARTEL was found to be a more effective preprocessing method in our experiment.

With respect to noise, also, it was found that the DARTEL approach was more robust to additive Gaussian noise at different intensities with a stable accuracy range or 69%-71%. Conversely, in the approach with affine preprocessing, the accuracy range was adversely affected leading to low ranges around 56%.

Finally, there are various limitations associated with this study that should be acknowledged. First, although multiple hyperparameteroptions exist for training and fine-tuning, we cannot rule out that other hyperparameter selections may lead to better accuracy. As ResNet used in this research is publicly available, other researchers may attempt to improve upon this. Second, we have not verified the robustness between modalities of the MRI equipment. We created the noise type with a Gaussian filter, but we have not verified this study with images containing actual artifacts. Third, The amount of data used in this experiment was not sufficient to confirm the generalization of the models. In the future, as more data becomes available, we will be able to clarify the relationship between classification accuracy, number of features, and the amount of data required for generalization.

REFERENCES

[1] World Health Organization(2012). Dementia: a public health priority. World Health Organization.

[2] McKhann G, Drachman D, Folstein M, Katzman R, Price D, Stadlan EM. Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. Neurology. 1984 Jul;34(7):939-44.

[3] Magnin B, Mesrob L, Kinkingnéhun S, Pélégrini-Issac M, Colliot O, Sarazin M, Dubois B, Lehéricy S, Benali H. Support vector machine-based classification of Alzheimer's disease from whole-brain anatomical MRI. Neuroradiology. 2009 Feb;51(2):73-83.

[4] A.V. Lebedev, E. Westman, G.J.P. Van Westen, M.G. Kramberger, A. Lundervold, D. Aarsland, H. Soininen, I. Kłoszewska, P. Mecocci, M. Tsolaki, B. Vellas, S. Lovestone, A. Simmons, Random Forest ensembles for detection and prediction of Alzheimer's disease with a good between-cohort robustness, NeuroImage: Clinical, Volume 6, 2014, Pages 115-125.

[5] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. 2012. ImageNet classification with deep convolutional neural networks. In Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1 (NIPS'12). Curran Associates Inc., Red Hook, NY, USA, 1097–1105.

[6] Revett K. (2011) An Introduction to Magnetic Resonance Imaging: From Image Acquisition to Clinical Diagnosis. In: Kwaśnicka H., Jain L.C. (eds) Innovations in Intelligent Image Analysis. Studies in Computational Intelligence, vol 339. Springer, Berlin, Heidelberg.

[7] Jose Bernal, Kaisar Kushibar, Daniel S. Asfaw, Sergi Valverde, Arnau Oliver, Robert Martí, Xavier Lladó, Deep convolutional neural networks for brain image analysis on magnetic resonance imaging: a review, Artificial Intelligence in Medicine,Volume 95,2019,Pages 64-81,ISSN 0933-3657,

[8] Razzak M.I., Naz S., Zaib A. (2018) Deep Learning for Medical Image Processing: Overview, Challenges and the Future. In: Dey N., Ashour A., Borra S. (eds) Classification in BioApps. Lecture Notes in Computational Vision and Biomechanics,

[9] Wen D, Wei Z, Zhou Y, Li G, Zhang X, Han W. Deep Learning Methods to Process fMRI Data and Their Application in the Diagnosis of Cognitive Impairment: A Brief Overview and Our Opinion. Front Neuroinform. 2018 Apr 26;12:23.

[10] LeCun, Y., Bengio, Y., Hinton, G. Deep learning. Nature 521, 436–444 (2015).

[11] Najafabadi, M.M., Villanustre, F., Khoshgoftaar, T.M. et al. Deep learning applications and challenges in big data analytics.

[12] Wei Zhao, Research on the Deep Learning of the Small Sample Data based on Transfer Learning,AIP Conference Proceedings 1864, 020018 (2017)

[13] Shen D, Wu G, Suk HI. Deep Learning in Medical Image Analysis. Annu Rev Biomed Eng. 2017 Jun 21;19:221-248. Epub 2017 Mar 9. PMID: 28301734; PMCID: PMC5479722.

[14] Hinton, G. E. and Salakhutdinov, R. R. Reducing the Dimensionality of Data with Neural Networks American Association for the Advancement of Science. 2006.

[15] Liu M, Cheng D, Wang K, Wang Y; Alzheimer ' s Disease Neuroimaging Initiative. Multi-Modality Cascaded Convolutional Neural Networks for Alzheimer's Disease Diagnosis. Neuroinformatics. 2018 Oct;16(3-4):295-308.

[16] Manhua Liu, Fan Li, Hao Yan, Kundong Wang, Yixin Ma, Li Shen, Mingqing Xu, A multi-model deep convolutional neural network for automatic hippocampus segmentation and classification in Alzheimer ' s disease, NeuroImage, Volume 208, 2020, 116459, ISSN 1053-8119,

[17] Liu M, Zhang J, Adeli E, Shen D. Landmark-based deep multi-instance learning for brain disease diagnosis. Med Image Anal. 2018 Jan;43:157-168.

[18] Ashburner J. A fast diffeomorphic image registration algorithm. Neuroimage. 2007 Oct 15;38(1):95-113.

[19] Basaia S, Agosta F, Wagner L, Canu E, Magnani G, Santangelo R, Filippi M; Alzheimer's Disease Neuroimaging Initiative. Automated classification of Alzheimer's disease and mild cognitive impairment using a single MRI and deep neural networks. Neuroimage Clin. 2019;21:101645.

[20] Tustison NJ, Avants BB, Cook PA, Zheng Y, Egan A, Yushkevich PA, Gee JC. N4ITK: improved N3 bias correction. IEEE Trans Med Imaging. 2010 Jun;29(6):1310-20.

[21] Tustison NJ, Cook PA, Klein A, Song G, Das SR, Duda JT, Kandel BM, van Strien N, Stone JR, Gee JC, Avants BB. Large-scale evaluation of ANTs and FreeSurfer cortical thickness measurements. Neuroimage. 2014 Oct 1;99:166-79.

[22] Avants BB, Tustison NJ, Song G, Cook PA, Klein A, Gee JC. A reproducible evaluation of ANTs similarity metric performance in brain image registration. Neuroimage. 2011 Feb 1;54(3):2033-44.

[23] Wang Z, Das SR, Xie SX, Arnold SE, Detre JA, Wolk DA; Alzheimer's Disease Neuroimaging Initiative. Arterial spin labeled MRI in prodromal Alzheimer's disease: A multi-site study. Neuroimage Clin. 2013 Apr 30;2:630-6.

[24] Martino ME, de Villoria JG, Lacalle-Aurioles M, Olazarán J, Cruz I, Navarro E, García-Vázquez V, Carreras JL, Desco M. Comparison of different methods of spatial normalization of FDG-PET brain images in the voxel-wise analysis of MCI patients and controls. Ann Nucl Med. 2013 Aug;27(7):600-9.

[25] Fernando Pérez-García, Rachel Sparks, Sébastien Ourselin,TorchIO: a Python library for efficient loading, preprocessing, augmentation and patch-based sampling of medical images in deep learning,Computer Methods and Programs in Biomedicine,2021,106236,ISSN 0169-2607.

[26] Ashburner J, Barnes G, Chen C, et al. SPM8 Manual. Statistical parametric mapping.

[27] K. He, X. Zhang, S. Ren and J. Sun, Deep Residual Learning for Image Recognition, 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, 2016, pp. 770-778.

[28] B. Zhou, A. Lapedriza, A. Khosla, A. Oliva and A. Torralba, "Places: A 10 Million Image Database for Scene Recognition," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 40, no. 6, pp. 1452-1464, 1 June 2018.

[29] Adam Paszke and Sam Gross and Francisco Massa and Adam Lerer and James Bradbury and Gregory Chanan and Trevor Killeen and Zeming Lin and Natalia Gimelshein and Luca Antiga and Alban Desmaison and Andreas Köpf and Edward Yang and Zach DeVito and Martin Raison and Alykhan Tejani and Sasank Chilamkurthy and Benoit Steiner and Lu Fang and Junjie Bai and Soumith Chintala,PyTorch: An Imperative Style, High-Performance Deep Learning Library,CoRR,abs/1912.01703,2019,arXiv,1912.01703

[30] Serra L, Cercignani M, Lenzi D, Perri R, Fadda L, Caltagirone C, Macaluso E, Bozzali M. Grey and white matter changes at different stages of Alzheimer's disease. J Alzheimers Dis. 2010;19(1):147-59.