3D Transformer based on deformable patch location for differential diagnosis between Alzheimer's disease and Frontotemporal dementia

Huy-Dung Nguyen, Michaël Clément, Boris Mansencal, Pierrick Coupé

3D Transformer based on deformable patch location for differential diagnosis between Alzheimer's disease and Frontotemporal dementia

Huy-Dung Nguyen, Michaël Clément, Boris Mansencal, and Pierrick Coupé

Univ. Bordeaux, CNRS, Bordeaux INP, LaBRI, UMR 5800, 33400 Talence, France
huy-dung.nguyen@u-bordeaux.com

Abstract. Alzheimer's disease and Frontotemporal dementia are common types of neurodegenerative disorders that present overlapping clinical symptoms, making their differential diagnosis very challenging. Numerous efforts have been done for the diagnosis of each disease but the problem of multi-class differential diagnosis has not been actively explored. In recent years, transformer-based models have demonstrated remarkable success in various computer vision tasks. However, their use in disease diagnostic is uncommon due to the limited amount of 3D medical data given the large size of such models. In this paper, we present a novel 3D transformer-based architecture using a deformable patch location module to improve the differential diagnosis of Alzheimer's disease and Frontotemporal dementia. Moreover, to overcome the problem of data scarcity, we propose an efficient combination of various data augmentation techniques, adapted for training transformer-based models on 3D structural magnetic resonance imaging data. Finally, we propose to combine our transformer-based model with a traditional machine learning model using brain structure volumes to better exploit the available data. Our experiments demonstrate the effectiveness of the proposed approach, showing competitive results compared to state-of-the-art methods. Moreover, the deformable patch locations can be visualized, revealing the most relevant brain regions used to establish the diagnosis of each disease.

Keywords: Deformable Patch Location · 3D Transformer · Differential diagnosis · Alzheimer's Disease · Frontotemporal Dementia

1 Introduction

Alzheimer's disease (AD) and Frontotemporal dementia (FTD) are the two most prevalent types of neurodegenerative disorders. They are the main cause of cognitive impairment and dementia [2]. Therefore, their differential diagnosis is crucial for determining appropriate interventions and treatment plans. However, these diseases share several overlapping symptoms such as memory loss and behavior changes, making their differential diagnosis challenging even when they have different clinical diagnostic criteria [28]. Indeed, several studies have demonstrated

the limitations of cognitive tests in distinguishing patients with FTD from those with AD [13,38]. Furthermore, cognitively normal (CN) people may also exhibit some changes in behavior and memory as a result of the natural aging process. Consequently, an automatic tool for multi-class diagnosis (*i.e.*, AD *vs.* FTD *vs.* CN) is highly valuable in a real clinical context.

Several works have reported that AD and FTD are associated with brain structure atrophy [27,29], which can be visualized using structural magnetic resonance imaging (sMRI) [9,24]. This modality has been used to extract structure volumes [9] or used as input of convolutional neural networks (CNN) [11,26] for differential diagnosis. In recent years, transformer-based models appear to be a promising alternative to CNN-based models in computer vision tasks. However, their application in disease diagnostic (*e.g.*, differential diagnosis) is still limited due to their computational demands and data requirements.

To alleviate computation problems, classification can be considered as a 2D problem. Lyu *et al.* and Jang *et al.* used 2D features extracted from MRI, both using a vision transformer (ViT) [8] for AD classification [15,19]. However, the lack of spatial information in such 2D approaches may not be optimal. Regarding 3D methods, for AD diagnosis, Li *et al.* downsampled the input image before feeding it to their transformer [16], Zhang *et al.* reduced the feature map dimension by setting a big patch size for embedding [40]. However, these strategies may reduce the details of local regions. For natural image classification, other techniques to reduce computation are local attention [17] and deformable attention [37]. The idea of both methods is to reduce the size of the attention matrix by decreasing the number of query, key, and value points. In the case of deformable attention mechanism, key points can be visualized for better interpretation.

Transformer-based models are known to require a large amount of data to achieve high performance [8]. In medical imaging, the limited number of labeled sMRI makes it difficult to train these models effectively. In this situation, data augmentation plays an important role in the model generalization. While data augmentation has been shown to be effective for transformer in natural image classification [33], its effectiveness in medical imaging has not been investigated.

In this paper, we first propose a 3D transformer-based architecture using a deformable patch location (DPL) module for the problem of multi-class differential diagnosis (*i.e.*, AD *vs.* FTD *vs.* CN). In the backbone, we employ local attention [17] instead of global one to reduce the computation. Our DPL module is inspired from the deformable attention [37], however, unlike the original model, deformable points in DPL are determined for each sub-volume of the image rather than being shared across the entire image. Second, to alleviate data scarcity, we propose an efficient combination of various data augmentation techniques. The exploration of data augmentation for 3D transformer-based classification using sMRI has remained relatively unexplored until now, and our strategy aims to fill this gap. Moreover, our data augmentation allows a multi-scale prediction, improving our model performance. Finally, we propose to combine our transformer-based method with a support vector machine (SVM) using structure volumes to even better exploit the limited training data. As a result,

Table 1. Number of participants.

	Dataset	CN	AD	FTD
In-domain	ADNI2	180	149	
	NIFD	136		150
Out-of-domain	NACC	2182	485	37

our framework shows competitive results compared to state-of-the-art methods for multi-class differential diagnosis.

2 Materials and method

2.1 Datasets and preprocessing

Table 1 describes the number of participants used in this study. The data consisted of 3319 subjects from multiple studies: the Alzheimer's Disease Neuroimaging Initiative (ADNI) [14], the Frontotemporal lobar Degeneration Neuroimaging Initiative (NIFD) [1] and the National Alzheimer's Coordinating Center (NACC) [3]. We only used T1-weighted MRIs at the baseline acquired with 3 Tesla machines. For the NIFD dataset, we only selected the behavior variant, progressive non-fluent aphasia, and semantic variant sub-types. The ADNI2 and NIFD datasets constituted our in-domain dataset while the NACC constituted our out-of-domain dataset. The in-domain dataset was used to perform a 10-fold cross-validation. The out-of-domain was used as an external dataset for evaluating the generalization capacity of the trained models.

The T1w MRI was preprocessed in 5 steps, which included (1) denoising [22], (2) inhomogeneity correction [35], (3) affine registration into MNI152 space ($181 \times 217 \times 181$ voxels at $1mm \times 1mm \times 1mm$) [1], (4) intensity standardization [21] and (5) intracranial cavity (ICC) extraction [23]. After that, we cropped at the image center a volume of size $144 \times 168 \times 144$ voxels to remove empty spaces. The brain structure volumes (*i.e.*, normalized volume in % of ICC) were measured using a brain segmentation predicted by AssemblyNet [7]. These volume features were used as input for our SVM.

2.2 Method

Overview Figure 1 shows an overview of our proposed model. Our model is composed of four parts: a volume embedding (VE), N blocks of a patch multi-head self-attention (P-MSA) followed by a shift patch multi-head self-attention (SP-MSA - the main building block of Swin [17]), a deformable patch location multi-head self-attention module (DPL-MSA) and a local patch averaging layer followed by a multi-layer perceptron (MLP). Intuitively, the VE module encodes

[1] Available at `https://ida.loni.usc.edu/`

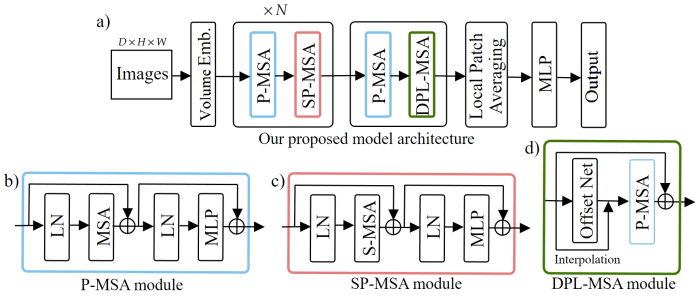

Fig. 1. The architecture of our proposed model

an MRI to a 3D volume of tokens. The N blocks of P-MSA and SP-MSA process these tokens as a attention-based feature extractor. Then, the DPL-MSA block predicts deformable patch locations and performs attention on them. While standard transformer-based approaches perform a global average of all the patches together [17], in our method we perform local average of patches in the same area (*i.e.*, sub-volume). To this end, we divide the brain feature map into 27 sub-volumes ($3 \times 3 \times 3$ areas evenly distributed along 3 axis). This is because different brain locations may be affected by a disease differently, thus should be weighted differently in the model decision. Finally, we use an MLP for classification.

Volume embedding We start with a preprocessed image of size $144 \times 168 \times 144$ (at $1mm^3$) (see Figure 1a). The VE module uses a CNN (similar to [34]) to embed the input into token vectors (with an embedding dimension of 96). This results in a 96-channel 3D feature map of size $36 \times 42 \times 36$.

Feature extractor The obtained 3D feature map is fed into three (P-MSA + SP-MSA) blocks. The details of each block are presented in 1b,c. Our implementation of these blocks is based on [17]. The local attention size is set to $6 \times 7 \times 6$. By using attention mechanism, the size of feature maps remain unchanged.

DPL block Taking the output of the feature extractor, we first update the feature map with a P-MSA module (see 1a). We then split it into $6 \times 6 \times 6$ reference patches of size $(p_x, p_y, p_z) = (6 \times 7 \times 6)$. Their centers are denoted as: $(x^i_{ct}, y^i_{ct}, z^i_{ct})$. The coordinates of these points are normalized in $[0, 1]$. Each reference patch is used as input of an offset network (see Figure 1d) to predict the offset logits $(\delta^i_x, \delta^i_y, \delta^i_z)$. The deformable patch center $(x^i_{Dct}, y^i_{Dct}, z^i_{Dct})$ is then calculated by: $x^i_{Dct} = x^i_{ct} + \tanh \delta^i_x / (2 \times p_x)$ (idem for y^i_{Dct} and z^i_{Dct}). Based on the deformable patch centers, we interpolate our feature map to obtain the corresponding deformable patches of size $p_x \times p_y \times p_z$. After that, we apply a P-MSA module to these deformable patches. Finally, a shortcut from reference patches is added to the output of the P-MSA module (see Figure 1d).

Local patch averaging We consider the obtained 96-channel 3D brain feature map (of size $36 \times 42 \times 36$) as a $3 \times 3 \times 3$ areas of size $12 \times 14 \times 12$ voxels, which are evenly distributed along 3 dimensions. We first average each deformable patch to a 96-channel mean token (of size $1 \times 1 \times 1$). Then, all the mean tokens located

in a same area are averaged. Finally, we concatenate the obtained tokens and feed it into a MLP for classification.

2.3 Data augmentation

In this part, we describe our combination of data augmentation techniques. We start with mixup, which has been known to reduce overfitting in various applications [39]. Following this, we apply a series of affine transformations, including rotation and scaling, commonly used in medical imaging applications [12, 25]. To further enhance our augmentation process, we randomly crop images at an arbitrary position (with a probability p) and resize them to match the input resolution. This technique, similar to "Random resized crop" in 2D imaging [32], mitigates overfitting and allows evaluation at both global and local views of an image. During inference, we ensemble predictions from multiple views to improve the model performance. In Section 3.1, we demonstrate the importance of each of these techniques on our framework accuracy.

2.4 Validation framework and ensembling

When evaluating our models, we made two predictions for each image: one for the whole image and one for a crop of that image. The cropping position was selected from nine cropping positions: a center crop and eight crops at corners. For each trained model, the crop position that produced the lowest loss on the validation set was selected. Finally, we averaged the two obtained results.

To further exploit the limited amount of training data, we combined (*i.e.*, average) the transformer prediction with SVM prediction based on brain structures volumes (see Section 2.5).

2.5 Implementation details

The offset network consisted of 3 layers: 3D convolution with 24 channels, kernel = (6, 7, 6), GELU activation [10] and another 3D convolution with 3 channels, kernel = 1. For data augmentation, rotation range was $\pm 0.05 rad$ and scale range was $[0.9, 1.1]$, the crop size was (132, 154, 132), the probability $p = 0.7$. The model was trained for 300 epochs using AdamW optimizer [18], cosine learning rate scheduler (start at 3e-4 and end at 5e-5). To train the SVM models, we used a grid search of three kernels (linear, polynomial, and gaussian) and 50 values of the hyper-parameter C in $[10^{-2}, 10^{2}]$ on the validation for tuning hyper-parameters. The SVM models used the same train/validation/test (70%/20%/10%) splits of in-domain data during cross-validation than our deep learning models.

3 Experimental results

In this study, we first performed a 10-fold cross-validation on in-domain dataset. This resulted in 20 models (10 Transformers and 10 SVM models). We concatenated the prediction of 10 test folds to compute the global in-domain performance. For out-of-domain evaluation, we averaged all 10 predictions to estimate

Table 2. Ablation study of the model performance. Results obtained using the data augmentation described in 2.3. Gray text, symbols: that option is the same as in the previous experiment. Red, Blue: best, second result.

No.	2D/3D	Local patch averaging	Nonlinear VE	DPL module	Multi-scale prediction	Combination with SVM	In-domain			Out-of-domain		
							ACC	BACC	AUC	ACC	BACC	AUC
1	2D	✗	✗	✗	✗	✗	68.8	64.1	81.1	77.4	63.3	78.4
2	3D	✗	✗	✗	✗	✗	78.4	74.7	90.1	81.5	75.2	87.8
3	3D	✓	✗	✗	✗	✗	82.9	79.5	92.7	85.4	78.2	89.3
4	3D	✓	✓	✗	✗	✗	83.6	80.3	92.5	86.6	79.7	89.9
5	3D	✓	✓	✓	✗	✗	83.4	80.7	93.4	87.1	80.1	90.5
6	3D	✓	✓	✓	✓	✗	85.2	82.5	94.1	87.7	80.7	91.0
7	3D	✓	✓	✓	✓	✓	86.2	83.4	94.5	89.3	82.8	91.6

the model performance. We used 3 metrics to assess the model performance: accuracy (ACC), balanced accuracy (BACC) and area under curve (AUC).

3.1 Ablation study

Performance study In this part, we studied the impact of each contribution on our model performance. These factors could be organized into 4 groups: Input type (2D/3D), architecture (local patch averaging, non linear volume embedding), validation framework (multi-scale prediction) and ensemble (combination with SVM). The used data augmentation schema was described in 2.3. Table 2 showed the results of the comparison.

First, we implemented a basic 2D transformer-based architecture (exp. 1) and its 3D version (exp. 2) to see if the spatial information from 3D input is valuable. We observed that the 3D version was better than the 2D version in all metrics. Second, using local patch averaging (exp. 3) improved our model performance, confirming the effectiveness of assigning different weights to different brain areas. Third, the nonlinear volume embedding (exp. 4) could also improve the performance of transformer, which was inline with [34]. Then, the DPL module demonstrated an improvement in performance across almost all metrics (exp. 5). Finally, the multi-scale prediction (exp. 6) and ensembling (exp. 7) increased even more our model performance in both in-domain and out-of-domain data.

Data augmentation study Table 3 shows the contribution of each data augmentation technique to our model performance. The ensembling with SVM was removed for analysis and the multi-scale evaluation was applied only when multi-crop was used. First, without any data augmentation, the obtained result (exp. 1) was lower than in other experiments. Second, combining different augmentations

Table 3. Ablation study of the data augmentation. Gray symbols: that option is the same as in the previous experiment. Red, Blue: best, second result.

No.	Mixup	Rand. affine	Multi crops	In-domain ACC	In-domain BACC	In-domain AUC	Out-of-domain ACC	Out-of-domain BACC	Out-of-domain AUC
1	✗	✗	✗	74.6	69.0	87.8	84.3	73.3	87.3
2	✓	✗	✗	77.6	72.0	88.4	84.8	76.0	87.4
3	✓	✓	✗	82.1	78.9	91.5	86.2	78.6	90.0
4	✓	✓	✓	85.2	82.5	94.1	87.7	80.7	91.0

(exp. 2, 3, 4) progressively improved the model's generalization. This showed the effectiveness of our data augmentation for medical imaging applications.

3.2 Comparison with state-of-the-art methods

In this section, we compare our results with current state-of-the-art methods for the multi-class diagnosis AD *vs.* FTD *vs.* CN. Hu *et al.* proposed an CNN-based architecture inspired by Resnet which processes the whole 3D MRI for classification [11]. Ma *et al.* used a MLP with cortical thickness (Cth) and brain structure volumes extracted from a 3D MRI [20]. They also used a generative adversarial network to generate new data to prevent over-fitting. More recently, Nguyen *et al.* used a large number of CNN to grade brain regions. The grading values were then averaged for each brain structure and used as input of a MLP for classification [26]. For a fair comparison, we reimplemented these methods and trained them under the same training setting as our method and on the same data. Table 4 shows the results of the comparison.

Overall, our method presented most of the time the best performance in all metrics (*i.e.*, ACC, BACC and AUC) and for both in-domain and out-of-domain data. Moreover, our method was the only method based on the transformer mechanism. This suggested that transformer-based methods can obtain competitive results compared to CNN-based networks even with a limited amount of data.

3.3 Visualization of deformable patch location

Figure 2 shows the centers of deformable patch locations for patients with AD and FTD. For each patient group, the patch center positions are calculated as the averaged center locations from our ten models. To enrich visual comprehension, we utilized GradCAM to attribute an importance score within the range of $[0, 1]$ to each patch. Patches obtaining an importance score above 0.3 are displayed. Furthermore, a higher importance score is visually represented by a larger circle, and the warmth of the circle's color increases with the score.

Table 4. Comparison with state-of-the-art methods. Red, Blue: best, second result.

Method	In-domain			Out-of-domain		
	ACC	BACC	AUC	ACC	BACC	AUC
CNN on intensities [11]	76.3	72.5	90.0	85.2	68.8	86.5
MLP on Cth and volumes [20]	77.1	75.9	86.4	69.1	74.6	87.5
3D Grading [26]	86.0	84.7	93.8	87.1	81.6	91.6
Our method	86.2	83.4	94.5	89.3	82.8	91.6

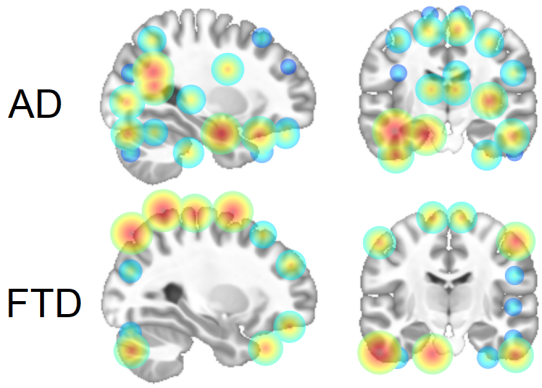

Fig. 2. Visualization of deformable patch locations. The importance of each patch was estimated with GradCAM. Warmer color, larger radius mean higher importance score.

The obtained results were coherent with the current knowledge about these diseases. Indeed, for AD patients, the structures that obtained higher score were the left hippocampus [30], bilateral entorhinal cortex, bilateral ventricle [6] and parietal lobe [31]. In FTD patients, the frontal pole [4], superior frontal gyrus [5] and left temporal cortex [36] were highlighted.

4 Conclusion

Our study presents a novel 3D transformer model, which incorporates a deformable patch location module for the differential diagnosis between cognitively normal subjects, patients with Alzheimer's disease and patients with Frontotemporal dementia. The proposed module enhances the model's accuracy and provides useful visualizations that reveal insights into each disease. To address the problem of limited training data, we designed a combination common data augmentations for training transformer models using 3D MRI. Furthermore, we proposed to combine both our deep learning model and an SVM using brain structure volumes to even better exploit the limited data. As a result, our framework showed competitive performance compared to state-of-the-art methods.

References

1. Avants, B.B., et al.: A reproducible evaluation of ANTs similarity metric performance in brain image registration. NeuroImage **54**, 2033–2044 (2011)
2. Bang, J., et al.: Frontotemporal dementia. The Lancet **386**, 1672–1682 (2015)
3. Beekly, D.L., et al.: The National Alzheimer's Coordinating Center (NACC) Database: The Uniform Data Set. Alzheimer Disease & Associated Disorders **21**, 249–258 (2007)
4. Boeve, B.F., et al.: Advances and controversies in frontotemporal dementia: diagnosis, biomarkers, and therapeutic considerations. The Lancet Neurology **21**, 258–272 (2022)
5. Brambati, S.M., et al.: A tensor based morphometry study of longitudinal gray matter contraction in FTD. Neuroimage **35**(3), 998–1003 (2007)
6. Coupé, P., et al.: Lifespan Changes of the Human Brain In Alzheimer's Disease. Scientific reports **9**, 3998 (2019)
7. Coupé, P., et al.: AssemblyNet: A large ensemble of CNNs for 3D whole brain MRI segmentation. NeuroImage **219**, 117026 (2020)
8. Dosovitskiy, A., et al.: An image is worth 16x16 words: Transformers for image recognition at scale. arXiv:2010.11929 (2020)
9. Du, A.T., et al.: Different regional patterns of cortical thinning in Alzheimer's disease and frontotemporal dementia. Brain **130**, 1159–1166 (2006)
10. Hendrycks, D., Gimpel, K.: Gaussian error linear units (gelus). arXiv preprint arXiv:1606.08415 (2016)
11. Hu, J., et al.: Deep learning-based classification and voxel-based visualization of Frontotemporal dementia and Alzheimer's disease. Frontiers in Neuroscience **14**, 626154 (2021)
12. Hussain, Z., Gimenez, F., Yi, D., Rubin, D.: Differential data augmentation techniques for medical imaging classification tasks. In: AMIA annual symposium proceedings. vol. 2017, p. 979 (2017)
13. Hutchinson, A.D., et al.: Neuropsychological deficits in frontotemporal dementia and Alzheimer's disease: a meta-analytic review. Journal of Neurology, Neurosurgery and Psychiatry **78**, 917–928 (2007)
14. Jack, C.R., et al.: The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods. Journal of Magnetic Resonance Imaging **27**, 685–691 (2008)
15. Jang, J., Hwang, D.: M3t: three-dimensional Medical image classifier using Multi-plane and Multi-slice Transformer. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 20718–20729 (2022)
16. Li, C., et al.: Trans-ResNet: Integrating Transformers and CNNs for Alzheimer's disease classification. In: 2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI). pp. 1–5 (2022)
17. Liu, Z., et al.: Swin transformer: Hierarchical vision transformer using shifted windows. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 10012–10022 (2021)
18. Loshchilov, I., Hutter, F.: Decoupled weight decay regularization. arXiv preprint arXiv:1711.05101 (2017)
19. Lyu, Y., et al.: Classification of Alzheimer's disease via Vision Transformer. In: Proceedings of the 15th International Conference on PErvasive Technologies Related to Assistive Environments. pp. 463–468 (2022)

20. Ma, D., et al.: Differential diagnosis of Frontotemporal dementia, Alzheimer's disease, and normal aging using a multi-scale multi-type feature Generative adversarial deep neural network on structural magnetic resonance images. Frontiers in Neuroscience **14**, 853 (2020)
21. Manjón, J.V., et al.: Robust MRI brain tissue parameter estimation by multistage outlier rejection. Magnetic Resonance in Medicine **59**, 866–873 (2008)
22. Manjón, J.V., et al.: Adaptive non-local means denoising of MR images with spatially varying noise levels: Spatially Adaptive Nonlocal Denoising. Journal of Magnetic Resonance Imaging **31**, 192–203 (2010)
23. Manjón, J.V., et al.: Nonlocal Intracranial Cavity Extraction. International Journal of Biomedical Imaging **2014**, 1–11 (2014)
24. Möller, C., et al.: Alzheimer Disease and Behavioral Variant Frontotemporal Dementia: Automatic Classification Based on Cortical Atrophy for Single-Subject Diagnosis. Radiology **279**, 838–848 (2016)
25. Nalepa, J., Marcinkiewicz, M., Kawulok, M.: Data augmentation for brain-tumor segmentation: a review. Frontiers in computational neuroscience **13**, 83 (2019)
26. Nguyen, H., et al.: Interpretable Differential Diagnosis for Alzheimer's disease and Frontotemporal dementia. In: Medical Image Computing and Computer Assisted Intervention. pp. 61–69 (2022)
27. Pini, L., et al.: Brain atrophy in Alzheimer's disease and aging. Ageing research reviews **30**, 25–48 (2016)
28. Rascovsky, K., et al.: Sensitivity of revised diagnostic criteria for the behavioural variant of frontotemporal dementia. Brain **134**, 2456–2477 (2011)
29. Rosen, H.J., et al.: Patterns of brain atrophy in frontotemporal dementia and semantic dementia. Neurology **58**(2), 198–208 (2002)
30. Schuff, N., et al.: MRI of hippocampal volume loss in early Alzheimer's disease in relation to ApoE genotype and biomarkers. Brain **132**, 1067–1077 (2009)
31. Silhan, D., et al.: The parietal atrophy score on brain magnetic resonance imaging is a reliable visual scale. Current Alzheimer Research **17**(6), 534–539 (2020)
32. Touvron, H., Vedaldi, A., Douze, M., Jégou, H.: Fixing the train-test resolution discrepancy. Advances in neural information processing systems **32** (2019)
33. Touvron, H., et al.: Training data-efficient image transformers & distillation through attention. In: International Conference on Machine Learning. pp. 10347–10357 (2021)
34. Touvron, H., et al.: Three things everyone should know about vision transformers. In: Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XXIV. pp. 497–515. Springer (2022)
35. Tustison, N.J., et al.: N4ITK: Improved N3 Bias Correction. IEEE Transactions on Medical Imaging **29**, 1310–1320 (2010)
36. Whitwell, J.L., et al.: Distinct anatomical subtypes of the behavioural variant of frontotemporal dementia: a cluster analysis study. Brain **132**, 2932–2946 (2009)
37. Xia, Z., et al.: Vision transformer with deformable attention. In: Conference on computer vision and pattern recognition. pp. 4794–4803 (2022)
38. Yew, B., et al.: Lost and forgotten? Orientation versus memory in Alzheimer's disease and Frontotemporal dementia. Journal of Alzheimer's disease: JAD **33**, 473–481 (2013)
39. Zhang, H., et al.: mixup: Beyond Empirical Risk Minimization. arXiv:1710.09412 (2018)
40. Zhang, S., et al.: 3D Global Fourier Network for Alzheimer's disease diagnosis using structural MRI. In: Medical Image Computing and Computer Assisted Intervention. pp. 34–43 (2022)