ADAPT: Adaptive Profiling Transformers for Efficient Alzheimer's Disease Diagnosis

Yifeng Wang[1], Ke Chen[2], and Haohan Wang[3⋆]

[1] Department of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, USA
[2] Center for Data Science, New York University, New York, USA
[3] School of Information Sciences, University of Illinois Urbana-Champaign, Champaign, USA
yifengw3@andrew.cmu.edu,kc6544@nyu.edu,haohanw@illinois.edu

Abstract. Accurate and computationally efficient diagnosis of Alzheimer's Disease (AD) from MRI remains a challenge in medical imaging. While 3D deep learning models capture volumetric biomarkers, they require substantial GPU memory. Conversely, 2D approaches are efficient but fail to capture spatial context, limiting their clinical applicability. We present **ADAPT**, a clinically-inspired diagnostic framework that achieves 3D-level accuracy with 2D computational costs through careful co-design of three innovations: (1) a *View-specific transformer architecture* with four encoder families that separately learn within-slice, within-view, and cross-view relationships; (2) an *adaptive view-profiling strategy* that dynamically allocates more slices to medically informative views based on learned attention scores, providing more context in limited burden; and (3) a *pathology-driven morphology augmentation* that simulates progressive atrophy patterns to improve model robustness. Experiments across various AD datasets demonstrate that ADAPT achieves improved performance over CNN and transformer baselines while costing **75.8%** lower GFLOPs. Beyond standard AD vs. NC classification, ADAPT also achieves strong performance on the *early-stage* task of NC vs. MCI, further supporting its relevance for real-world application. These results highlight ADAPT as a practical, interpretable, and computationally efficient alternative to previous models for cross-institutional Alzheimer's Disease diagnosis.

Keywords: Alzheimer's Disease · Brain Atrophy · Multi-view Learning · Memory-efficient Deep Learning

1 Introduction

Alzheimer's Disease (AD) is a progressive neurodegenerative disorder characterized by structural brain changes including ventricular enlargement, cortical thinning, and hippocampal atrophy. Structural magnetic resonance imaging (MRI) provides high-resolution visualization of these anatomical patterns and remains

⋆ Corresponding author.

a primary clinical tool for AD diagnosis [25]. However, manual inspection of 3D MRI volumes is labor-intensive and subject to inter-reader variability, motivating the development of automated systems to support clinical decision-making [9].

Limitations of Fully 3D Modeling. Deep learning approaches for AD diagnosis have largely relied on 3D convolutional neural networks (CNNs) to capture volumetric structure [17,4,20]. While effective, fully 3D models suffer from substantial computational and memory costs, limiting their feasibility in resource-constrained clinical environments. Reducing resolution or extracting local patches can relieve the burden, but often at the expense of losing global anatomical context, which is essential for modeling distributed atrophy patterns [11,2]. Transformer-based models offer improved modeling of long-range dependencies [19], yet directly applying vision transformers to 3D volumes dramatically increases token count and computational load. Hybrid strategies have therefore emerged to balance efficiency and context modeling. For example, MTr-Net [21] combines 2D and 2.5D streams for volumetric learning, while M3T [8] extracts slice-level features from axial, sagittal, and coronal planes and fuses them using standard transformer encoders. These multi-view approaches reduce some 3D overhead, but they typically (i) treat anatomical views uniformly, (ii) use fixed numbers of slices per plane, and (iii) lack mechanisms to adapt computation based on pathology-specific relevance.

Mismatch with Clinical Practice. In clinical workflows, clinicians do *not* treat all anatomical planes as equally informative. Ventricular enlargement is more apparent in axial and coronal views, whereas hippocampal atrophy is often assessed across multiple planes. Diagnostic reasoning therefore involves selectively focusing on anatomically informative slices and integrating cross-view evidence. Existing multi-view models do not explicitly model this variable view importance and instead allocate equal computational resources across planes, thus inefficiency.

Our Approach. To address these limitations, we propose ADAPT, an efficient 2.5D multi-view framework for AD diagnosis from structural MRI. Rather than modeling the full 3D volume directly, ADAPT factorizes a scan into axial, coronal, and sagittal slices and performs adaptive cross-view reasoning using a clinically motivated transformer design. Unlike prior 2.5D models that apply uniform processing across planes, ADAPT introduces view-aware modeling and adaptive training while maintaining global representation. This design enables ADAPT to approximate 3D contextual diagnosis with substantially reduced memory and computational cost. The framework integrates three components:

- **Pathology-driven morphology augmentation**, which simulates progressive atrophy patterns grounded in AD neuropathology to improve robustness and generalization under limited data.
- **View-specialized transformer encoders** that model slice-level, intra-/inter-view relationships through self-/fusion-attention mechanisms.
- **Adaptive view profiling**, which dynamically reallocates slice counts across anatomical planes based on learned importance scores, enabling the model to concentrate computation on diagnostically informative views.

2 Methodology

ADAPT consists of three components: (1) pathology-driven morphology augmentation, (2) a clinically motivated multi-view transformer with view-specialized encoders, and (3) adaptive view profiling that reallocates slices across planes. Given a 3D MRI volume, we factorize it into three slice sequences according to sagittal, coronal, and axial views. We sample total n slices per volume with n_t slices from view $t \in \{\text{axial}, \text{coronal}, \text{sagittal}\}$. Morphology augmentation is applied only during training (Sec. 2.1). The sampled slices are then encoded with view-specialized blocks to capture within-slice, inter/intra-view relationships (Sec. 2.2). Finally, adaptive view profiling uses view-importance scores from the final cross-view encoder to update $\{n_t\}$ for the next iteration (Sec. 2.4).

2.1 Pathology-Driven Morphology Augmentation

Alzheimer's pathology manifests in MRI as ventricular enlargement and shrinkage of cortical/hippocampal structures, producing "atrophy regions" that expand with disease progression. We propose *morphology augmentation* to simulate atrophy progression/regression by locally expanding or contracting relative regions on each 2D slice. Let f be an input slice and b_N be a structuring element controlling the degree of deformation. We define morphological erosion and dilation as

$$[f \ominus b_N](x, y) = \min_{(\alpha,\beta)\in b_N} \{f(x + \alpha, y + \beta) - b_N(\alpha, \beta)\}, \tag{1}$$

$$[f \oplus b_N](x, y) = \max_{(\alpha,\beta)\in b_N} \{f(x - \alpha, y - \beta) + b_N(\alpha, \beta)\}. \tag{2}$$

Augmentation Strategy. We apply *atrophy expansion* ($\oplus$) to AD and MCI scans and retain the AD label to reflect progression. Conversely, we apply *atrophy reduction* ($\ominus$) to NC and MCI scans and label them as NC. Augmenting MCI encourages the classifier to improve accuracy and robustness in AD diagnosis. Note, we do not claim an MCI subject becomes AD/NC; rather, we use bidirectional morphology perturbations as a regularizer to model the NC–MCI–AD continuum. This augmentation increases intra-class variability, injects meaningful morphology changes without requiring segmentation, and encourages the model to focus on atrophy-related regions.

2.2 Clinically Motivated Multi-view Transformer Encoders

Radiologists review MRI by browsing slices across sagittal/coronal/axial planes and integrating evidence across views. ADAPT follows this with a 2.5D transformer that processes slices from each view and fuses information across slices and views using four complementary encoders: **SAE** (shared self-attention), **DS-AE** (dimension-specific self-attention), **IntraCAE** (intra-view cross-attention), and **InterCAE** (inter-view cross-attention). Figure 1 illustrates this architecture.

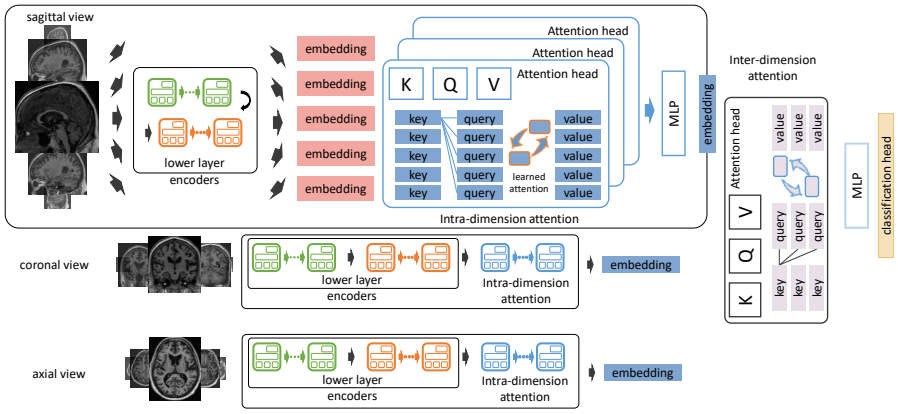

Fig. 1. The detailed architecture of ADAPT blocks, which consists of four complementary encoders for feature extraction and information gathering, across slices and views.

Given sampled slices, we tokenize each slice into patch embeddings. To inject shared global context, we add a learnable guide embedding $\mathbf{x}_{\text{guide}}$ to patch tokens and use positional embeddings $\mathbf{E}_{pos}$ to preserve ordering. The input sequence is

$$\mathbf{S}_0 = \Big[\mathbf{x}_{class}; \underbrace{\mathbf{x}_{p_1} + \mathbf{x}_{guide}; \cdots; \mathbf{x}_{p_n} + \mathbf{x}_{guide}}_{\text{sagittal}}; \underbrace{\cdots; \mathbf{x}_{p_{2n}} + \mathbf{x}_{guide}}_{\text{coronal}}; \underbrace{\cdots; \mathbf{x}_{p_{3n}} + \mathbf{x}_{guide}}_{\text{axial}}\Big] + \mathbf{E}_{pos}; \quad \mathbf{E}_{pos} \in \mathbb{R}^{(3 \cdot n \cdot N+1) \times D}. \tag{3}$$

SAE. SAE captures long-range spatial cues across slices. These encoders help maintain global dependencies inherent in the 3D MRI. The networks are implemented as Siamese networks [7] with shared weights:

$$\mathbf{S}_0^s = [\mathbf{x}_{class}^s; \mathbf{x}_{p_s}], \ \mathbf{S}_l^s = \text{SAE}(\mathbf{S}_{l\text{-}1}^s); \qquad s \in (1, n), \ l = 1...L_{\text{SAE}} \tag{4}$$

DS-AE. To enhance model's ability to capture local view-dependent cues, DS-AE applies view-specific self-attention to the slice representations, which mitigates ViT's [2] tendency to under-emphasize fine-grained anatomical features:

$$\mathbf{S}_l^{t \cdot s} = \text{DSAE}_{\text{t}}(\mathbf{S}_{l\text{-}1}^{t \cdot s}), \quad s \in (1, n_t), l = (L_{\text{SAE}} + 1)...(L_{\text{SAE}} + L_{\text{DSAE}}) \tag{5}$$

IntraCAE and InterCAE. To aggregate complementary information across slices within a single view, IntraCAE employs fusion-attention (Section 2.3), enabling slice-to-slice information sharing. While InterCAE integrates representations across three views, enabling the model to capture cross-plane structural patterns characteristic (e.g., ventricular enlargement visible across axial and coronal slices):

$$\mathbf{S}_l^{t \cdot s} = \text{IntraCAE}_{\text{t}}(\mathbf{S}_{l\text{-}1}^{t \cdot s}), \ \mathbf{S}_l^t = \text{InterCAE}_{\text{t}}(\mathbf{S}_{l\text{-}1}^t); \quad s \in (1, n_t). \tag{6}$$

Algorithm 1 ADAPT Training Strategy

Input: 3D MRI Training set T, initial slice number list ψ, model ADAPT Θ, total slice number n_{total}

Output: Updated model Θ, final list ψ

while *Training* **do**
  With T and ψ, sample 2D data δ_a, δ_c,δ_s on axial, coronal and sagittal views
  δ_a,δ_c,δ_s = SAE(δ_a,δ_c,δ_s)
  δ_a,δ_c,δ_s = DSAE$_a$(δ_a), DSAE$_c$(δ_c), DSAE$_s$(δ_s)
  δ_a,δ_c,δ_s = IntraCAE$_a$(δ_a), IntraCAE$_c$(δ_c), IntraCAE$_s$(δ_s)
  δ_a,δ_c,δ_s = InterCAE(δ_a,δ_c,δ_s)
  Calculate score using $H_{\text{dim}} = \text{softmax}\left(\frac{Q_{\text{class\_dim}} K^\top}{\sqrt{d_k}}\right) V$ for three dimensions
  Calculate cross-entropy loss and update Θ
  if p **then**
    Update ψ according to $n_t = \text{round}\left(\frac{\hat{H}_{\text{dim}}}{\sum_{r \in \psi} r}\, n\right)$.
  else
    INITIALIZE(ψ)
  end if
end while

Classification. We average across view-wise `[class]` tokens after InterCAE, and feed the result to an MLP classifier for AD vs. NC (and NC vs. MCI) prediction.

2.3 Fusion Attention Mechanism

Alzheimer's pathology often manifests as distributed structural changes that require information across slices and views. Standard self-attention mixes information only through learned Q, K, V projections from the same slice/view. We propose a lightweight *fusion attention* used inside IntraCAE and InterCAE to inject shared representations while preserving slice-/view-specific information.

Fusion Operation. For a view t with n_t slices, let $\{\mathbf{x}_{p_{t,1}}, \ldots, \mathbf{x}_{p_{t,n_t}}\}$ denote patch embeddings. We form a fused token by summing patch embeddings within a view $\mathbf{f}_t = \sum_{s=1}^{n_t} \mathbf{x}_{p_{t,s}}$, and update the slice representation $\mathbf{S}_l^{t,s} = [\mathbf{x}_{class}^{t,s},\ \mathbf{f}_t]$. This retains slice-specific structure via $\mathbf{x}_{class}^{t,s}$ while providing shared contextual evidence via $\mathbf{f}_t$. InterCAE applies the same mechanism but fuses across views (forming fused tokens at the view level) to enable cross-plane integration.

2.4 Adaptive View Profiling

Different views provide unequal diagnostic value for Alzheimer's Disease. For instance, sagittal slices better capture hippocampal atrophy, whereas axial and coronal slices emphasize ventricular enlargement. To allow the model to dynamically emphasize the most informative views during training, we introduce an attention score–based adaptive training strategy. After the final InterCAE layer,

we compute a view score H_{dim} from each view-wise `[class]` token using its attention response, which provides a principled, model-internal estimate of how much information the model extracts from each slice sequence. Given normalized attention scores $\hat{H}_{\text{dim}}$, we update the number of slices n_t assigned to each view with the total number of slices n. To avoid suboptimal view allocation, we reinitialize the slice number list ψ with probability p to encourage exploration of different slice allocations while allowing exploitation of stable attention patterns as training progresses. Algorithm 1 summarizes the full training strategy.

3 Experiments

3.1 Experimental Settings

Datasets. We train and select models on ADNI [14], and evaluate cross-institutional generalization on held-out external cohorts including AIBL [3], MIRIAD [12], and OASIS [13]. All images are single-channel T1-weighted 3D MRI. For AD vs. NC, ADNI is split at the subject level with no subject overlap (train: 878 NC / 884 AD / 1565 MCI; val: 72 NC / 81 AD; test: 266 NC / 145 AD). The external test cohorts contain AIBL: 413 scans (363 NC / 50 AD), MIRIAD: 523 scans (177 NC / 346 AD), and OASIS: 2157 scans (1692 NC / 465 AD), providing cohorts with different sample sizes and class balances for evaluating generalization beyond the development dataset. For NC vs. MCI, we use ADNI NC and MCI subjects with subject-disjoint splits (train: 878 NC / 1130 MCI; val: 72 NC / 93 MCI; test: 266 NC / 342 MCI). All data are preprocessed following protocols from medical community [20] with standard augmentations via TorchIO [15].
Training Details. All MRI volumes are preprocessed using bias-field correction, registration to MNI space, background cropping, resizing to $224 \times 224 \times 224$, and intensity normalization, following the medical imaging protocol in [20]. Each preprocessed volume is decomposed into sagittal/coronal or axial slice sequences. We initialize with 16 equidistant slices per view (48 total) sampled from the central region of the volume. We optimize using AdamW ($\text{lr} = 5 \times 10^{-5}$) with cosine scheduling [10] and cross-entropy loss [24]. The transformer uses 6 attention layers total (SAE 1, DS-AE 1, IntraCAE 2, InterCAE 2) with 4 heads and embedding dimension 256. Adaptive view profiling uses reinitialization probability $p = 0.8$. The classifier averages the three view-wise `[class]` tokens and uses an MLP head for final prediction.

3.2 Evaluation Between Baselines

In this subsection, ADAPT was compared with various baseline models for Alzheimer's diagnosis. We choose different models to show the capability of ADAPT, including: CRATE [23] designed for natural 3D data; MedicalNet [1] designed for diseases other than AD diagnosis; FCNlinksCNN [16], Knowledge4D [25], M3T [8] and nnMamba [6] designed for AD diagnosis.

Each experiment is conducted 3 times with 5-fold cross-validation. The quantitative performance is presented in Table 1. We select the best-performing

Table 1. Comparison of accuracy on multi-institutional dataset. The numerical numbers of models with morphology augmentation are bolded with better performance.

Model name	Model size (#params)	GFLOPs	Morphology Aug	ADNI		AIBL	MIRIAD	OASIS
				val acc.	test acc.	test acc.	test acc.	test acc.
MedicalNet	17,723,458	451.4	No	0.855 ± 0.015	0.851 ± 0.016	0.880 ± 0.007	0.845 ± 0.007	0.802 ± 0.002
MedicalNet [1]	**17,723,458**	**451.4**	**Yes**	**0.857**±0.013	**0.871**±0.009	**0.889**±0.011	**0.849**±0.016	**0.834**±0.013
Knowledge4D	33,162,880	1267.8	No	0.805 ± 0.005	0.864 ± 0.002	0.844 ± 0.003	0.826 ± 0.002	0.799 ± 0.006
Knowledge4D [25]	**33,162,880**	**1267.8**	**Yes**	**0.875**±0.010	0.862 ± 0.008	**0.889**±0.002	**0.864**±0.005	**0.809**±0.003
FCNlinksCNN	310,488,372	751.2	No	0.802 ± 0.008	0.807 ± 0.005	0.827 ± 0.008	0.852 ± 0.012	0.820 ± 0.007
FCNlinksCNN [16]	**310,488,372**	**751.2**	**Yes**	**0.847**±0.006	**0.836**±0.016	**0.828**±0.004	**0.864**±0.006	**0.832**±0.011
CRATE	27,567,300	489.2	No	0.804 ± 0.011	0.822 ± 0.009	0.808 ± 0.003	0.814 ± 0.002	0.794 ± 0.007
CRATE [23]	**27,567,300**	**489.2**	**Yes**	**0.856**±0.003	**0.826**±0.005	**0.818**±0.003	**0.834**±0.003	**0.842**±0.007
M3T	29,229,500	934.4	No	0.848 ± 0.002	0.858 ± 0.003	0.870 ± 0.004	0.852 ± 0.006	0.814 ± 0.001
M3T [8]	**29,229,500**	**934.4**	**Yes**	**0.894**±0.003	**0.902**±0.004	**0.884**±0.002	**0.876**±0.005	**0.844**±0.001
nnMamba	12,876,400	471.8	No	0.850 ± 0.002	0.862 ± 0.004	0.862 ± 0.008	0.858 ± 0.002	0.810 ± 0.003
nnMamba [6]	**12,876,400**	**471.8**	**Yes**	**0.900**±0.002	**0.882**±0.006	**0.876**±0.006	**0.878**±0.003	**0.826**±0.005
ADAPT	12,104,400	92.5	No	0.855 ± 0.005	0.862 ± 0.003	0.905 ± 0.004	0.856 ± 0.006	0.820 ± 0.007
ADAPT	**12,104,400**	**92.5**	**Yes**	**0.924**±0.007	**0.928**±0.002	**0.921**±0.005	**0.913**±0.004	**0.868**±0.002

Table 2. Comparison of accuracy, Brier score, specificity, and ROC-AUC for various CNN-based and transformer-based models on the ADNI NC vs MCI classification task.

Model name	ADNI							
	Val				Test			
	acc.	brier	specificity	roc	acc.	brier	specificity	roc
MedicalNet	0.752±0.003	0.252±0.004	0.617±0.003	0.719±0.002	0.728±0.005	0.248±0.002	0.612±0.003	0.688±0.001
Knowledge4D	0.775±0.006	0.248±0.003	0.703±0.003	0.758±0.005	0.748±0.004	0.245±0.005	0.728±0.002	0.742±0.004
FCNlinksCNN	0.758±0.003	0.236±0.006	0.726±0.002	0.738±0.003	0.746±0.003	0.240±0.008	0.700±0.004	0.719±0.004
CRATE	0.762±0.002	0.236±0.003	0.686±0.003	0.760±0.005	0.714±0.005	0.228±0.004	0.700±0.004	0.752±0.003
M3T	0.782±0.004	0.216±0.006	0.780±0.005	0.780±0.003	0.762±0.005	0.190±0.004	0.724±0.007	0.768±0.006
nnMamba	0.786±0.002	0.202±0.005	0.766±0.004	0.758±0.003	0.744±0.007	0.196±0.004	0.743±0.001	0.752±0.005
ADAPT	0.794±0.002	0.202±0.004	0.772±0.003	0.806±0.001	0.780±0.003	0.218±0.005	0.746±0.004	0.770±0.003

model on ADNI and evaluate it on various AD datasets to verify if the model has learned well on the knowledge that is highly transferable across different datasets collected by different institutions. We also report the total parameters and GFLOPs for measuring computational burdens with an ablation study on Morphology Augmentation. Overall, ADAPT achieves the best performance on both i.i.d scenario (ADNI) and out-of-domain testing scenarios (AIBL, MIRIAD, and OASIS), demonstrating robustness across datasets collected from different facilities. This confirms ADAPT can learn the actual AD-relevant brain features. At the same time, our model has the fewest parameters and GFLOPs, proving the efficiency of ADAPT. Beyond GFLOPs, we additionally profiled inference under matched hardware settings. ADAPT runs at approximately 52 ms per volume with 0.43 GiB peak GPU memory; compared with M3T, it is about 4.9× faster (52 vs. 254 ms), uses about 56× less peak memory (0.43 vs. 24.3 GiB), and requires about one-tenth the GFLOPs (92.5 vs. 934.4), supporting a favorable accuracy–memory–complexity trade-off. By analyzing the morphology augmentation result (bolded one), we found that it can greatly improve the diagnosis accuracy on most models, demonstrating its generalizability among different models. The NC vs MCI classification task shows a similar result in Table 2. These results highlight that ADAPT provides more reliable discrimination between NC and MCI, a clinically important early-stage diagnosis task where anatomical differences are subtle. Additionally, the best performance is achieved when ADAPT chooses 14

Table 3. Ablation study of the main components of ADAPT.

	AD vs NC (train: ADNI)					**NC vs MCI** (ADNI)	
Component	**ADNI** Val	**ADNI** Test	**AIBL** Test	**MIRIAD** Test	**OASIS** Test	**Val**	**Test**
w/o SAE	0.872±0.005	0.905±0.003	0.914±0.001	0.869±0.003	0.848±0.006	0.762±0.002	0.740±0.005
w/o DS-AE	0.859±0.002	0.859±0.002	0.901±0.003	0.781±0.001	0.817±0.004	0.718±0.003	0.720±0.001
w/o IntraCAE	0.885±0.004	0.867±0.001	0.904±0.004	0.885±0.006	0.827±0.003	0.744±0.002	0.736±0.003
w/o InterCAE	0.872±0.006	0.864±0.007	0.877±0.002	0.870±0.004	0.851±0.003	0.757±0.003	0.748±0.005
w/o Cross-Attention	0.719±0.004	0.627±0.010	0.810±0.008	0.710±0.002	0.675±0.007	0.603±0.007	0.588±0.010
w/o Adaptive Training	0.855±0.004	0.836±0.003	0.883±0.002	0.882±0.002	0.807±0.005	0.746±0.001	0.740±0.003
ADAPT	0.924±0.007	**0.928**±0.002	**0.921**±0.005	**0.913**±0.004	**0.868**±0.002	**0.794**±0.002	**0.780**±0.003

slices from sagittal view, and 17 slices from coronal and axial views. As compared with table 3 w/o adaptive training, we found that coronal and axial views may contain more differential relationships about the cortex and ventricle of AD and NC, which may help the model learn the specific attention features accurately.

3.3 Ablation Study and Visualization Result

To evaluate how effective each block is, we ablated ADAPT, changing one setting each time as shown in table 3. We find that each component is vital for the final performance. In general, Fusion-Attention contributes the most because of its role in information gathering. Among the four encoders, DS-AE shows its superiority because it extracts detailed features from each 2D slice from different views, not only leading the whole model to focus on special features but also guiding the adaptive training strategy to determine which view is more important.

We visualize the activated area of ADAPT based. Figure 2 shows a NC-related attention map in 3D MRI images from ADNI dataset in sagittal, coronal and axial views. We visualize the attention result after each encoder block. We found the attention mostly focused on some special brain tissues, such as hippocampus, cortex, ventricle and frontal lobe. Disruption of the frontal lobes and its associated networks are a common consequence of neurodegenerative

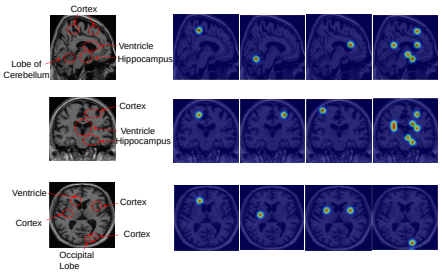

Fig. 2. Attention map for Normal Control result. Each line corresponds to one view dimension: sagittal, coronal and axial.

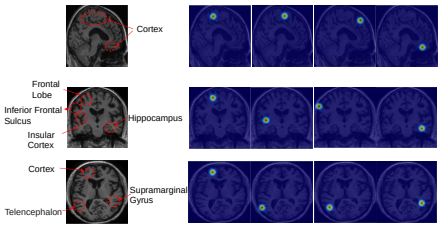

Fig. 3. Attention map for Alzheimer's Disease result. Each line corresponds to one view dimension: sagittal, coronal and axial.

disorders [18], as well as the hippocampus is most notably damaged by AD [22]. Based on these understandings of Alzheimer's pathology [5], ADAPT successfully captured the AD-related part. The hippocampus and cortex begin to atrophy, and the ventricle begins to expand, which can serve as an evidence of morphology augmentation and confirm the reliability of our proposed ADAPT.

4 Conclusions

We presented ADAPT, a clinically-inspired framework for AD diagnosis that integrates pathology-driven augmentation, view-specific encoders, and adaptive training, enabling effective modeling of within and cross-view structural relationships in MRI. Mirroring clinical practice, ADAPT allocates different attention to axial, coronal, sagittal views, allowing the model to prioritize relevant anatomical cues. Experiments show that ADAPT achieves improved accuracy while maintaining significantly lower memory compared to conventional approaches. Visualization analysis further confirms that ADAPT focuses on AD-related regions such as hippocampus and frontal lobe. In the future, we would extend ADAPT to broader clinical applications and brain-related disease diagnosis.

Acknowledgments. The project is partially supported by NIH R03OD038389.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Chen, S., Ma, K., Zheng, Y.: Med3d: Transfer learning for 3d medical image analysis. arXiv preprint arXiv:1904.00625 (2019)
2. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., et al.: An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929 (2020)
3. Ellis, J., Nathan, P.J., Villemagne, V.L., Mulligan, R., Saunder, T., Young, K., Smith, C.L., Welch, J., Woodward, M., Wesnes, K.A., et al.: Galantamine-induced improvements in cognitive function are not related to alterations in α 4 β 2 nicotinic receptors in early alzheimer's disease as measured in vivo by 2-[18 f] fluoro-a-85380 pet. Psychopharmacology **202**, 79–91 (2009)
4. Farooq, A., Anwar, S., Awais, M., Rehman, S.: A deep cnn based multi-class classification of alzheimer's disease using mri. In: 2017 IEEE International Conference on Imaging systems and techniques (IST). pp. 1–6. IEEE (2017)
5. Frisoni, G.B., Fox, N.C., Jack Jr, C.R., Scheltens, P., Thompson, P.M.: The clinical use of structural mri in alzheimer disease. Nature Reviews Neurology **6**(2), 67–77 (2010)
6. Gong, H., Kang, L., Wang, Y., Wan, X., Li, H.: nnmamba: 3d biomedical image segmentation, classification and landmark detection with state space model. arXiv preprint arXiv:2402.03526 (2024)

7. Guo, Q., Feng, W., Zhou, C., Huang, R., Wan, L., Wang, S.: Learning dynamic siamese network for visual object tracking. In: Proceedings of the IEEE international conference on computer vision. pp. 1763–1771 (2017)
8. 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)
9. Jo, T., Nho, K., Saykin, A.J.: Deep learning in alzheimer's disease: diagnostic classification and prognostic prediction using neuroimaging data. Frontiers in aging neuroscience **11**, 220 (2019)
10. Loshchilov, I., Hutter, F.: Sgdr: Stochastic gradient descent with warm restarts. arXiv preprint arXiv:1608.03983 (2016)
11. Luo, W., Li, Y., Urtasun, R., Zemel, R.: Understanding the effective receptive field in deep convolutional neural networks. Advances in neural information processing systems **29** (2016)
12. Malone, I.B., Cash, D., Ridgway, G.R., MacManus, D.G., Ourselin, S., Fox, N.C., Schott, J.M.: Miriad—public release of a multiple time point alzheimer's mr imaging dataset. NeuroImage **70**, 33–36 (2013)
13. Marcus, D.S., Wang, T.H., Parker, J., Csernansky, J.G., Morris, J.C., Buckner, R.L.: Open access series of imaging studies (oasis): cross-sectional mri data in young, middle aged, nondemented, and demented older adults. Journal of cognitive neuroscience **19**(9), 1498–1507 (2007)
14. Mueller, S.G., Weiner, M.W., Thal, L.J., Petersen, R.C., Jack, C.R., Jagust, W., Trojanowski, J.Q., Toga, A.W., Beckett, L.: Ways toward an early diagnosis in alzheimer's disease: the alzheimer's disease neuroimaging initiative (adni). Alzheimer's & Dementia **1**(1), 55–66 (2005)
15. Pérez-García, F., Sparks, R., Ourselin, S.: 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 **208**, 106236 (2021)
16. Qiu, S., Joshi, P.S., Miller, M.I., Xue, C., Zhou, X., Karjadi, C., Chang, G.H., Joshi, A.S., Dwyer, B., Zhu, S., et al.: Development and validation of an interpretable deep learning framework for alzheimer's disease classification. Brain **143**(6), 1920–1933 (2020)
17. Salehi, A.W., Baglat, P., Sharma, B.B., Gupta, G., Upadhya, A.: A cnn model: earlier diagnosis and classification of alzheimer disease using mri. In: 2020 International Conference on Smart Electronics and Communication (ICOSEC). pp. 156–161. IEEE (2020)
18. Sawyer, R.P., Rodriguez-Porcel, F., Hagen, M., Shatz, R., Espay, A.J.: Diagnosing the frontal variant of alzheimer's disease: a clinician's yellow brick road. Journal of clinical movement disorders **4**, 1–9 (2017)
19. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I.: Attention is all you need. Advances in neural information processing systems **30** (2017)
20. Wen, J., Thibeau-Sutre, E., Diaz-Melo, M., Samper-González, J., Routier, A., Bottani, S., Dormont, D., Durrleman, S., Burgos, N., Colliot, O., et al.: Convolutional neural networks for classification of alzheimer's disease: Overview and reproducible evaluation. Medical image analysis **63**, 101694 (2020)
21. Xia, F., Peng, Y., Wang, J., Chen, X.: A 2.5 d multi-path fusion network framework with focusing on z-axis 3d joint for medical image segmentation. Biomedical Signal Processing and Control **91**, 106049 (2024)

22. Xu, F., Ono, M., Ito, T., Uchiumi, O., Wang, F., Zhang, Y., Sun, P., Zhang, Q., Yamaki, S., Yamamoto, R., et al.: Remodeling of projections from ventral hippocampus to prefrontal cortex in alzheimer's mice. Journal of Comparative Neurology **529**(7), 1486–1498 (2021)
23. Yu, Y., Buchanan, S., Pai, D., Chu, T., Wu, Z., Tong, S., Haeffele, B., Ma, Y.: White-box transformers via sparse rate reduction. Advances in Neural Information Processing Systems **36**, 9422–9457 (2023)
24. Zhang, Z., Sabuncu, M.: Generalized cross entropy loss for training deep neural networks with noisy labels. Advances in neural information processing systems **31** (2018)
25. Zhou, Y., Li, Y., Zhou, F., Liu, Y., Tu, L.: Learning with domain-knowledge for generalizable prediction of alzheimer's disease from multi-site structural mri. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 452–461. Springer (2023)