DiaMond: Dementia Diagnosis with Multi-Modal Vision Transformers Using MRI and PET

Yitong Li[1,2] Morteza Ghahremani[1,2] Youssef Wally[1] Christian Wachinger[1,2]

[1]Lab for Artificial Intelligence in Medical Imaging, Technical University of Munich (TUM), Germany
[2]Munich Center for Machine Learning (MCML), Germany
yi_tong.li@tum.de

A. Supplementary Material

A.1. Data Preprocessing and Splitting

The data acquired from the ADNI dataset included all subjects from ADNI 1, ADNI 2, ADNI 3, ADNI GO that have paired MRI and FDG-PET. Data has been through pre-processing steps as shown at the ADNI website for FDG-PET[1] and for MRI[2]. FDG-PET scans were additionally processed using SPM12[3] and CAT12[4]: first, the origin of all images was set to the anterior commissure region which is required for the normalization function in SPM; secondly, the scans are normalized to the MNI space using a 4-th degree B-spline interpolation; thirdly, the normalized images were registered to the MNI152 template with a voxel size of $1.5mm^3$. In the end, all MRI scans are registered to the corresponding PET scans. MRI scans were additionally processed using the CAT12 toolbox with the standard VBM pipeline, consisting the initial and the refined voxel-based processing. The initial pipeline applies a denoising filter, followed by internal resampling. The data is then bias-corrected, affine-registered and lastly followed by the standard SPM unified segmentation. In the second stage, skull-stripping is performed and the brain is parcellated before the final AMAP segmentation step. Finally, the tissue segments are spatially normalized to a common reference space using Geodesic Shooting. Fig. A.1 concludes the whole data preprocessing procedure.

To evenly distribute age, gender, and diagnosis across training, validation, and test data splits, we adapt the data split method from ClinicaDL [42]. First, we assess the balance of each split by computing the propensity score, which represents the probability of a sample being in the training set based on a logistic regression model that includes the known confounders [37, 41]. We then compare the percentiles of the propensity score distribution across the training, validation, and test sets, using the maximum deviation across all percentiles as a measure of imbalance [39]. This process is repeated for 1000 randomly selected partitions and the partition with the minimum imbalance is finally selected.

A.2. Model Parameters

Tab. A.1 presents detailed information on the parameters used in our model and training procedure. We use the AdamW optimizer with a learning rate of 5×10^{-4}, a weight decay of 1×10^{-5}, a dropout rate of 0.0, a batch size of 16, and cosine annealing as the learning rate scheduler. The models are trained on one NVIDIA A100 GPU with 40 GByte memory for 3,800 iterations, with early-stop to prevent overfitting. The ViT-based backbone of DiaMond adopts a patch size of 8, a model depth of 4, a feature embedding dimension of 512, 8 attention heads, τ of 0.01. In total, the model includes $30M$ parameters.

A.3. Training Strategies

We conducted an additional ablation study on different training strategies for DiaMond. DiaMond's three independent branches allow us to initialize individual modality-specific branches with pretrained weights, potentially leveraging prior knowledge. During the training process, we explore two distinct strategies: (1) keeping the pretrained branches static (frozen) to preserve the learned representations, and (2) allowing the branches to continue learning and adapt further (continual learning). We compare these approaches to training the model entirely from scratch using 5-fold cross-validation on the ADNI dataset, specifically focusing on the classification task between CN and AD. The results indicate that incorporating pretrained models provides no substantial advantage over training the model from scratch. This suggests that the model is able to learn the relevant representations adequately through training alone.

[1] https://adni.loni.usc.edu/methods/pet-analysis-method/pet-analysis/
[2] https://adni.loni.usc.edu/methods/mri-tool/mri-analysis/
[3] https://www.fil.ion.ucl.ac.uk/spm/software/spm12/
[4] https://neuro-jena.github.io/cat12-help/

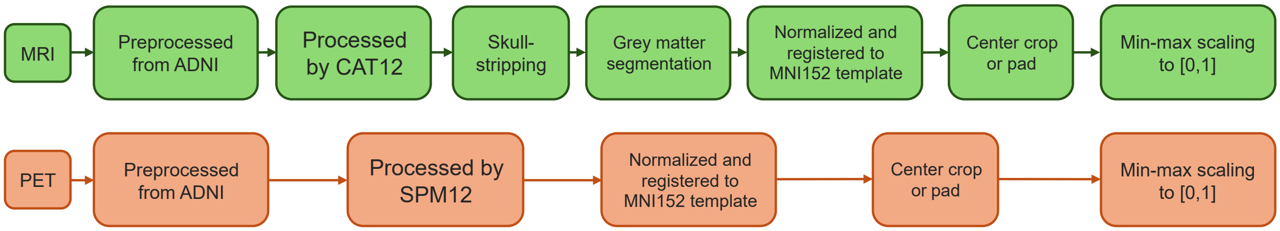

Figure A.1. Dataset preprocessing of MRI and PET data.

Table A.1. Parameters for DiaMond.

	Parameter	Value
Model	Patch size	8
	Depth	4
	Embedding dim.	512
	Attention heads	8
	Model size	30M
	τ	0.01
Training	Batch size	16
	Optimizer	AdamW
	Scheduler	Cosine Annealing
	Learning rate	1×10^{-4}
	Weight Decay	1×10^{-5}
	Dropout	0.0
	Training Iterations	3.8K
	Hardware	one NVIDIA A100 GPU

Table A.2. Ablation on different training strategies.

Training strategy	Frozen	Continual Learning	From Scratch
BACC (%)	91.62 ± 2.67	92.36 ± 2.41	92.42 ± 2.63
AUC (%)	96.50 ± 1.73	97.17 ± 1.25	97.11 ± 1.47

A.4. Differential Diagnosis of Dementia with FreeSurfer Features

Ma et al. [40] uses whole-brain cortical thickness and volume features derived from FreeSurfer [38] in a multi-scale deep neural network (MDNN) for the differential diagnosis of dementia. We further incorporate this method as a benchmark for the differential diagnosis between CN, AD, and FTD. This shape-based method achieves a balanced accuracy of 71.41%, yet it remains lower than the performance of DiaMond.

Table A.3. Differential diagnosis between CN, AD, and FTD on the in-house multi-dementia dataset.

Method	CN vs. AD vs. FTD			
	BACC	F1-Score	Precision	Recall
MDNN [40]	71.41 ± 2.24	71.99 ± 2.12	73.24 ± 1.77	73.50 ± 2.02
DiaMond	**76.46 ± 3.33**	**75.53 ± 4.38**	**76.76 ± 4.88**	**75.39 ± 3.23**

References

[37] Josephine Barnes, Gerard R Ridgway, Jonathan Bartlett, Susie MD Henley, Manja Lehmann, Nicola Hobbs, Matthew J Clarkson, David G MacManus, Sebastien Ourselin, and Nick C Fox. Head size, age and gender adjustment in mri studies: a necessary nuisance? *Neuroimage*, 53(4):1244–1255, 2010. 1

[38] Bruce Fischl, Martin I Sereno, and Anders M Dale. Cortical surface-based analysis: Ii: inflation, flattening, and a surface-based coordinate system. *Neuroimage*, 9(2):195–207, 1999. 2

[39] Daniel E Ho, Kosuke Imai, Gary King, and Elizabeth A Stuart. Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Political analysis*, 15(3):199–236, 2007. 1

[40] Da Ma, Donghuan Lu, Karteek Popuri, Lei Wang, Mirza Faisal Beg, and Alzheimer's Disease Neuroimaging Initiative. 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. 2

[41] Yaakov Stern, Eider M Arenaza-Urquijo, David Bartrés-Faz, Sylvie Belleville, Marc Cantilon, Gael Chetelat, Michael Ewers, Nicolai Franzmeier, Gerd Kempermann, William S Kremen, et al. Whitepaper: Defining and investigating cognitive reserve, brain reserve, and brain maintenance. *Alzheimer's & Dementia*, 16(9):1305–1311, 2020. 1

[42] Elina Thibeau-Sutre, Mauricio Diaz, Ravi Hassanaly, Alexandre M Routier, Didier Dormont, Olivier Colliot, and Ninon Burgos. ClinicaDL: an open-source deep learning software for reproducible neuroimaging processing. *Computer Methods and Programs in Biomedicine*, 220:106818, June 2022. 1