

# Detecting Beta-Amyloid via Cross-Modal Knowledge Distillation from PET to MRI

Francesco Chiumento<sup>1,5</sup>, Julia Dietlmeier<sup>2</sup>, Ronan P. Killeen<sup>3,4</sup>,  
Kathleen M. Curran<sup>2,4,5</sup>, Noel E. O'Connor<sup>1,2,5</sup>, and Mingming Liu<sup>1,2,5</sup>

<sup>1</sup> School of Electronic Engineering, Dublin City University, Dublin, Ireland  
`francesco.chiumento2@mail.dcu.ie, mingming.liu@dcu.ie`

<sup>2</sup> Insight Research Ireland Centre for Data Analytics, DCU, Dublin, Ireland  
`{julia.dietlmeier,noel.oconnor}@insight-centre.org`

<sup>3</sup> St. Vincent's University Hospital, Dublin, Ireland  
`ronan.killeen@svuh.ie`

<sup>4</sup> School of Medicine, University College Dublin, Dublin, Ireland  
`kathleen.curran@ucd.ie`

<sup>5</sup> Research Ireland Centre for Research Training in Machine Learning (ML-Labs),  
Dublin, Ireland

**Abstract.** We approach the task of beta-amyloid detection in Alzheimer's patients and propose a multimodal contrastive and distillation-based learning framework integrating PET (using AV45 and PIB tracers) and 3T MRI data from 790 participants in the OASIS-3 cohort. Built on BiomedCLIP with efficient LoRA fine-tuning, our model leverages cross-modal and self-attention mechanisms with soft triplet loss and adaptive margin to align PET–MRI embeddings. Through knowledge distillation, we transfer PET-guided representations into MRI-only embeddings. A lightweight MLP predicts amyloid positivity from PET-guided MRI representations and MRI embeddings. Our preliminary results demonstrate that PET-guided MRI successfully transfers knowledge to an MRI-only model, showing substantial relative improvement (+8.3% F1 score) compared to models trained without distilled knowledge.

**Keywords:** Contrastive learning · Cross-modal attention · Knowledge distillation · MRI · PET · Transformer

## 1 Introduction

Alzheimer's Disease (AD) is among the most common types of dementia and affects more than 55 million individuals, with this number expected to reach 78 million by 2030 [2, 14]. Beta-amyloid accumulation, the initial neuropathological event in AD, can be identified through Positron Emission Tomography (PET) utilizing radiotracers [1]. However, structural MRI is more widely available, has higher spatial resolution, and is radiation free. Therefore, some studies, such as [9, 11], synthesize PET data from MRI. Contrary to existing works [2], we propose a PET-guided cross-attention approach with knowledge distillation [10] that enhances MRI embeddings for beta-amyloid detection while enabling

MRI-only inference. Consequently, its application to detect beta-amyloid plaques holds significant promise for earlier AD diagnosis and better outcomes.

## 2 Methods

This study used the OASIS-3 dataset [8], with AV45/PIB scans and *T1-weighted MRI scans* aligned within  $\pm 365$  days (790 matched pairs, with a 68:32 class imbalance). MRIs underwent N4 bias correction [7], HD-BET skull-stripping [6] and intensity normalization (clipping [0.5-99.5 percentile], scaling to [0,1], gamma=0.9) [3]. Volumes were reoriented to RAS. Standardized Uptake Value Ratio (SUVR) maps (50–70 min post-injection) were processed using the PUP pipeline and, with MRI volumes, registered to MNI152 space using ANTsPy [4, 13]. Informative sagittal slices were selected by removing low-intensity regions.

### 2.1 Model Architecture and Training

*Training* The model involves two phases: (1) contrastive learning with triplet loss to align MRI-PET representations in a teacher model, followed by (2) knowledge distillation transferring PET-guided features to an MRI-only student model with integrated classification.

*Multimodal Contrastive Pretraining* A multimodal **contrastive learning** framework aligns MRI and PET embeddings through a slice-level Transformer architecture. **BiomedCLIP**’s visual encoder [15] was adapted using LoRA (rank 10, scale 16, and dropout 0.2) [5] applied to attention projections across all 12 self-attention layers with the first 6 blocks frozen. The model encodes 15 uniformly sampled 2D sagittal slices ( $224 \times 224$ ) from each modality. A projection head (**LayerNorm**, **GELU**, **dropout**) maps each slice embedding to 128-space. **Self-attention** enables intra-modality aggregation, while **cross-modal attention** guides MRI features using PET queries. A **soft triplet loss** with adaptive margin aligned embeddings using *anchor* (PET-guided MRI), *positive* (matched PET), and *negative* (PET differing in Centiloid SUVR by 5 units), with distillation losses (MSE and cosine similarity) with epoch-dependent weighting. L2 regularization was applied with AdamW optimization (lr:  $5 \cdot 10^{-6}$  encoder,  $2 \cdot 10^{-5}$  projection head), weight decay of  $10^{-2}$ , cosine annealing learning rate, AMP, and gradient accumulation. This teacher model was trained for 15 epochs with early stopping (patience 5) using CLIP-normalized data augmentation.

*Knowledge Distillation and Classification* In the second phase, knowledge from the PET-guided teacher model was distilled into an MRI-only student model. The student’s self-attention module was initialized with the teacher’s cross-attention weights. A classifier head ( $128 \rightarrow 64 \rightarrow 1$ ), with **LayerNorm**, **ReLU**, and **dropout** predicted amyloid status. Training combined feature distillation loss (MSE between normalized embeddings) and a specialized **MarginFocalLoss** for classification. The weighting between distillation and classification evolved

during training, gradually shifting from distillation (0.5) toward classification (0.8). Optimization used AdamW with differential learning rates ( $2 \cdot 10^{-5}$  for attention,  $8 \cdot 10^{-6}$  for projection, and  $5 \cdot 10^{-5}$  for the classifier), and a ReduceLROnPlateau scheduler based on validation **F1**. The student model was trained for 100 epochs with early stopping (patience 25).

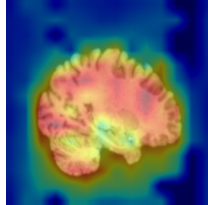
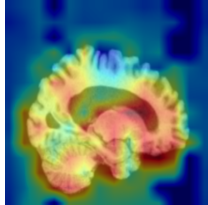
### 3 Results

*Contrastive Pretraining Performance* During training, **soft triplet loss** enforced MRI-PET alignment. Mean *anchor-positive* distance decreased (from 4.15 to 2.02), while *anchor-negative* distance increased (from 6.41 to 9.91), maintaining high cosine similarity between *anchor-positive pairs* (0.85-0.88) while decreasing similarity between *anchor-negative pairs* (from 0.72 to -0.43).

*MRI-based Classification of Amyloid Positivity* The multimodal MRI-PET model achieved high performance (Acc: 0.920, F1: 0.857, opt.  $\theta$ ) with balanced precision-recall. Our distilled MRI-only model outperformed the baseline (opt.  $\theta$ ), improving F1 from 0.482 to 0.522 via enhanced recall (0.756 to 0.795), demonstrating successful transfer of PET-guided knowledge to MRI-only inference.

*Model Interpretability via GradCAM* During training, the PET-guided model attention shifts to more localized and clinically relevant regions (Figure 1) [12].

Metric	MRI-PET		MRI-Dist		MRI-Base	
	$\theta=0.5$	Opt	$\theta=0.5$	Opt	$\theta=0.5$	Opt
Acc	0.913	0.920	<b>0.693</b>	<b>0.573</b>	0.567	0.513
Prec	0.881	0.923	<b>0.455</b>	<b>0.389</b>	0.353	0.354
Rec	0.822	0.800	0.227	<b>0.795</b>	<b>0.533</b>	0.756
F1	0.851	0.857	0.303	<b>0.522</b>	<b>0.425</b>	0.482
Spec	0.952	0.971	<b>0.887</b>	<b>0.481</b>	0.581	0.410
NPV	0.926	0.919	0.734	<b>0.850</b>	<b>0.744</b>	0.796

(a) Epoch 1
(b) Final Epoch

Fig. 1: Performance and GradCAM. Opt = F1-optimized thresholds vs  $\theta = 0.5$ .

### 4 Conclusion

Our cross-modal attention framework for beta-amyloid detection shows promising results in significantly enhancing MRI-only model performance. We trained a teacher model using cross-modal attention to learn PET-MRI relationships, extracting clinically relevant features. Knowledge from this teacher was distilled into an MRI-only student model, enabling accurate amyloid detection without PET at inference time. This approach addresses a critical need where PET is

unavailable. Despite the class imbalance, our distilled MRI-only model demonstrates improvements in F1 score (+8.3%), recall (+5.2%), and NPV (+6.8%) compared to the non-distilled baseline, reducing missed amyloid-positive cases.

**Acknowledgments.** This research was supported by Research Ireland under Grant Research Ireland/12/RC/2289\_P2 and co-funded by the European Regional Development Fund through the Insight Research Ireland Centre for Data Analytics at DCU. F. Chiumento was also supported by the Research Ireland Centre for Research Training in Machine Learning (ML-Labs) at DCU.

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

## References

1. Bao, Y.W., Wang, Z.J., Shea, Y.F., Chiu, P.K.C., Kwan, J.S., Chan, F.H.W., Mak, H.K.F.: Combined quantitative amyloid- $\beta$  pet and structural mri features improve alzheimer’s disease classification in random forest model - a multicenter study. *Academic Radiology* **31**(12), 5154–5163 (2024)
2. Chattopadhyay, T., Ozarkar, S.S., Buwa, K., Joshy, N.A., Komandur, D., Naik, J., Thomopoulos, S.I., Ver Steeg, G., Ambite, J.L., Thompson, P.M.: Comparison of deep learning architectures for predicting amyloid positivity in Alzheimer’s disease, mild cognitive impairment, and healthy aging, from T1-weighted brain structural MRI. *Frontiers in Neuroscience* **18** (Jul 2024). <https://doi.org/10.3389/fnins.2024.1387196>
3. Dang, K., Vo, T., Ngo, L., Ha, H.: A deep learning framework integrating MRI image preprocessing methods for brain tumor segmentation and classification. *IBRO Neuroscience Reports* **13**, 523–532 (Dec 2022). <https://doi.org/10.1016/j.ibneur.2022.10.014>
4. Developers, A.: ANTsX/ANTsPy: Advanced normalization tools in python. <https://github.com/ANTsX/ANTsPy> (2025)
5. Hu, E.J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L., Chen, W.: LoRA: Low-Rank Adaptation of Large Language Models (Oct 2021). <https://doi.org/10.48550/arXiv.2106.09685>
6. Isensee, F., Schell, M., Pflueger, I., Brugnara, G., Bonekamp, D., Neuberger, U., Wick, A., Schlemmer, H.P., Heiland, S., Wick, W., Bendszus, M., Maier-Hein, K.H., Kickingeder, P.: Automated brain extraction of multisequence MRI using artificial neural networks. *Human Brain Mapping* **40**(17), 4952–4964 (2019). <https://doi.org/10.1002/hbm.24750>
7. Kanakaraj, P., Yao, T., Cai, L.Y., Lee, H.H., Newlin, N.R., Kim, M.E., Gao, C., Pechman, K.R., Archer, D., Hohman, T., Jefferson, A., Beason-Held, L.L., Resnick, S.M., Garyfallidis, E., Anderson, A., Schilling, K.G., Landman, B.A., Moyer, D.: DeepN4: Learning N4ITK Bias Field Correction for T1-weighted Images. *Research Square* pp. rs.3.rs-3585882 (2023). <https://doi.org/10.21203/rs.3.rs-3585882/v1>
8. LaMontagne, P.J., Benzinger, T.L., Morris, J.C., Keefe, S., Hornbeck, R., Xiong, C., Grant, E., Hassenstab, J., Moulder, K., Vlassenko, A.G., Raichle, M.E., Cruchaga, C., Marcus, D.: OASIS-3: Longitudinal Neuroimaging, Clinical, and Cognitive Dataset for Normal Aging and Alzheimer Disease (2019). <https://doi.org/10.1101/2019.12.13.19014902>

9. Lyu, Q., Kim, J.Y., Kim, J., Whitlow, C.T.: Synthesizing beta-amyloid pet images from t1-weighted structural mri: A preliminary study. *arXiv* (2024)
10. Moslemi, A., Briskina, A., Dang, Z., Li, J.: A survey on knowledge distillation: Recent advancements. *Machine Learning with Applications* **18**, 100605 (Dec 2024). <https://doi.org/10.1016/j.mlwa.2024.100605>
11. Ou, Z., Pan, Y., Xie, F., Guo, Q., Shen, D.: Image-and-label conditioning latent diffusion model: Synthesizing  $\alpha/\beta$ -pet from mri for detecting amyloid status. *IEEE Journal of Biomedical and Health Informatics* **29**(2), 1221 – 1231 (2025)
12. Palmqvist, S., Schöll, M., Strandberg, O., Mattsson, N., Stomrud, E., Zetterberg, H., Blennow, K., Landau, S., Jagust, W., Hansson, O.: Earliest accumulation of  $\beta$ -amyloid occurs within the default-mode network and concurrently affects brain connectivity. *Nature Communications* **8**(1), 1214 (2017). <https://doi.org/10.1038/s41467-017-01150-x>
13. Vega, F., Addeh, A., Ganesh, A., Smith, E.E., MacDonald, M.E.: Image Translation for Estimating Two-Dimensional Axial Amyloid-Beta PET From Structural MRI. *Journal of Magnetic Resonance Imaging* **59**(3), 1021–1031 (2024). <https://doi.org/10.1002/jmri.29070>
14. World Health Organization: Dementia. <https://www.who.int/news-room/fact-sheets/detail/dementia> (2025)
15. Zhang, S., Xu, Y., Usuyama, N., Xu, H., Bagga, J., Tinn, R., Preston, S., Rao, R., Wei, M., Valluri, N., Wong, C., Tupini, A., Wang, Y., Mazzola, M., Shukla, S., Liden, L., Gao, J., Crabtree, A., Piening, B., Bifulco, C., Lungren, M.P., Naumann, T., Wang, S., Poon, H.: BiomedCLIP: A multimodal biomedical foundation model pretrained from fifteen million scientific image-text pairs (Jan 2025). <https://doi.org/10.48550/arXiv.2303.00915>