Detecting Beta-Amyloid via Transformer-Based Multimodal Integration of MRI and PET Imaging

Francesco Chiumento[1,5] FRANCESCO.CHIUMENTO2@MAIL.DCU.IE
[1] *School of Electronic Engineering, Dublin City University, Dublin, Ireland*
Julia Dietlmeier[2] JULIA.DIETLMEIER@INSIGHT-CENTRE.ORG
[2] *Insight Research Ireland Centre for Data Analytics, DCU, Dublin, Ireland*
Ronan P. Killeen[3,4] RONAN.KILLEEN@SVUH.IE
[3] *St Vincent's University Hospital, Dublin, Ireland*
Kathleen M. Curran[2,4,5] KATHLEEN.CURRAN@UCD.IE
[4] *School of Medicine, University College Dublin, Dublin, Ireland*
[5] *Research Ireland Centre for Research Training in Machine Learning (ML-Labs), Dublin, Ireland*
Noel E. O'Connor[1,2,5] NOEL.OCONNOR@INSIGHT-CENTRE.ORG
Mingming Liu[1,2,5] MINGMING.LIU@DCU.IE

Editors: Under Review for MIDL 2025

Abstract

In this paper, we propose a multimodal contrastive learning framework that integrates AV45 PET and 3T MRI data from 511 baseline participants in the OASIS-3 cohort. Built on BiomedCLIP, our model incorporates cross-modal attention and a soft triplet loss with adaptive margin to align PET–MRI embeddings. After contrastive pretraining, a lightweight MLP predicts amyloid positivity using PET-guided MRI representations. Results show our approach learns robust MRI features that capture PET-derived signals for reliable beta-amyloid prediction.

Keywords: Contrastive Learning, Cross-modal Attention, MRI, PET, Transformer

1. Introduction

Alzheimer's Disease is among the most common types of dementia and affects more than 55 million individuals, with the number expected to reach 78 million by the year 2030 (Chattopadhyay et al., 2024; World Health Organization, 2025). The accumulation of beta-amyloid is considered the initial neuropathological event in the brain during the course of Alzheimer's Disease. The presence of beta-amyloid deposits can be identified through Positron Emission Tomography (PET) utilizing radiotracers (Bao et al., 2024). However, structural MRI is more widely available, and some studies, such as (Lyu et al., 2024; Ou et al., 2025), synthesize PET data from MRI. Contrary to existing works (Chattopadhyay et al., 2024), we propose a PET-guided cross-attention approach to enhance MRI embeddings through contrastive learning for beta-amyloid detection.

2. Methods

This study used the OASIS-3 dataset (LaMontagne et al., 2019), including AV45 scans and the corresponding *T1-weighted MRI scans* aligned within ±365 days, with a total of 511 matched pairs after preprocessing. All T1-weighted MRIs underwent **N4 bias correction** (Kanakaraj et al., 2023), HD-BET skull-stripping (Isensee et al., 2019) and intensity normalization (clipping [0.5-99.5 percentile], scaling to [0,1], gamma=0.9) (Dang et al., 2022). 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** (Developers, 2025; Vega et al., 2024). Informative sagittal slices were selected by removing low-intensity regions.

2.1. Model Architecture and Training

The training involves: (1) contrastive pretraining to align MRI and PET in a shared embedding space, and (2) supervised classification based on PET-guided MRI embeddings. All experiments were conducted on a workstation with an NVIDIA GeForce RTX 4090 GPU.

Multimodal Contrastive Pretraining A multimodal **contrastive learning** framework was implemented to align MRI and PET embeddings through a slice-level Transformer architecture. The visual encoder from **BiomedCLIP** (Zhang et al., 2025) was adapted to the neuroimaging domain using **Low-Rank Adaptation (LoRA)** with rank 10, scale 16 and dropout 0.2 (Hu et al., 2021), applied to the *query*, *key*, and *value* projections across all 12 self-attention layers of the vision Transformer. The first 6 Transformer blocks were frozen. The model encodes 15 uniformly sampled 2D sagittal slices (224×224) from each modality. A projection head (`LayerNorm`, `GELU`, `dropout`) maps each slice embedding to 128-dimensional space. Two attention modules were used: **self-attention** for intra-modality aggregation, and **cross-modal attention** to guide MRI features using PET queries.

A **soft triplet loss** with an adaptive margin aligned embeddings using *anchor* (PET-guided MRI), positive (matched PET) and negative (PET differing in Centiloid SUVR by 5 units). The margin scaled with the average batch-wise inter-modality distance. L2 regularization was applied and the model was trained using `AdamW` (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. The model was trained for 40 epochs with early stopping (patience 5) based on triplet loss and total loss. Data augmentation was applied, and images were normalized using CLIP statistics. At the end of training, PET-guided MRI embeddings were extracted via cross-modal attention using PET as *query* and MRI as *key*/*value*.

MRI-based Classification from PET-guided Embeddings A feedforward neural network was trained to predict amyloid positivity from PET-guided MRI embeddings. The architecture included three fully connected layers (128, 64, 1 units) with batch normalization, ReLU activations, and dropout (0.6, 0.5). The `BCEWithLogitsLoss` function was used for binary classification, incorporating class imbalance through a dynamic `pos_weight` ratio. Optimization used `Adam` (learning rate of $1 \cdot 10^{-4}$ and a weight decay of $1 \cdot 10^{-3}$). Training was conducted for up to 500 epochs with early stopping (patience = 30), based on the validation **F1 score**. The *optimal classification threshold* (θ) was selected at each epoch to maximize the F1 score over a range of candidate thresholds.

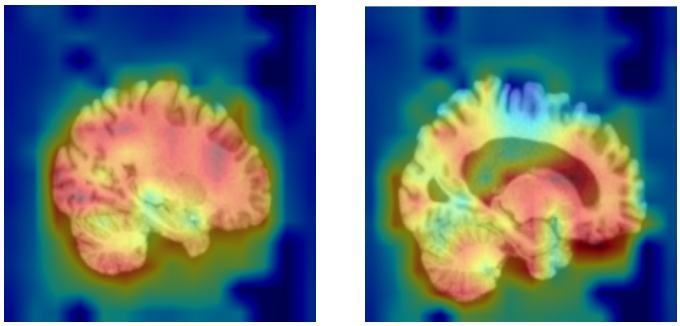
3. Results

Contrastive Pretraining Performance During the contrastive training phase, the soft triplet loss enforced alignment between MRI and PET embeddings. The mean anchor–positive distance progressively decreased (from 4.9243 to 2.5767), while the anchor–negative distance increased (from 8.0486 to 16.4347), also resulting in an increase in cosine similarity between anchor and positive (from 0.8868 to 0.9292) and a decrease in similarity between anchor and negative (from 0.7184 to -0.6840).

MRI-based Classification of Amyloid Positivity Using the fixed threshold (0.5), the model achieved high precision (0.909) and good overall balance. Optimizing the threshold led to an increase in recall (from 0.833 to 0.889) and Negative Predictive Value (NPV) (from 0.913 to 0.934), but at the cost of lower precision, resulting in a decrease in the F1 score.

Model Interpretability via GradCAM As shown in Figure 1, during training epochs, the model progressively focuses its attention on more localized and clinically relevant regions (Palmqvist et al., 2017; Zhou et al., 2022).

Metric	$\boldsymbol{\theta = 0.5}$	**Opt** $\boldsymbol{\theta}$
Accuracy	0.912	0.873
Precision	0.909	0.780
Recall	0.833	**0.889**
F1 Score	**0.870**	0.831
Specificity	**0.955**	0.864
NPV	0.913	**0.934**



(*a*) Epoch 1 (*b*) Final Epoch

Figure 1: Metrics (left) and GradCAM attention shift over training (right).

4. Conclusion

Our cross-modal attention framework for beta-amyloid detection demonstrates high precision (0.909) and strong F1 score (0.870) on 511 OASIS-3 patients. GradCAM visualizations show the model focusing on brain regions linked to amyloid deposition, confirming that contrastive learning effectively aligns MRI and PET embeddings. This alignment enables the extraction of clinically relevant features. Future work will expand cohort diversity, explore alternative tracers, and develop methods that use cross-modal knowledge during training but rely only on MRI at inference, enhancing clinical applicability and early detection.

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.

References

Y.-W. Bao, Z.-J. Wang, Y.-F. Shea, P. K.-C. Chiu, J. S.K. Kwan, F. H.-W. Chan, and H. K.-F. Mak. Combined quantitative amyloid-β pet and structural mri features improve alzheimer's disease classification in random forest model - a multicenter study. *Academic Radiology*, 31(12):5154–5163, 2024.

Tamoghna Chattopadhyay, Saket S. Ozarkar, Ketaki Buwa, Neha Ann Joshy, Dheeraj Komandur, Jayati Naik, Sophia I. Thomopoulos, Greg Ver Steeg, Jose Luis Ambite, and Paul M. Thompson. 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, July 2024. doi: 10.3389/fnins.2024.1387196.

Khiet Dang, Toi Vo, Lua Ngo, and Huong Ha. A deep learning framework integrating MRI image preprocessing methods for brain tumor segmentation and classification. *IBRO Neuroscience Reports*, 13:523–532, December 2022. doi: 10.1016/j.ibneur.2022.10.014.

ANTsX Developers. ANTsX/ANTsPy: Advanced normalization tools in python. https://github.com/ANTsX/ANTsPy, 2025.

Edward J. Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. LoRA: Low-Rank Adaptation of Large Language Models, October 2021.

Fabian Isensee, Marianne Schell, Irada Pflueger, Gianluca Brugnara, David Bonekamp, Ulf Neuberger, Antje Wick, Heinz-Peter Schlemmer, Sabine Heiland, Wolfgang Wick, Martin Bendszus, Klaus H. Maier-Hein, and Philipp Kickingereder. Automated brain extraction of multisequence MRI using artificial neural networks. *Human Brain Mapping*, 40(17): 4952–4964, 2019. doi: 10.1002/hbm.24750.

Praitayini Kanakaraj, Tianyuan Yao, Leon Y. Cai, Ho Hin Lee, Nancy R. Newlin, Michael E. Kim, Chenyu Gao, Kimberly R. Pechman, Derek Archer, Timothy Hohman, Angela Jefferson, Lori L. Beason-Held, Susan M. Resnick, Eleftherios Garyfallidis, Adam Anderson, Kurt G. Schilling, Bennett A. Landman, and Daniel Moyer. DeepN4: Learning N4ITK Bias Field Correction for T1-weighted Images. *Research Square*, pages rs.3.rs–3585882, 2023. doi: 10.21203/rs.3.rs-3585882/v1.

Pamela J. LaMontagne, Tammie LS Benzinger, John C. Morris, Sarah Keefe, Russ Hornbeck, Chengjie Xiong, Elizabeth Grant, Jason Hassenstab, Krista Moulder, Andrei G. Vlassenko, Marcus E. Raichle, Carlos Cruchaga, and Daniel Marcus. OASIS-3: Longitudinal Neuroimaging, Clinical, and Cognitive Dataset for Normal Aging and Alzheimer Disease, 2019.

Q. Lyu, J. Y. Kim, J. Kim, and C. T. Whitlow. Synthesizing beta-amyloid pet images from t1-weighted structural mri: A preliminary study. *arXiv*, 2024.

Z. Ou, Y. Pan, F. Xie, Q. Guo, and D. Shen. Image-and-label conditioning latent diffusion model: Synthesizing aβ-pet from mri for detecting amyloid status. *IEEE Journal of Biomedical and Health Informatics*, 29(2):1221 – 1231, 2025.

Sebastian Palmqvist, Michael Schöll, Olof Strandberg, Niklas Mattsson, Erik Stomrud, Henrik Zetterberg, Kaj Blennow, Susan Landau, William Jagust, and Oskar Hansson. Earliest accumulation of β-amyloid occurs within the default-mode network and concurrently affects brain connectivity. *Nature Communications*, 8(1):1214, 2017. doi: 10.1038/s41467-017-01150-x.

Fernando Vega, Abdoljalil Addeh, Aravind Ganesh, Eric E. Smith, and M. Ethan MacDonald. Image Translation for Estimating Two-Dimensional Axial Amyloid-Beta PET From Structural MRI. *Journal of Magnetic Resonance Imaging*, 59(3):1021–1031, 2024. doi: 10.1002/jmri.29070.

World Health Organization. Dementia. https://www.who.int/news-room/fact-sheets/detail/dementia, 2025. Accessed: 2025-04-09.

Sheng Zhang, Yanbo Xu, Naoto Usuyama, Hanwen Xu, Jaspreet Bagga, Robert Tinn, Sam Preston, Rajesh Rao, Mu Wei, Naveen Valluri, Cliff Wong, Andrea Tupini, Yu Wang, Matt Mazzola, Swadheen Shukla, Lars Liden, Jianfeng Gao, Angela Crabtree, Brian Piening, Carlo Bifulco, Matthew P. Lungren, Tristan Naumann, Sheng Wang, and Hoifung Poon. BiomedCLIP: A multimodal biomedical foundation model pretrained from fifteen million scientific image-text pairs, January 2025.

Da-An Zhou, Kai Xu, Xiaobin Zhao, Qian Chen, Feng Sang, Di Fan, Li Su, Zhanjun Zhang, Lin Ai, and Yaojing Chen. Spatial Distribution and Hierarchical Clustering of β-Amyloid and Glucose Metabolism in Alzheimer's Disease. *Frontiers in Aging Neuroscience*, 14: 788567, June 2022. doi: 10.3389/fnagi.2022.788567.