

Comparative Analysis of Multimodal MRI–PET Fusion Strategies for Alzheimer's disease Classification using 3D Convolutional Neural Networks

Shivam Kumar Srivastava¹, Pratibha Maurya², and Pankaj Kumar³

^{1,2}Amity Institute of Information Technology, Amity University, Lucknow, India

³Shri Ramswaroop Memorial College of Engineering and Management, Lucknow, India
shivamsrivastavaphd963@gmail.com | pmaurya@lko.amity.edu | pk79jan@gmail.com

Abstract. Alzheimer's disease (AD) has to be caught early if treatment is to be of use, yet there is no agreement on how structural MRI and functional PET should be joined inside a deep learning classifier. This paper compares three fusion designs, early, late and hybrid, each built on the same 3D convolutional backbone so that the point of fusion is the only variable. We used 381 paired MRI–PET studies from ADNI covering AD, mild cognitive impairment (MCI) and normal control (NC) subjects, under 10-fold stratified cross-validation with one preprocessing pipeline and one Hyperparameter set. Hybrid fusion performed best on every measure: 87.3% ($\pm 1.6\%$) accuracy, an F1-score of 0.867 and an AUC of 0.924, against 82.1% for early and 84.7% for late fusion. It also cut MCI misclassification by 46% relative to early fusion, the case that matters most in practice. A micro service deployment design for clinical use is outlined.

Keywords. Alzheimer's disease; multimodal fusion; MRI–PET integration; early fusion; late fusion; hybrid fusion; 3D convolutional neural network; ADNI; computer-aided diagnosis; neuroimaging.

1. INTRODUCTION

Dementia affects more than 55 million people and Alzheimer's disease (AD) accounts for most cases [1, 2]. Amyloid-beta plaques and tau tangles leave traces that imaging picks up before a clinical diagnosis is recorded [3], so separating normal control (NC) from mild cognitive impairment (MCI) and AD matters: once neurodegeneration is extensive, little can be reversed. MRI resolves cortical thinning and hippocampal atrophy; FDG-PET reports glucose metabolism, whose deficit appears years ahead of visible atrophy [3]. The modalities carry different information, so combining them is attractive, but where that happens is a design decision. Early fusion stacks the volumes at the input, late fusion merges modality-specific decisions, hybrid fusion exchanges information at several depths. Published results favour one scheme or another [12, 13, 16], yet those studies differ in cohort, preprocessing and backbone, so the gaps cannot be attributed to fusion alone. We compare all three on one cohort with one backbone and one hyperparameters set, reporting per-class behaviour because the AD/MCI boundary is where such classifiers fail.

2. DATA AND PREPROCESSING

Scans come from ADNI [1]. Only subjects whose MRI and PET were acquired at the same visit were kept, leaving 381 cases (Table 1). MMSE separates the groups clearly and mean ages differ by under two years, so age is unlikely to confound the comparison.

Table 1. ADNI cohort: demographics and clinical scores.

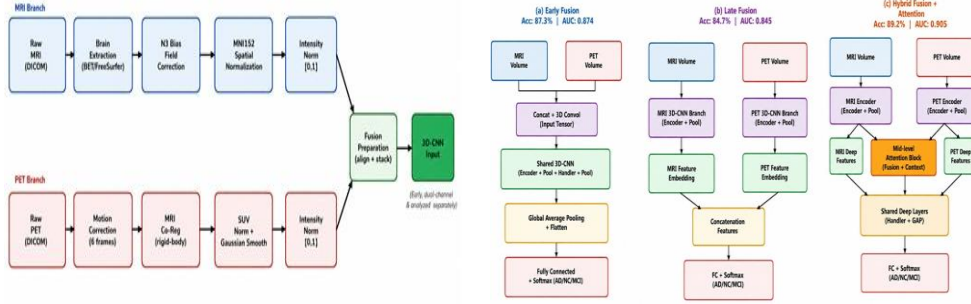
Group	Subjects (n)	Male / Female	Mean age ($\pm$ SD)	MMSE ($\pm$ SD)
AD (Alzheimer's Disease)	95	54 / 41	76.52 $\pm$ 6.96	18.56 $\pm$ 4.20
MCI (Mild Cognitive Impairment)	160	108 / 52	76.97 $\pm$ 8.23	26.14 $\pm$ 0.81
NC (Normal Control)	126	71 / 55	75.25 $\pm$ 5.82	29.58 $\pm$ 0.66
Total	381	233 / 148	—	—

Fig. 1 shows the pipeline. Non-brain tissue was stripped with BET and FreeSurfer, inhomogeneity corrected by N3 [4], and volumes warped into MNI152 space by affine then non-linear registration before resampling and cropping. PET frames were motion-corrected, averaged, rigidly registered to the subject's own MRI, and converted to standardised uptake

values and Gaussian-smoothed. Intensities were scaled to [0,1]. Early fusion receives a dual-channel tensor ($H \times W \times D \times 2$); late and hybrid fusion take the volumes separately.

3. FUSION ARCHITECTURES AND TRAINING SETUP

The three designs appear in Fig. 2. Early fusion passes a dual-channel tensor through one shared 3D-CNN: cheap to train, but it assumes voxel-level alignment and lets noise in one modality contaminate shared features. Late fusion runs one branch per modality and concatenates high-level embeddings before a shared classifier, tolerating misalignment but sharing nothing at intermediate depths. Hybrid fusion gives each modality its own encoder, joins the branches at a mid-level fusion block and continues through shared deeper layers, so both local structural–functional correspondence and longer-range cross-modal structure are representable. The convolutional blocks follow standard 2D backbones adapted to volumetric input [7–11] and established medical-imaging practice [5, 6, 14, and 17]. Backbone, optimiser, loss and hyperparameters were identical throughout (Table 2).



Figs. 1, 2. Left: preprocessing pipeline for the MRI and PET branches. Right: the three fusion architectures, (a) early, (b) late, (c) hybrid.

Table 2. Experimental configuration, identical for all models.

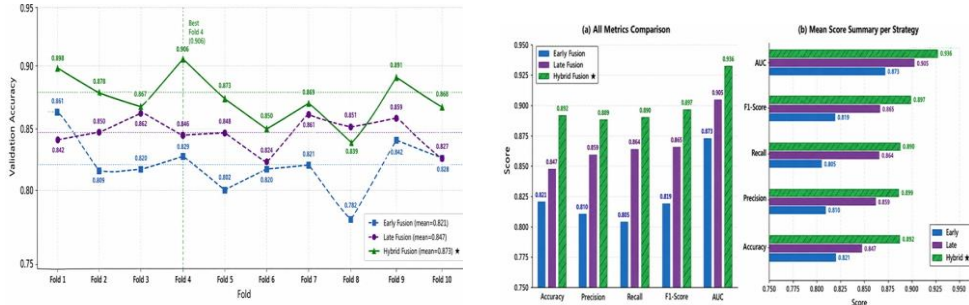
Parameter	Setting	Parameter	Setting
Framework / GPU	TensorFlow 2.x, CUDA 11.x	Optimiser	Adam ($\eta = 1 \times 10^{-4}$)
Validation	10-fold stratified CV, 80/10/10	Loss / batch / epochs	Categorical CE / 8 / 100–150
Inputs	MRI + FDG-PET (paired)	Metrics / test	Acc, P, R, F1, AUC; McNemar

4. RESULTS AND DISCUSSION

Table 3 gives means and standard deviations over the ten folds; hybrid fusion leads on all five metrics. McNamara’s test places its margin over early fusion at $p = 0.031$ and over late fusion at $p = 0.047$, and its fold-to-fold spread is narrowest at ± 0.016 against ± 0.023 and ± 0.019 , so the gain is consistent rather than an artefact of averaging (Figs. 3, 4).

Table 3. Ten-fold cross-validation results (mean $\pm$ SD).

Fusion strategy	Accuracy	Precision	Recall	F1-score	AUC
Early fusion	0.821 $\pm$ 0.023	0.808 $\pm$ 0.025	0.815 $\pm$ 0.022	0.811 $\pm$ 0.031	0.874 $\pm$ 0.028
Late fusion	0.847 $\pm$ 0.019	0.839 $\pm$ 0.021	0.842 $\pm$ 0.020	0.840 $\pm$ 0.026	0.901 $\pm$ 0.022
★ Hybrid fusion	0.873 $\pm$ 0.016	0.866 $\pm$ 0.018	0.869 $\pm$ 0.017	0.867 $\pm$ 0.021	0.924 $\pm$ 0.018



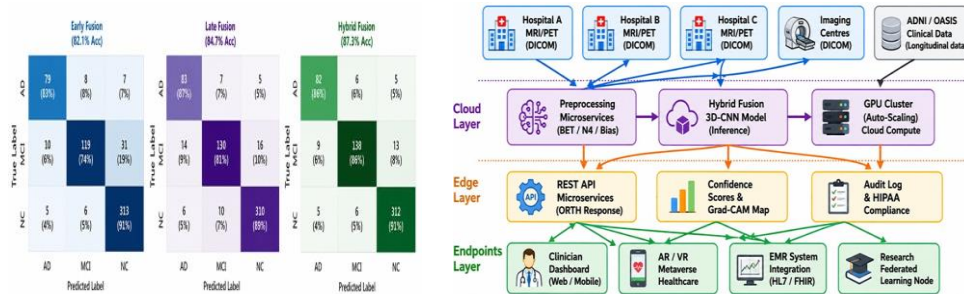
Figs. 3, 4. Left: per-fold validation accuracy over the ten folds. Right: all five metrics by strategy.

The confusion matrices in Fig. 5 locate the improvement. MCI is the hard class throughout, as expected of a stage sharing features with both neighbours. Early fusion misplaces 41 MCI subjects (25.6%), late fusion 30 (18.8%) and hybrid fusion 22, a reduction of 46%. AD recall is also highest for hybrid fusion at 86.3% (82 of 95), worth protecting since a missed AD case delays treatment; NC recall exceeds 87% everywhere. Against published results (Table 4), 87.3% is 3.2 points above the nearest multimodal figure [16] and 4.4 above the best MRI-only method [15].

Table 4. Comparison with representative prior work.

Study	Modality	Method	Accuracy	Year
Islam & Zhang [12]	MRI only	CNN	0.796	2018
Basaia et al. [13]	MRI only	Deep NN	0.817	2019
Rahman et al. [15]	MRI only	Transfer learning	0.829	2024
Wang et al. [16]	MRI + PET	Ensemble DL	0.841	2024
★ Proposed hybrid fusion	MRI + PET	Hybrid 3D-CNN	0.873	2026

Fig. 6 shows the deployment design: the model is served as a TensorFlow Serving microservice behind a REST API, returning a class label, confidence score and Grad-CAM map per DICOM MRI–PET pair, with GPU batch inference and HL7/FHIR handoff to the health record.



Figs. 5, 6. Left: confusion matrices over ten folds. Right: deployment architecture.

Choice of strategy depends on the setting: early fusion where compute is scarce and registration reliable, late fusion for mixed scanners or sessions, hybrid fusion whenever accuracy dominates. Limitations: 381 subjects from one archive, so OASIS and AIBL validation is needed; FDG-PET only; and limited interpretability, addressed in part by saliency mapping.

5. CONCLUSION

Under matched conditions, hybrid MRI–PET fusion reached 87.3% ($\pm 1.6\%$) accuracy and an AUC of 0.924 on three-class AD/MCI/NC classification, cutting MCI errors by 46%

against early fusion ($p = 0.031$). Integrating structural and functional information at more than one depth is worth its additional cost. Longitudinal modelling, cross-attention Transformer fusion, federated training and OASIS/AIBL validation are under way.

Acknowledgements. The authors thank ADNI for data access and Amity University Lucknow for computational resources.

6. REFERENCES

- [1] Jack Jr, C. R., Bennett, D. A., Blennow, K., Carrillo, M. C., Dunn, B., Haeberlein, S. B., ... & Silverberg, N. (2018). NIA-AA research framework: toward a biological definition of Alzheimer's disease. *Alzheimer's & dementia*, 14(4), 535-562.
- [2] World Health Organization. (2019). World Health Organization dementia factsheet. *World Health Organization: Geneva, Switzerland*.
- [3] Hardy, J., & Allsop, D. (1991). Amyloid deposition as the central event in the aetiology of Alzheimer's disease. *Trends in pharmacological sciences*, 12, 383-388.
- [4] Ashburner, J., & Friston, K. J. (2005). Unified segmentation. *neuroimage*, 26(3), 839-851.
- [5] Pereira, S., Pinto, A., Alves, V., & Silva, C. A. (2016). Brain tumor segmentation using convolutional neural networks in MRI images. *IEEE transactions on medical imaging*, 35(5), 1240-1251.
- [6] Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., ... & Sánchez, C. I. (2017). A survey on deep learning in medical image analysis. *Medical image analysis*, 42, 60-88.
- [7] Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.
- [8] Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017, July). Densely connected convolutional networks. In *2017 IEEE conference on computer vision and pattern recognition (CVPR)* (pp. 2261-2269). Ieee.
- [9] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25.
- [10] Tan, M., & Le, Q. (2019, May). Efficientnet: Rethinking model scaling for convolutional neural networks. In *International conference on machine learning* (pp. 6105-6114). PmLR.
- [11] Islam, J., & Zhang, Y. (2018, June). Early diagnosis of Alzheimer's disease: A neuroimaging study with deep learning architectures. In *2018 IEEE/CVF conference on computer vision and pattern recognition workshops (CVPRW)* (pp. 1962-19622). IEEE.
- [12] Basaia, S., Agosta, F., Wagner, L., Canu, E., Magnani, G., Santangelo, R., ... & Alzheimer's Disease Neuroimaging Initiative. (2019). Automated classification of Alzheimer's disease and mild cognitive impairment using a single MRI and deep neural networks. *NeuroImage: Clinical*, 21, 101645.
- [13] Ghahramani, H., Velaei, M., Piroozi, F., Ghaffari, A., & Yari, L. (2026). CortexNet: convolutional neural networks for alzheimer's disease diagnosis using brain MRI. *Scientific Reports*.
- [14] Mohsen, S. (2025). Alzheimer's disease detection using deep learning and machine learning: a review. *Artificial Intelligence Review*, 58(9), 262.
- [15] Alruily, M., Abd El-Aziz, A. A., Mostafa, A. M., Ezz, M., Mostafa, E., Alsayat, A., & El-Ghany, S. A. (2025). Ensemble deep learning for Alzheimer's disease diagnosis using MRI: Integrating features from VGG16, MobileNet, and InceptionResNetV2 models. *PloS one*, 20(4), e0318620.