Structural MRI–Informed Multimodal Fusion for Robust Alzheimer's Disease Prediction

Anonymous Author(s)
Affiliation
Address
email

Abstract

Alzheimer's Disease (AD) is a devastating neurodegenerative disorder, and accurate prediction remains a critical challenge. Multiple modalities can inform AD prediction, with public resources such as the Alzheimer's Disease Neuroimaging Initiative (ADNI) providing access to diverse datasets. However, prior studies often rely on single modalities, limiting clinical applicability, or struggle to integrate multimodal data effectively. In this work, we introduce MOIRA (Multi-Omics Integration with Robustness to Absent modalities), a predictive framework that leverages the strong discriminative power of structural Magnetic Resonance Imaging (sMRI) while flexibly incorporating additional modalities to boost performance. MOIRA achieves 0.91 accuracy, substantially surpassing existing approaches. Notably, we show that our model trained with sMRI can still improve prediction without sMRI data at inference, supporting potentially cost-efficient diagnostic strategies in clinical settings. Our findings highlight the value of sMRI-informed multimodal integration for advancing robust, translational AD prediction.

1 Introduction

Among progressive neurodegenerative conditions, AD is marked by amyloid-beta accumulation and related pathological processes that ultimately lead to neuronal loss [4, 28]. Clinical diagnosis of AD typically draws on diverse sources, including medical records, cognitive assessments, genetic history, and neuroimaging [9], which are now collected at unprecedented scale in public datasets. Leveraging these multimodal inputs can offer a more holistic and thorough understanding of disease progression.

Neuroimaging techniques such as sMRI offer a non-invasive means to assess neurodegeneration through detailed brain structural analysis. Notably, sMRI enables accurate in vivo quantification of brain regions associated with AD [11]. However, sMRI-based diagnosis is limited to detecting anatomical changes, leaving the underlying pathogenesis largely unresolved [24]. High-throughput genomic profiling offers valuable insights into the molecular mechanisms of AD and holds promise for early-stage detection. Moreover, cerebrospinal fluid (CSF) biomarkers are often recommended in diagnostically challenging or atypical cases [25]. Nevertheless, relying solely on biomarkers is insufficient for definitive diagnosis or precise prediction of disease progression [5]. Consequently, integrating neuroimaging with omics data enables a more comprehensive characterization of heterogeneous AD phenotypes by capturing complementary morphological and molecular information.

Despite growing interest in multimodal integration for AD prediction, prior work has often focused on single modalities, leading to substantial data exclusion and limiting applicability in clinical practice. More recent approaches have attempted to integrate multiple data types, yet they often fail to achieve stable prediction of phenotypes. For example, Flex-MoE [27] introduced a Sparse Mixture of Experts (SMoE) framework capable of handling arbitrary combinations of input modalities and addressing missing modality scenarios. In this approach, each available modality is processed by a dedicated

Submitted to 39th Conference on Neural Information Processing Systems (NeurIPS 2025). Do not distribute.

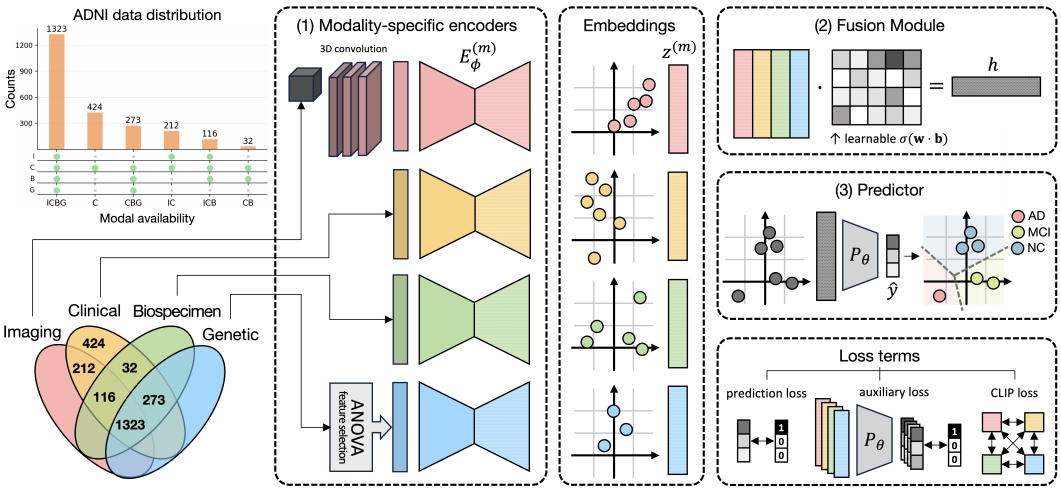

Figure 1: Overview of the ADNI dataset and MOIRA architecture. Four modalities were collected: Imaging (I), Clinical (C), Biospecimen (B), and Genetic (G). Sample sizes for their combinations are shown (top left). Imaging and Genetic data are preprocessed using 3D CNN and ANOVA feature selection, respectively. Each modality is processed by (1) modality-specific encoders to produce embeddings, which are then passed to (2) a fusion module to form an aggregated embedding. This is input to the (3) predictor for classification. Training optimizes a combination of prediction, auxiliary, and CLIP losses. AD: Alzheimer's Disease; MCI: Mild cognitive impairment; NC: Normal control

encoder, while missing ones are approximated via a modality bank. However, incorporating sMRI data, which directly reflects neurodegenerative changes [18], resulted in only marginal performance gains, suggesting that their model failed to fully exploit the rich information of imaging features.

To address these limitations, we introduce MOIRA, a multimodal framework that integrates imaging, omics, and other data by projecting them into a shared representational space. This design allows modalities to complement each other during training, enabling the model to learn more robust and generalizable features. In particular, sMRI proved highly informative: beyond its substantial contribution to classification performance, it also enhanced the representation learning of other modalities by serving as a stable anatomical reference. These results highlight the central role of sMRI in improving both predictive accuracy and cross-modal alignment. The main contributions of this paper can be summarized as follows:

- We propose a framework that effectively incorporates incomplete multimodal heterogeneous data, including 3D sMRI scans and genomics.
- We achieve state-of-the-art performance in three-way AD classification while leveraging a larger portion of the ADNI database than prior models.
- We show that high-fidelity imaging features facilitate the utility of understudied modalities, with potential cost-saving application in clinical and public health settings.

2 Materials & Methods

2.1 Dataset

We utilized the ADNI dataset [16], which provides large-scale multimodal data encompassing neuroimaging, genetics, cognitive assessments, and biomarkers [19]. Following prior studies [26, 27], we categorized the data into four modalities: Imaging, Clinical, Biospecimen, and Genomics. For Imaging modality, we employed MP-RAGE and IR-FSPGR 3D T1-weighted sequences. They were preprocessed via a standard pipeline including reorientation, skull-stripping, affine registration, bias field correction, and intensity normalization. The Clinical modality was built by merging patient history from the MEDHIST, NEUROEXM, PTDEMOG, RECCMEDS, and VITALS csv files into a single tabular dataset. The Biospecimen modality integrated CSF biomarker measurements such

as amyloid-beta, total tau, and phosphorylated tau from the UPENNBIOMK_ROCHE_ELECSYS dataset with ApoE genotype data from the APOERES dataset. These biomarkers are established as highly informative AD biomarkers [21]. The Genomics modality consists of SNP (single nucleotide polymorphism) data in PLINK format from the ADNI-1, GO/2, and ADNI-3 studies.

2.2 Model

Our framework comprises three key components: **(1) modality-specific encoders** that map each input modality into a latent embedding space, **(2) a fusion module** that integrates information across modalities, and **(3) a predictor** for the final classification (Figure 1).

Table 1: Number of features and samples in the ADNI dataset.

	Imaging	**Clinical**	**Biospecimen**	**Genomics**
# features	128	1,496	147	135,595
# samples	1,651	2,380	1,744	1,596

Modality-Specific Encoders Each modality-specific encoder received inputs whose dimensions are summarized in Table 1. For the Imaging modality, sMRI scans are processed by a 3D CNN with three convolutional layers with 16, 32, and 64 channels, each followed by a LeakyReLU activation and 3D max pooling. Afterwards, it is flattened and passed through two fully connected layers with dimensions $128 \to 3$, where the 128-dimensional output serves as the image embedding. For Genomics, we used an ANOVA F-test to select the top 2,000 features to reduce dimensions. For other modalities (Clinical and Biospecimen), we directly feed the raw features into encoder networks.

Let $\mathbf{x}^{(m)} \in \mathbb{R}^{d_m}$ denote the input features from modality m, where $m \in \{1, \ldots, M\}$, d_m is the input dimension, and k is the embedding dimension. Each modality is processed by an encoder $E_\phi^{(m)}$:

$$\mathbf{z}^{(m)} = E_\phi^{(m)}\big(\mathbf{x}^{(m)}\big), \quad \mathbf{z}^{(m)} \in \mathbb{R}^k.$$

Each encoder consists of a two-layer MLP with LeakyReLU activations and dropout, and follows an autoencoder-style architecture to learn unsupervised, lower-dimensional representations of the modality [23]. Hence, each encoder was paired with a decoder and pre-trained until convergence.

Fusion Module Let $\mathbf{b} \in \{0, 1\}^M$ be an indicator vector denoting presence of each modality. We introduce a learnable weight matrix $\mathbf{W} \in \mathbb{R}^{M \times k}$ where the m-th row $\mathbf{w}^{(m)}$ corresponds to modality m. The masked weights are then computed:

$$\tilde{\mathbf{w}}^{(m)} = \mathbf{w}^{(m)} \cdot \mathbf{b}^{(m)},$$

ensuring that absent modalities have minimal contribution. We then normalize the weights across modalities using a softmax along the modality dimension:

$$\alpha^{(m)} = \frac{\exp(\tilde{\mathbf{w}}^{(m)})}{\sum_{n=1}^{M} \exp(\tilde{\mathbf{w}}^{(n)})}, \quad \alpha^{(m)} \in \mathbb{R}^k.$$

Finally, the aggregated embedding is obtained via a weighted sum of the modality-specific ones:

$$\mathbf{h} = \sum_{m=1}^{M} \alpha^{(m)} \odot \mathbf{z}^{(m)},$$

where $\odot$ denotes the Hadamard product. We ℓ_2-normalize $\mathbf{h}$ before passing it to the predictor. The fusion module allows to adaptively weight available modalities and compensate for missing ones.

Predictor The aggregated representation $\mathbf{h}$ is passed to a predictor network P_θ parameterized by θ:

$$\hat{\mathbf{y}} = P_\theta(\mathbf{h}), \quad \hat{\mathbf{y}} \in \mathbb{R}^C,$$

where C is the number of classes and $\hat{\mathbf{y}}$ denotes the predicted class probabilities. The predictor consists of a two-layer MLP with LeakyReLU activations and dropout.

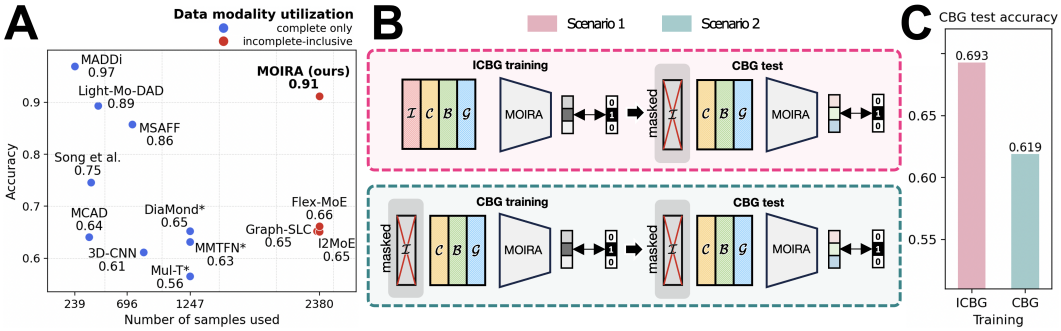

Figure 2: The evaluation results of MOIRA using the ADNI dataset. **(A)** Three-way classification accuracy versus number of samples used across multimodal learning methods. Blue points represent *complete only* models (requiring all modalities), while red points represent *incomplete-inclusive* models (handling missing modalities). The latter utilize more data during training and inference. Asterisks on DiaMond*, MMTFN*, and Mul-T* indicate that their scores reflect balanced accuracy [12], unlike others reporting standard accuracy. **(B)** Experimental setups evaluating whether imaging enhances learning from other modalities. **(C)** Accuracy comparison from Figure 2B. Model trained with ICBG outperforms the one trained with CBG, even when imaging is absent at inference time.

Loss Function The training objective combines three terms: (i) a prediction loss of cross-entropy between ground-truth labels and predictions from aggregated embedding from the fusion module; (ii) an auxiliary loss of the same cross-entropy but using each modality-specific embedding; and (iii) a contrastive loss [17] over all modality pairs to guide cross-modal alignment (Figure 1).

3 Results

A total of 2,380 subjects were used in this study, with splits of 70%, 15%, and 15% for training, validation, and test sets, respectively. We compared our model mainly with Flex-MoE [27], which is currently the strongest model in terms of handling missing modalities in ADNI. To ensure fair comparison, we aligned our data splits with those used in Flex-MoE. The embedding dimension was set to 1,000, with encoder and predictor dropout rates of 0.5 and 0.1. Models were trained for 200 epochs. All experiments were conducted using NVIDIA A40 GPUs. Each experiment was run three times with different seeds to ensure reproducibility, and the results were averaged.

MOIRA was evaluated against other three-way classification approaches on the ADNI dataset. Figure 2A summarizes accuracy and the number of samples utilized. Most prior methods are limited to patients with complete modalities, restricting their usable sample size to no more than 1,247 [3, 8, 9, 10, 12, 13, 22, 29]. By contrast, models that incorporate incomplete multimodal data extend coverage to 2,380 patients. However, such models generally report lower accuracy when using this larger cohort [15, 26, 27]. In contrast, our model achieves both broad utilization of incomplete data and high classification accuracy, without the trade-off observed in prior methods.

Table 2: Comparison of MOIRA to Flex-MoE across modality sets in Accuracy, AUC, and F1

Modals	Accuracy		AUC		F1	
	Flex-MoE	MOIRA	Flex-MoE	MOIRA	Flex-MoE	MOIRA
$\mathcal{I},\mathcal{C},\mathcal{B},\mathcal{G}$	66.11 ± 1.14	91.13 ± 0.74	81.67 ± 0.54	98.12 ± 0.07	64.73 ± 2.01	90.58 ± 0.87
$\mathcal{I},\mathcal{C},\mathcal{B}$	64.05 ± 1.78	91.78 ± 0.66	80.55 ± 1.26	98.03 ± 0.05	61.60 ± 1.46	91.03 ± 0.22
$\mathcal{I},\mathcal{C},\mathcal{G}$	63.21 ± 1.73	90.94 ± 0.13	79.55 ± 1.69	98.00 ± 0.13	61.98 ± 1.04	90.12 ± 0.19
$\mathcal{I},\mathcal{B},\mathcal{G}$	62.28 ± 2.75	91.60 ± 0.23	79.27 ± 0.65	98.13 ± 0.23	59.45 ± 3.14	91.00 ± 0.22
$\mathcal{C},\mathcal{B},\mathcal{G}$	65.36 ± 1.38	71.34 ± 0.92	81.67 ± 0.59	87.20 ± 0.25	64.15 ± 1.69	69.34 ± 0.97

* $\mathcal{I}$: Imaging; $\mathcal{C}$: Clinical; $\mathcal{B}$: Biospecimen; $\mathcal{G}$: Genomics
** Results are reported as mean $\pm$ std values.

To assess the contribution of individual modalities, we conducted ablation studies and compared our model with Flex-MoE. Table 2 shows that the performance of Flex-MoE decreased only by 1.1% without Imaging modality, suggesting limited use of sMRI data. In contrast, our model exhibited a 21.7% drop without the Imaging modality, signifying MOIRA's utilization of sMRI in AD prediction.

We further replaced the 3D CNN–extracted sMRI embeddings with descriptors derived from the UCSF Cross-Sectional FreeSurfer (7.x) dataset [ADNI1, GO, 2, 3, 4]. The data contains cortical thickness, volumetric measurement, and other anatomical information summarized using the FreeSurfer software [6]. As shown in Table 3, this led to a substantial drop in predictive performance compared to models using the original 3D CNN embeddings. The finding suggests that sMRI data is more effective through approaches that fully preserve and use its rich spatial and morphological structure [2].

Table 3: Classification performance using FreeSurfer dataset (w/o CNN) and sMRI (w/ CNN).

Metric	**MOIRA (w/o CNN)**	**MOIRA (w/ CNN)**
ACC	72.36 ± 1.13	91.13 ± 0.74
AUC	88.33 ± 1.28	98.12 ± 0.07

Beyond imaging ($\mathcal{I}$), ablation of other modalities led to only marginal changes in classification performance (Table 2), raising the question of whether the model relies predominantly on sMRI. To test this, we compared two settings: (1) training with all four modalities but masking $\mathcal{I}$ during inference, and (2) using only $\mathcal{C}$, $\mathcal{B}$, and $\mathcal{G}$ throughout training, validation, and testing (Figure 2B). If predictions depended solely on $\mathcal{I}$, then masking $\mathcal{I}$ at inference (Scenario 1) would at best yield similar performance to the $\mathcal{CBG}$-only baseline (Scenario 2). However, Figure 2C shows that Scenario 1 outperformed Scenario 2 by 10.7% in accuracy, despite having no access to imaging at inference. This indicates that incorporating sMRI during training improves the model's ability to capture patterns in other modalities. The limited gains observed from Table 2 when including non-imaging modalities are not due to an absence of informative signals, but to the prevailing predictive strength of imaging data. Clinically, this suggests that the model trained with imaging modality retains predictive power even when future patients have no sMRI data, thereby enhancing its applicability in real-world prognostic settings where incomplete data modalities are common.

4 Conclusion

Integrating heterogeneous multimodal data is crucial for deepening our understanding of complex diseases such as AD, as it enables a systems-level perspective that encompasses both molecular and structural pathology. However, this integration remains challenging due to the frequent absence of certain modalities and the difficulty of harmonizing diverse data types [14].

In this study, we present MOIRA, a multimodal integration framework that handles diverse modalities, including sMRI and various non-imaging data. The fusion module aligns features extracted from modality-specific encoders in a shared embedding space. Hence, unlike models that rely on samples with complete modalities, MOIRA can flexibly leverage incomplete datasets.

Furthermore, we showed that image embeddings provide high-fidelity anatomical information and serve a dual role: contributing directly to inference and guiding the integration of less informative or missing modalities. This supervisory role enhances performance even when imaging data are absent at inference. Our finding that incorporating sMRI data during training can improve prediction for patients missing them is particularly relevant since key AD biomarkers such as amyloid-beta, tau, or clinical symptoms may precede the onset of overt disease by over a decade in some cases [1, 20]. This has important implications for pre-emptive intervention, which are increasingly emphasized in emerging AD treatment strategies [7]. Beyond AD, MOIRA generalizes to other conditions where imaging and omics data are available, highlighting its potential as an extensible framework for multimodal disease modeling.

In future work, we will improve the integration of modalities such as genotypes and CSF data, investigate the underlying mechanisms by which sMRI contributes to the $\mathcal{CBG}$-only inference (Figure 2B), and conduct feature importance analyses to identify biomarkers for AD diagnosis. We hope our work inspire further research into incomplete-inclusive learning strategies and contribute to the development of robust diagnostic models for AD and related neurodegenerative disorders.

References

[1] Hélène Amieva, Mélanie Le Goff, Xavier Millet, Jean Marc Orgogozo, Karine Pérès, Pascale Barberger-Gateau, Hélène Jacqmin-Gadda, and Jean François Dartigues. Prodromal alzheimer's disease: successive emergence of the clinical symptoms. *Annals of Neurology: Official Journal of the American Neurological Association and the Child Neurology Society*, 64(5):492–498, 2008.

[2] Ida Arvidsson, Olof Strandberg, Sebastian Palmqvist, Erik Stomrud, Nicholas Cullen, Shorena Janelidze, Pontus Tideman, Anders Heyden, Karl Åström, Oskar Hansson, et al. Comparing a pre-defined versus deep learning approach for extracting brain atrophy patterns to predict cognitive decline due to alzheimer's disease in patients with mild cognitive symptoms. *Alzheimer's Research & Therapy*, 16(1):61, 2024.

[3] Xia-An Bi, Wenzhuo Shen, Yinglu Shan, Dayou Chen, Luyun Xu, Ke Chen, and Zhonghua Liu. Msaff: Multi-way soft attention fusion framework with the large foundation models for the diagnosis of alzheimer's disease. *IEEE Transactions on Neural Networks and Learning Systems*, 2025.

[4] Zeinab Breijyeh and Rafik Karaman. Comprehensive review on alzheimer's disease: causes and treatment. *Molecules*, 25(24):5789, 2020.

[5] Bruno Dubois, Christine AF von Arnim, Nerida Burnie, Sasha Bozeat, and Jeffrey Cummings. Biomarkers in alzheimer's disease: role in early and differential diagnosis and recognition of atypical variants. *Alzheimer's Research & Therapy*, 15(1):175, 2023.

[6] Bruce Fischl. Freesurfer. *Neuroimage*, 62(2):774–781, 2012.

[7] Kristian Steen Frederiksen, Xavier Morato Arus, Henrik Zetterberg, Serge Gauthier, Mercè Boada, Vanesa Pytel, Julie Hahn-Pedersen, Luis Rafael Solís Tarazona, and Soeren Mattke. Focusing on earlier diagnosis of alzheimer's disease. *Future Neurology*, 19(1):2337452, 2024.

[8] Xingyu Gao, Feng Shi, Dinggang Shen, and Manhua Liu. Multimodal transformer network for incomplete image generation and diagnosis of alzheimer's disease. *Computerized Medical Imaging and Graphics*, 110:102303, 2023.

[9] Michal Golovanevsky, Carsten Eickhoff, and Ritambhara Singh. Multimodal attention-based deep learning for alzheimer's disease diagnosis. *Journal of the American Medical Informatics Association*, 29(12):2014–2022, 2022.

[10] Bhoomi Gupta, Ganesh Kanna Jegannathan, Mohammad Shabbir Alam, Kottala Sri Yogi, Janjhyam Venkata Naga Ramesh, Vemula Jasmine Sowmya, and Isa Bayhan. Multimodal lightweight neural network for alzheimer's disease diagnosis integrating neuroimaging and cognitive scores. *Neuroscience Informatics*, page 100218, 2025.

[11] Lanlan Li, Xianfeng Yu, Can Sheng, Xueyan Jiang, Qi Zhang, Ying Han, and Jiehui Jiang. A review of brain imaging biomarker genomics in alzheimer's disease: implementation and perspectives. *Translational Neurodegeneration*, 11(1):42, 2022.

[12] Yitong Li, Morteza Ghahremani, Youssef Wally, and Christian Wachinger. Diamond: Dementia diagnosis with multi-modal vision transformers using mri and pet. In *2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 107–116. IEEE, 2025.

[13] Shang Miao, Qun Xu, Weimin Li, Chao Yang, Bin Sheng, Fangyu Liu, Tsigabu T Bezabih, and Xiao Yu. Mmtfn: Multi-modal multi-scale transformer fusion network for alzheimer's disease diagnosis. *International Journal of Imaging Systems and Technology*, 34(1):e22970, 2024.

[14] Minsik Oh, Sungjoon Park, Sun Kim, and Heejoon Chae. Machine learning-based analysis of multi-omics data on the cloud for investigating gene regulations. *Briefings in bioinformatics*, 22(1):66–76, 2021.

[15] Zaixin Ou, Caiwen Jiang, Yuxiao Liu, Yuanwang Zhang, Zhiming Cui, and Dinggang Shen. A graph-embedded latent space learning and clustering framework for incomplete multimodal multiclass alzheimer's disease diagnosis. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 45–55. Springer, 2024.

[16] Ronald Carl Petersen, Paul S Aisen, Laurel A Beckett, Michael C Donohue, Anthony Collins Gamst, Danielle J Harvey, CR Jack Jr, William J Jagust, Leslie M Shaw, Arthur W Toga, et al. Alzheimer's disease neuroimaging initiative (adni) clinical characterization. *Neurology*, 74(3):201–209, 2010.

[17] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever. Learning transferable visual models from natural language supervision, 2021.

[18] Shannon L Risacher and Andrew J Saykin. Neuroimaging biomarkers of neurodegenerative diseases and dementia. In *Seminars in neurology*, volume 33, pages 386–416. Thieme Medical Publishers, 2013.

[19] Andrew J Saykin, Li Shen, Xiaohui Yao, Sungeun Kim, Kwangsik Nho, Shannon L Risacher, Vijay K Ramanan, Tatiana M Foroud, Kelley M Faber, Nadeem Sarwar, et al. Genetic studies of quantitative mci and ad phenotypes in adni: Progress, opportunities, and plans. *Alzheimer's & Dementia*, 11(7):792–814, 2015.

[20] Dennis J Selkoe and John Hardy. The amyloid hypothesis of alzheimer's disease at 25 years. *EMBO molecular medicine*, 8(6):595–608, 2016.

[21] Leslie M Shaw, Hugo Vanderstichele, Malgorzata Knapik-Czajka, Christopher M Clark, Paul S Aisen, Ronald C Petersen, Kaj Blennow, Holly Soares, Adam Simon, Piotr Lewczuk, et al. Cerebrospinal fluid biomarker signature in alzheimer's disease neuroimaging initiative subjects. *Annals of neurology*, 65(4):403–413, 2009.

[22] Juan Song, Jian Zheng, Ping Li, Xiaoyuan Lu, Guangming Zhu, and Peiyi Shen. An effective multimodal image fusion method using mri and pet for alzheimer's disease diagnosis. *Frontiers in digital health*, 3:637386, 2021.

[23] Dustin Van Der Haar, Ahmed Moustafa, Samuel L Warren, Hany Alashwal, and Terence van Zyl. An alzheimer's disease category progression sub-grouping analysis using manifold learning on adni. *Scientific reports*, 13(1):10483, 2023.

[24] Jingru Wang, Shipeng Wen, Wenjie Liu, Xianglian Meng, and Zhuqing Jiao. Deep joint learning diagnosis of alzheimer's disease based on multimodal feature fusion. *BioData Mining*, 17(1):48, 2024.

[25] Ming-Yu Wang, Ke-Liang Chen, Yu-Yuan Huang, Shu-Fen Chen, Rong-Ze Wang, Yi Zhang, He-Ying Hu, Ling-Zhi Ma, Wei-Shi Liu, Jun Wang, et al. Clinical utility of cerebrospinal fluid alzheimer's disease biomarkers in the diagnostic workup of complex patients with cognitive impairment. *Translational Psychiatry*, 15(1):130, 2025.

[26] Jiayi Xin, Sukwon Yun, Jie Peng, Inyoung Choi, Jenna L Ballard, Tianlong Chen, and Qi Long. I2moe: Interpretable multimodal interaction-aware mixture-of-experts. *arXiv preprint arXiv:2505.19190*, 2025.

[27] Sukwon Yun, Inyoung Choi, Jie Peng, Yangfan Wu, Jingxuan Bao, Qiyiwen Zhang, Jiayi Xin, Qi Long, and Tianlong Chen. Flex-moe: Modeling arbitrary modality combination via the flexible mixture-of-experts. *Advances in Neural Information Processing Systems*, 37:98782–98805, 2024.

[28] Jifa Zhang, Yinglu Zhang, Jiaxing Wang, Yilin Xia, Jiaxian Zhang, and Lei Chen. Recent advances in alzheimer's disease: Mechanisms, clinical trials and new drug development strategies. *Signal transduction and targeted therapy*, 9(1):211, 2024.

[29] Jin Zhang, Xiaohai He, Yan Liu, Qingyan Cai, Honggang Chen, and Linbo Qing. Multi-modal cross-attention network for alzheimer's disease diagnosis with multi-modality data. *Computers in biology and medicine*, 162:107050, 2023.