

# Hierarchical Information Aggregation for Incomplete Multimodal Alzheimer’s Disease Diagnosis

Chengliang Liu<sup>1,2</sup>, Yuanxi Que<sup>1</sup>, Qihao Xu<sup>3</sup>, Yabo Liu<sup>4</sup>,  
Jie Wen<sup>3</sup>, Jinghua Wang<sup>3</sup>, Xiaoling Luo<sup>1\*</sup>

<sup>1</sup>College of Computer Science and Software Engineering, Shenzhen University

<sup>2</sup>Laboratory for Artificial Intelligence in Design, The Hong Kong Polytechnic University

<sup>3</sup>School of Computer Science and Technology, Harbin Institute of Technology, Shenzhen

<sup>4</sup>College of Artificial Intelligence, Ocean University of China

liucl1996@163.com, {queyuanxi, xqh51199597, xiaolingluo}@outlook.com,

yaboliu.ug@gmail.com, jiewen\_pr@126.com, wangjh2012@foxmail.com

## 1 Implementation Details

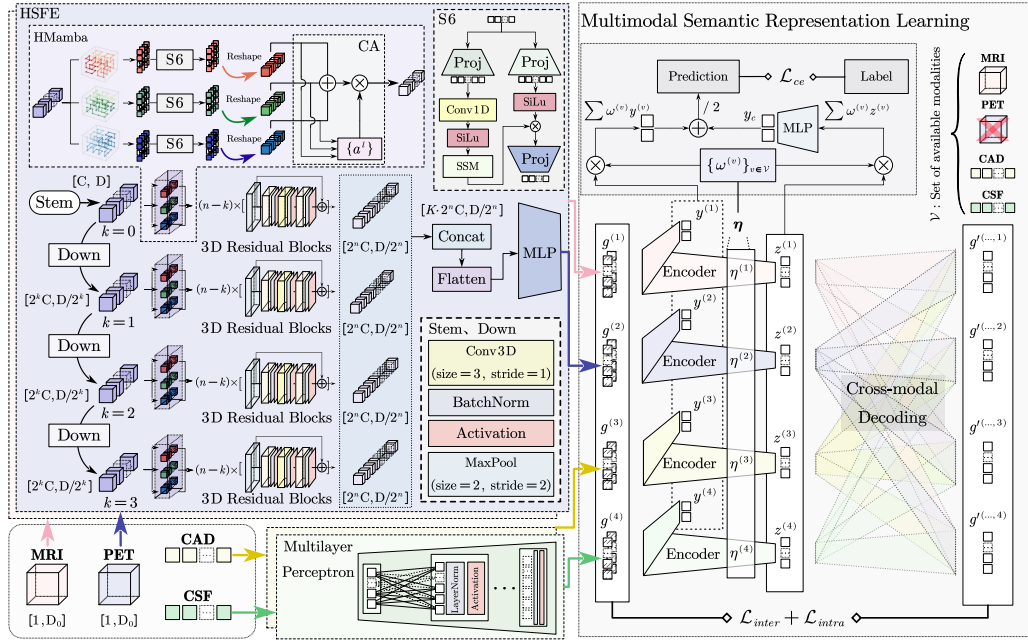


Figure 1: The schematic diagram of the architecture details of our HAD. “Stem” represents the standardized 3D image processing process. In our experiments, the number of channel  $C = 8$  and dimension of  $D = 64$ .

The overall network architecture is presented in Fig. 1. Original MRI and PET data with spatial dimension  $D_0 = 128$  is processed by the Stem module, expanding the number of channels  $C$  to 8 and reducing the spatial dimension  $D$  to 64 (“[ $C, D$ ]” in Fig. 1 means the dimension ( $C, D, D, D$ )). Our proposed HSFE module consists of four hierarchical levels ( $K = 4$ ), processing input images at progressively reduced resolutions of  $(64, 64, 64)$ ,  $(32, 32, 32)$ ,  $(16, 16, 16)$ , and  $(8, 8, 8)$ . At each

\*Corresponding author: Xiaoling Luo (email: xiaolingluo@outlook.com).

hierarchical level, we utilize our designed Mamba blocks to effectively capture spatial information. Specifically, each Mamba block adopts a multi-view scanning strategy with three distinct scanning views. Moreover, for each scanning direction, both forward and reverse scanning sequences are employed, resulting in a total of six scanning sequences per block. We implement our HAD using Python 3.9 and train the model with the Adam optimizer, setting an initial learning rate of 0.001. For all methods, we take the optimal result based on the validation set within 100 training epochs. All experiments are conducted on an Ubuntu 20.04 computing platform equipped with multiple NVIDIA RTX 4090 GPUs. The code is available at <https://github.com/YuanxiQue/HAD>.

## 2 Experimental Settings and Results

As mentioned above, not all existing methods are designed for incomplete heterogeneous multimodal AD diagnosis, so we need to make necessary modifications to them to meet the requirements of our task. These methods are modified as follows: (1) **LMVCAT**: we add ResNet-50 backbone for 3D image feature extraction and modify the multi-label classifier to a single-label classifier. (2) **Adapted**: It is combined with the LMVCAT as a plug-and-play parameter-efficient adaptation method. (3) **DMRNet**: we add ResNet-50 backbone for 3D image feature extraction. (4) **ShaSpec**: we replace the segmentation head with the classification head and add MLP for extracting CSF and CAD features. (5) **GMD**: we add ResNet-50 backbone for 3D image feature extraction. (6) **CM3T**: we replace the input demographic data with our CSF and PET data. (7) **TriMF**: we replace the input data to our four modalities and add ResNet-50 backbone for 3D image feature extraction. In Table 2 and 4, we fully present the experimental results of eight methods on three tasks with different missing rates.

To study the single-modality performance, we conduct additional experiments on the ADNI dataset (without missing modalities) for the MCI vs. CN classification task using each modality individually (MRI, PET, CAD, CSF), as well as our proposed HAD method in Table 5.

To provide more quantitative evidence, we conducted a new experiment: for each modality, we extracted the modality-specific features after training, fed them independently into the same classifier, and measured their classification performance. This analysis was performed under two training schemes: (a) mid-fusion only, and (b) our proposed hybrid late-fusion strategy.

## 3 Study on Execution Efficiency

To evaluate the execution efficiency, we have conducted a comprehensive comparison of model parameters (Params, in millions) and computational cost (FLOPs, in GigaFLOPs) with representative state-of-the-art methods. The results are summarized in Table 1:

Table 1: Model size and computational cost.

Model	Params (M)	FLOPs (G)
LMVCAT	97.33	81.43
CM3T	10.41	147.96
GMD	94.3	81.56
DMRNet	101.79	81.56
Adapted	97.33	81.43
ShaSpec	0.97	5.64
<b>Ours</b>	43.42	6.65

## 4 Hyperparameter Sensitivity Analysis

We perform sensitivity analyses to investigate the influence of key hyperparameters in our proposed method, namely the intra-modal reconstruction parameter ( $\lambda$ ), cross-modal reconstruction parameter ( $\gamma$ ), and the temperature parameter ( $\tau$ ). Specifically, we conduct a thorough grid search for  $\lambda$  and  $\gamma$ , across three classification tasks (AD vs. CN, AD vs. MCI, and MCI vs. CN) under a modality missing rate of 50%. As illustrated in Fig. 3, the optimal values of  $\lambda$  and  $\gamma$  differ substantially across

Table 2: Comparison of eight methods on five metrics under missing rate of 10%. The standard deviation is after the sign ' $\pm$ '.

Task	Metric	LMVCAT [1] AAAI 23	Adapted[2] TPAMI 24	DMRNet[3] ECCV 24	ShaSpec[4] CVPR 23	GMD[5] AAAI 24	CM3T[6] NeuroImage 23	TriMF[7] J. B. Infor 25	Ours
AD vs. CN	AUC	0.994 $\pm$ 0.003	0.994 $\pm$ 0.002	0.993 $\pm$ 0.004	0.991 $\pm$ 0.007	0.993 $\pm$ 0.008	0.995 $\pm$ 0.004	0.995 $\pm$ 0.004	<b>0.995<math>\pm</math>0.002</b>
	ACC	0.969 $\pm$ 0.010	0.969 $\pm$ 0.012	0.966 $\pm$ 0.014	0.970 $\pm$ 0.010	0.973 $\pm$ 0.011	0.972 $\pm$ 0.009	0.969 $\pm$ 0.013	<b>0.975<math>\pm</math>0.006</b>
	F1	0.944 $\pm$ 0.018	0.942 $\pm$ 0.021	0.937 $\pm$ 0.025	0.945 $\pm$ 0.017	0.951 $\pm$ 0.020	0.948 $\pm$ 0.016	0.942 $\pm$ 0.023	<b>0.953<math>\pm</math>0.011</b>
	SEN	0.926 $\pm$ 0.038	0.928 $\pm$ 0.033	0.918 $\pm$ 0.043	0.923 $\pm$ 0.041	0.951 $\pm$ 0.034	0.937 $\pm$ 0.028	0.915 $\pm$ 0.040	<b>0.954<math>\pm</math>0.030</b>
	SPE	0.986 $\pm$ 0.006	0.984 $\pm$ 0.011	0.984 $\pm$ 0.010	0.988 $\pm$ 0.010	0.982 $\pm$ 0.009	0.985 $\pm$ 0.006	<b>0.989<math>\pm</math>0.010</b>	0.982 $\pm$ 0.010
AD vs. MCI	AUC	0.936 $\pm$ 0.025	0.948 $\pm$ 0.011	0.926 $\pm$ 0.024	0.935 $\pm$ 0.016	0.938 $\pm$ 0.015	0.942 $\pm$ 0.017	<b>0.957<math>\pm</math>0.014</b>	0.953 $\pm$ 0.016
	ACC	0.892 $\pm$ 0.018	0.901 $\pm$ 0.013	0.881 $\pm$ 0.020	0.889 $\pm$ 0.024	0.889 $\pm$ 0.019	0.890 $\pm$ 0.021	0.897 $\pm$ 0.024	<b>0.901<math>\pm</math>0.025</b>
	F1	0.783 $\pm$ 0.050	<b>0.803<math>\pm</math>0.021</b>	0.767 $\pm$ 0.036	0.783 $\pm$ 0.038	0.776 $\pm$ 0.034	0.785 $\pm$ 0.035	0.778 $\pm$ 0.069	0.789 $\pm$ 0.046
	SEN	0.772 $\pm$ 0.118	<b>0.783<math>\pm</math>0.041</b>	0.763 $\pm$ 0.067	0.775 $\pm$ 0.048	0.747 $\pm$ 0.048	0.782 $\pm$ 0.056	0.721 $\pm$ 0.128	0.723 $\pm$ 0.067
	SPE	0.935 $\pm$ 0.022	0.942 $\pm$ 0.017	0.922 $\pm$ 0.023	0.928 $\pm$ 0.028	0.938 $\pm$ 0.021	0.928 $\pm$ 0.026	0.959 $\pm$ 0.018	<b>0.963<math>\pm</math>0.014</b>
MCI vs. CN	AUC	0.921 $\pm$ 0.019	0.927 $\pm$ 0.015	0.930 $\pm$ 0.011	0.924 $\pm$ 0.012	0.924 $\pm$ 0.013	0.920 $\pm$ 0.012	0.926 $\pm$ 0.008	<b>0.930<math>\pm</math>0.009</b>
	ACC	0.852 $\pm$ 0.019	<b>0.857<math>\pm</math>0.011</b>	0.852 $\pm$ 0.018	0.847 $\pm$ 0.020	0.850 $\pm$ 0.020	0.848 $\pm$ 0.011	0.853 $\pm$ 0.017	0.851 $\pm$ 0.022
	F1	0.855 $\pm$ 0.024	<b>0.863<math>\pm</math>0.012</b>	0.854 $\pm$ 0.022	0.851 $\pm$ 0.024	0.852 $\pm$ 0.021	0.852 $\pm$ 0.014	0.861 $\pm$ 0.026	0.848 $\pm$ 0.034
	SEN	0.838 $\pm$ 0.056	0.868 $\pm$ 0.039	0.836 $\pm$ 0.051	0.839 $\pm$ 0.034	0.827 $\pm$ 0.036	0.843 $\pm$ 0.041	<b>0.881<math>\pm</math>0.056</b>	0.810 $\pm$ 0.063
	SPE	0.871 $\pm$ 0.037	0.840 $\pm$ 0.059	0.875 $\pm$ 0.046	0.858 $\pm$ 0.047	0.878 $\pm$ 0.044	0.854 $\pm$ 0.054	0.821 $\pm$ 0.033	<b>0.894<math>\pm</math>0.037</b>

Table 3: Comparison of eight methods on five metrics under missing rate of 30%. The standard deviation is after the sign ' $\pm$ '.

Task	Metric	LMVCAT [1] AAAI 23	Adapted[2] TPAMI 24	DMRNet[3] ECCV 24	ShaSpec[4] CVPR 23	GMD[5] AAAI 24	CM3T[6] NeuroImage 23	TriMF[7] J. B. Infor 25	Ours
AD vs. CN	AUC	0.968 $\pm$ 0.017	0.971 $\pm$ 0.013	0.973 $\pm$ 0.014	0.976 $\pm$ 0.008	0.969 $\pm$ 0.013	0.968 $\pm$ 0.013	0.966 $\pm$ 0.013	<b>0.982<math>\pm</math>0.010</b>
	ACC	0.930 $\pm$ 0.023	<b>0.937<math>\pm</math>0.016</b>	0.933 $\pm$ 0.021	0.929 $\pm$ 0.019	0.928 $\pm$ 0.024	0.924 $\pm$ 0.015	0.919 $\pm$ 0.025	0.932 $\pm$ 0.013
	F1	0.862 $\pm$ 0.049	<b>0.878<math>\pm</math>0.034</b>	0.864 $\pm$ 0.048	0.860 $\pm$ 0.039	0.859 $\pm$ 0.046	0.847 $\pm$ 0.038	0.834 $\pm$ 0.055	0.865 $\pm$ 0.031
	SEN	0.799 $\pm$ 0.080	<b>0.824<math>\pm</math>0.052</b>	0.792 $\pm$ 0.095	0.799 $\pm$ 0.068	0.803 $\pm$ 0.078	0.781 $\pm$ 0.084	0.749 $\pm$ 0.068	0.802 $\pm$ 0.038
	SPE	0.980 $\pm$ 0.017	0.980 $\pm$ 0.010	<b>0.985<math>\pm</math>0.025</b>	0.978 $\pm$ 0.024	0.975 $\pm$ 0.025	0.977 $\pm$ 0.016	0.983 $\pm$ 0.016	0.981 $\pm$ 0.011
AD vs. MCI	AUC	0.891 $\pm$ 0.021	0.899 $\pm$ 0.022	0.885 $\pm$ 0.034	0.890 $\pm$ 0.022	0.895 $\pm$ 0.018	0.900 $\pm$ 0.015	0.891 $\pm$ 0.028	<b>0.915<math>\pm</math>0.014</b>
	ACC	0.842 $\pm$ 0.029	0.845 $\pm$ 0.025	0.847 $\pm$ 0.040	0.841 $\pm$ 0.027	0.843 $\pm$ 0.024	0.841 $\pm$ 0.017	0.846 $\pm$ 0.032	<b>0.857<math>\pm</math>0.010</b>
	F1	0.647 $\pm$ 0.055	0.676 $\pm$ 0.047	0.636 $\pm$ 0.105	0.640 $\pm$ 0.086	0.661 $\pm$ 0.056	<b>0.680<math>\pm</math>0.047</b>	0.618 $\pm$ 0.087	0.672 $\pm$ 0.030
	SEN	0.564 $\pm$ 0.065	0.630 $\pm$ 0.057	0.532 $\pm$ 0.125	0.564 $\pm$ 0.131	0.603 $\pm$ 0.102	<b>0.673<math>\pm</math>0.155</b>	0.491 $\pm$ 0.106	0.587 $\pm$ 0.060
	SPE	0.939 $\pm$ 0.031	0.919 $\pm$ 0.033	0.959 $\pm$ 0.021	0.939 $\pm$ 0.018	0.927 $\pm$ 0.024	0.902 $\pm$ 0.057	<b>0.970<math>\pm</math>0.017</b>	0.947 $\pm$ 0.021
MCI vs. CN	AUC	0.880 $\pm$ 0.012	0.867 $\pm$ 0.024	0.869 $\pm$ 0.019	0.878 $\pm$ 0.016	0.874 $\pm$ 0.011	0.866 $\pm$ 0.013	0.867 $\pm$ 0.018	<b>0.884<math>\pm</math>0.014</b>
	ACC	0.784 $\pm$ 0.013	0.754 $\pm$ 0.035	0.777 $\pm$ 0.030	0.780 $\pm$ 0.026	0.777 $\pm$ 0.016	0.745 $\pm$ 0.024	0.763 $\pm$ 0.031	<b>0.794<math>\pm</math>0.024</b>
	F1	0.794 $\pm$ 0.022	0.761 $\pm$ 0.047	0.786 $\pm$ 0.048	0.783 $\pm$ 0.035	0.782 $\pm$ 0.029	0.742 $\pm$ 0.053	0.781 $\pm$ 0.041	<b>0.810<math>\pm</math>0.019</b>
	SEN	0.806 $\pm$ 0.085	0.761 $\pm$ 0.116	0.804 $\pm$ 0.120	0.767 $\pm$ 0.074	0.777 $\pm$ 0.105	0.722 $\pm$ 0.120	<b>0.824<math>\pm</math>0.104</b>	0.817 $\pm$ 0.065
	SPE	0.761 $\pm$ 0.106	0.755 $\pm$ 0.133	0.745 $\pm$ 0.099	<b>0.799<math>\pm</math>0.099</b>	0.782 $\pm$ 0.130	0.763 $\pm$ 0.113	0.700 $\pm$ 0.101	0.779 $\pm$ 0.110

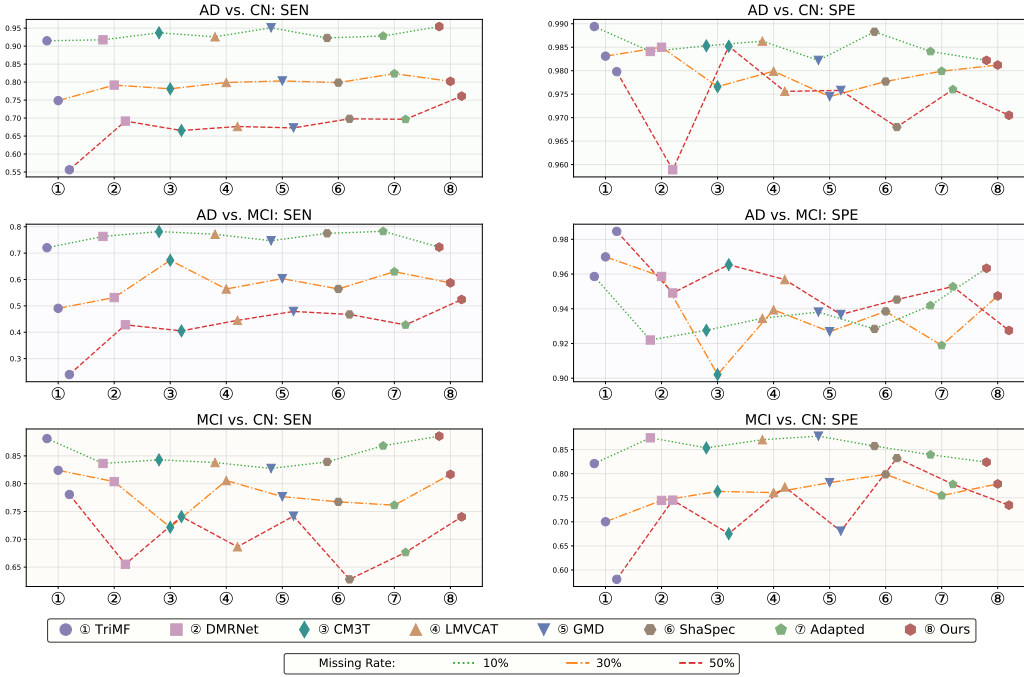


Figure 2: The comparison results of eight methods on three tasks with different missing rates.

Table 4: Comparison of eight methods on five metrics under missing rate of 50%. The standard deviation is after the sign ‘ $\pm$ ’.

Task	Metric	LMVCAT [1] AAAI 23	Adapted[2] TPAMI 24	DMRNet[3] ECCV 24	ShaSpec[4] CVPR 23	GMD[5] AAAI 24	CM3T[6] NeuroImage 23	TriMF[7] J. B. Infor 25	Ours
AD vs. CN	AUC	0.939 $\pm$ 0.018	0.952 $\pm$ 0.015	0.935 $\pm$ 0.021	0.948 $\pm$ 0.012	0.944 $\pm$ 0.015	0.943 $\pm$ 0.014	0.892 $\pm$ 0.013	<b>0.967<math>\pm</math>0.006</b>
	ACC	0.894 $\pm$ 0.016	0.899 $\pm$ 0.011	0.886 $\pm$ 0.025	0.895 $\pm$ 0.016	0.892 $\pm$ 0.019	0.898 $\pm$ 0.009	0.864 $\pm$ 0.009	<b>0.913<math>\pm</math>0.005</b>
	F1	0.778 $\pm$ 0.032	0.791 $\pm$ 0.033	0.766 $\pm$ 0.070	0.783 $\pm$ 0.050	0.773 $\pm$ 0.049	0.781 $\pm$ 0.024	0.690 $\pm$ 0.047	<b>0.827<math>\pm</math>0.018</b>
	SEN	0.677 $\pm$ 0.039	0.697 $\pm$ 0.026	0.692 $\pm$ 0.105	0.698 $\pm$ 0.098	0.673 $\pm$ 0.076	0.665 $\pm$ 0.035	0.556 $\pm$ 0.071	<b>0.761<math>\pm</math>0.040</b>
AD vs. MCI	SPE	0.976 $\pm$ 0.014	0.976 $\pm$ 0.013	0.959 $\pm$ 0.031	0.968 $\pm$ 0.032	0.976 $\pm$ 0.019	<b>0.985<math>\pm</math>0.009</b>	0.980 $\pm$ 0.018	0.971 $\pm$ 0.014
	AUC	0.854 $\pm$ 0.017	0.857 $\pm$ 0.031	0.843 $\pm$ 0.019	0.865 $\pm$ 0.019	0.861 $\pm$ 0.022	0.839 $\pm$ 0.050	0.791 $\pm$ 0.035	<b>0.876<math>\pm</math>0.021</b>
	ACC	<b>0.824<math>\pm</math>0.021</b>	0.817 $\pm$ 0.025	0.814 $\pm$ 0.021	0.821 $\pm$ 0.017	0.817 $\pm$ 0.029	0.820 $\pm$ 0.018	0.792 $\pm$ 0.028	0.823 $\pm$ 0.024
	F1	0.562 $\pm$ 0.062	0.542 $\pm$ 0.071	0.539 $\pm$ 0.048	0.569 $\pm$ 0.057	0.564 $\pm$ 0.104	0.534 $\pm$ 0.056	0.364 $\pm$ 0.112	<b>0.602<math>\pm</math>0.044</b>
MCI vs. CN	SEN	0.445 $\pm$ 0.092	0.428 $\pm$ 0.095	0.428 $\pm$ 0.088	0.467 $\pm$ 0.097	0.479 $\pm$ 0.136	0.405 $\pm$ 0.080	0.240 $\pm$ 0.096	<b>0.524<math>\pm</math>0.086</b>
	SPE	0.957 $\pm$ 0.028	0.953 $\pm$ 0.016	0.949 $\pm$ 0.043	0.945 $\pm$ 0.031	0.937 $\pm$ 0.026	0.965 $\pm$ 0.020	<b>0.985<math>\pm</math>0.014</b>	0.928 $\pm$ 0.032
	AUC	0.823 $\pm$ 0.022	0.818 $\pm$ 0.029	0.798 $\pm$ 0.028	0.826 $\pm$ 0.020	0.813 $\pm$ 0.029	0.814 $\pm$ 0.021	0.785 $\pm$ 0.016	<b>0.838<math>\pm</math>0.018</b>
	ACC	0.728 $\pm$ 0.013	0.722 $\pm$ 0.027	0.689 $\pm$ 0.038	0.726 $\pm$ 0.021	0.708 $\pm$ 0.032	0.701 $\pm$ 0.032	0.681 $\pm$ 0.016	<b>0.732<math>\pm</math>0.028</b>
	F1	0.721 $\pm$ 0.040	0.713 $\pm$ 0.046	0.689 $\pm$ 0.046	0.705 $\pm$ 0.017	0.725 $\pm$ 0.030	0.715 $\pm$ 0.052	0.717 $\pm$ 0.030	<b>0.740<math>\pm</math>0.032</b>
	SEN	0.687 $\pm$ 0.105	0.677 $\pm$ 0.129	0.655 $\pm$ 0.144	0.628 $\pm$ 0.021	0.742 $\pm$ 0.082	0.741 $\pm$ 0.175	<b>0.781<math>\pm</math>0.139</b>	0.740 $\pm$ 0.116
	SPE	0.773 $\pm$ 0.095	0.778 $\pm$ 0.111	0.745 $\pm$ 0.184	<b>0.832<math>\pm</math>0.018</b>	0.681 $\pm$ 0.096	0.675 $\pm$ 0.195	0.581 $\pm$ 0.158	0.735 $\pm$ 0.120

Table 5: Performance comparison across modalities. The standard deviation follows the sign “ $\pm$ ”.

Metric	MRI	PET	CAD	CSF	Ours (HAD)
<b>AUC</b>	0.731 $\pm$ 0.037	0.702 $\pm$ 0.024	0.946 $\pm$ 0.013	0.648 $\pm$ 0.036	<b>0.960<math>\pm</math>0.013</b>
<b>ACC</b>	0.700 $\pm$ 0.026	0.658 $\pm$ 0.015	0.873 $\pm$ 0.013	0.616 $\pm$ 0.054	<b>0.903<math>\pm</math>0.016</b>
<b>F1</b>	0.759 $\pm$ 0.038	0.723 $\pm$ 0.038	0.878 $\pm$ 0.015	0.671 $\pm$ 0.070	<b>0.917<math>\pm</math>0.016</b>
<b>SEN</b>	0.819 $\pm$ 0.069	0.750 $\pm$ 0.124	0.879 $\pm$ 0.019	0.689 $\pm$ 0.115	<b>0.925<math>\pm</math>0.027</b>
<b>SPE</b>	0.528 $\pm$ 0.065	0.528 $\pm$ 0.177	0.868 $\pm$ 0.042	0.508 $\pm$ 0.145	<b>0.874<math>\pm</math>0.023</b>

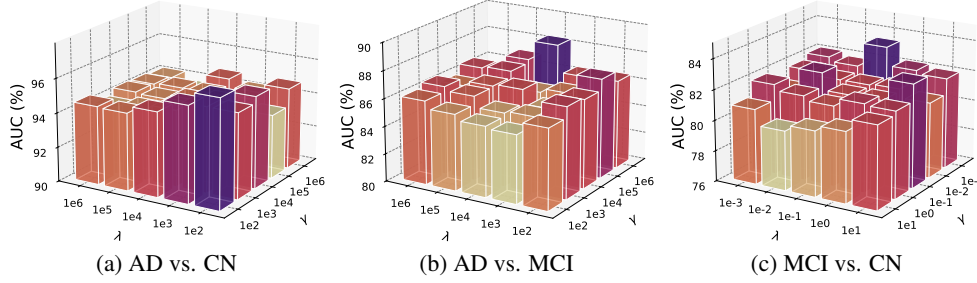


Figure 3: Hyperparameter sensitivity analysis on parameters  $\lambda$  and  $\gamma$  on three tasks with missing rate of 50%.

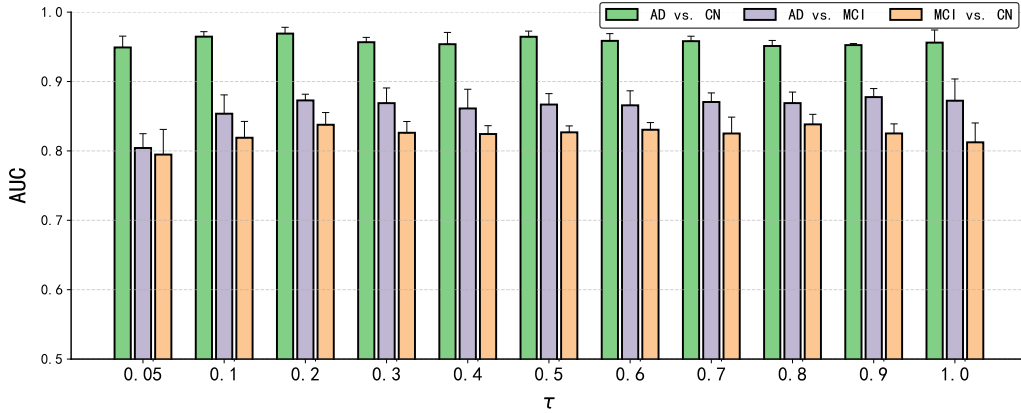


Figure 4: Hyperparameter sensitivity analysis on parameters  $\tau$  on three tasks with missing rate of 50%.

these tasks, indicating task-specific sensitivity. The best-performing hyperparameter pairs  $(\lambda, \gamma)$  are found to be  $(1e2, 1e2)$  for the AD vs. CN task,  $(1e4, 1e6)$  for the AD vs. MCI task, and  $(1e - 1, 1e - 3)$  for the MCI vs. CN task, respectively.

Additionally, we evaluate the impact of the temperature parameter  $\tau$  within the range  $[0.05, 1.0]$ . Results presented in Figure 4 demonstrate that our method exhibits relatively low sensitivity to the choice of  $\tau$ . Consequently, we uniformly set  $\tau = 0.2$  across all tasks for simplicity and consistency in the hybrid late-fusion. Note that we set  $\tau = 1$  constantly for MoE fusion in our experiments.

## 5 Limitations

**Computational inefficiency in extreme heterogeneity:** The decoder’s computational complexity scales quadratically with the number of modalities ( $m^2$ ), which becomes prohibitive for large-scale medical datasets containing high-resolution 3D images and tabular data. This bottleneck restricts real-time deployment in clinical scenarios where latency is critical.

**Imbalanced convergence dynamics:** Significant disparity exists between the slow convergence of high-resolution image backbones and faster-converging tabular modality learners [8]. This imbalance exacerbates modality competition during joint training, potentially leading to feature collapse in extreme heterogeneous settings. Extreme heterogeneity in data types (3D images vs. tables) causes imbalanced learning dynamics, risking model collapse when modality-specific convergence rates diverge significantly.

**Sensitivity to modality representation gaps:** The model struggles with extreme modality heterogeneity where spatial-semantic relationships differ fundamentally between modalities (e.g., pixel-level dependencies in 3D scans vs. discrete features). This can amplify noise propagation across modalities and degrade fusion quality. Future work should focus on developing adaptive computation mechanisms for modality-specific processing and investigating curriculum learning strategies to mitigate convergence imbalance.

## References

- [1] Chengliang Liu, Jie Wen, Xiaoling Luo, and Yong Xu. Incomplete multi-view multi-label learning via label-guided masked view-and category-aware transformers. In *Proceedings of the AAAI conference on artificial intelligence*, volume 37, pages 8816–8824, 2023.
- [2] Md Kaykobad Reza, Ashley Prater-Bennette, and M Salman Asif. Robust multimodal learning with missing modalities via parameter-efficient adaptation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 47(2):742–754, 2024.
- [3] Shicai Wei, Yang Luo, Yuji Wang, and Chunbo Luo. Robust multimodal learning via representation decoupling. In *European Conference on Computer Vision*, pages 38–54. Springer, 2024.
- [4] Hu Wang, Yuanhong Chen, Congbo Ma, Jodie Avery, Louise Hull, and Gustavo Carneiro. Multi-modal learning with missing modality via shared-specific feature modelling. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 15878–15887, June 2023.
- [5] Hao Wang, Shengda Luo, Guosheng Hu, and Jianguo Zhang. Gradient-guided modality decoupling for missing-modality robustness. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 15483–15491, 2024.
- [6] Linfeng Liu, Siyu Liu, Lu Zhang, Xuan Vinh To, Fatima Nasrallah, and Shekhar S Chandra. Cascaded multi-modal mixing transformers for alzheimer’s disease classification with incomplete data. *NeuroImage*, 277:120267, 2023.
- [7] Muyu Wang, Shiyu Fan, Yichen Li, Zhongrang Xie, and Hui Chen. Missing-modality enabled multi-modal fusion architecture for medical data. *Journal of Biomedical Informatics*, 164:104796, 2025.

- [8] Zhuang Qi, Lei Meng, Zhaochuan Li, Han Hu, and Xiangxu Meng. Cross-silo feature space alignment for federated learning on clients with imbalanced data. In *The 39th Annual AAAI Conference on Artificial Intelligence (AAAI-25)*, pages 19986–19994, 2025.