Adjacent-Aware Modality Recovery Based on Incomplete Multi-Modal Brain Disease Diagnosis

Jinrong Cui, *Member, IEEE*, Weihao Ye, Shengrong Li, Jie Wen, *Senior Member, IEEE*, and Qi Zhu, *Member, IEEE*

***Abstract*— Multi-modal learning is extensively applied to diagnose brain diseases such as epilepsy and Alzheimer's disease. However, incomplete multi-modal data, where some modalities are unavailable or difficult to collect, limits the effectiveness of conventional methods. Additionally, existing approaches often overlook semantic relationships between neighbors with the same-label and latent information in missing modalities. To address these challenges, we propose an adjacent-aware distillation recovery framework designed for incomplete multi-modal learning, with a focus on diagnosing representative brain diseases, *i.e.* epilepsy and Alzheimer's disease. The key novelty of our framework lies in its joint design of adjacent-aware modality recovery and multi-modal representation learning in a single end-to-end pipeline. Specifically, we introduce a label-guided adjacent-aware recovery module that uses a self-attention mechanism to exploit neighbor semantics and generate distribution-consistent features for high-quality modality reconstruction. The recovered features are then refined through a knowledge distillation pathway into a modality generator, enhancing generalization under severe data incompleteness. For multi-modal representation learning, the recovered modality information is fused with the original incomplete information to enhance feature extraction and representation. Extensive experiments demonstrate the effectiveness of our method in diagnosing epilepsy and Alzheimer's disease.**

***Index Terms*— Incomplete multi-modal brain image learning, adjacent-aware recovery, self-attention encoder, knowledge distillation, joint optimization.**

Received 7 August 2025; revised 12 September 2025 and 2 December 2025; accepted 10 January 2026. Date of publication 13 January 2026; date of current version 4 May 2026. This work was supported in part by Shenzhen Science and Technology Program under Grant JCYJ20240813105135047; in part by Guangdong Basic and Applied Basic Research Foundation under Grant 2024A1515030213; in part by the Open Project Program of State Key Laboratory of Virtual Reality Technology and Systems, Beihang University, under Grant VRLAB2025B01; and in part by the National Natural Science Foundation of China under Grant 62576140, Grant 62371234, and Grant 62076129. Recommended by Associate Editor T. Riklin Raviv. *(Corresponding authors: Jie Wen; Qi Zhu.)*

Jinrong Cui and Weihao Ye are with the College of Mathematics and Informatics, South China Agricultural University, Guangzhou 510642, China (e-mail: tweety1028@163.com; 516584369@qq.com).

Shengrong Li and Qi Zhu are with the College of Artificial Intelligence, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China (e-mail: lisrong@nuaa.edu.cn; zhuqinuaa@163.com).

Jie Wen is with the School of Computer Science and Technology, Harbin Institute of Technology, Shenzhen 518055, China (e-mail: jiewen_pr@126.com).

Digital Object Identifier 10.1109/TMI.2026.3654000

I. Introduction

THE human brain, with billions of neurons and synapses, enables complex behaviors and perceptions [1], while diseases like Alzheimer's disease and epilepsy disrupt these functions [2]. Epilepsy is a neurological disorder characterized by sudden seizures that affect movement and consciousness, often involving recurrent and severe convulsions without an identifiable trigger [3]. Alzheimer's disease is a progressive neurodegenerative disease leading to memory loss and cognitive decline, significantly impacting the elderly population [4]. Advances in brain imaging, particularly functional magnetic resonance imaging (fMRI) [5] and diffusion tensor imaging (DTI) [6], can help us better understand and diagnose these conditions by exploring their connection to the nervous system [7]. fMRI captures temporal coherence by detecting changes in brain activity associated with blood flow [8], while DTI reveals the structural connectivity of the brain through the diffusion patterns of water molecules [9]. Disrupted coupling between functional and structural brain networks in certain diseases [10] suggests that combining fMRI and DTI data enables more accurate detection of subtle pathophysiological abnormalities compared to single-modality approaches [11].

In recent years, multi-modal deep neural network technology [12], [13] has overcome the limitations of traditional unimodal diagnostic methods, sparking widespread research interest and significantly advancing the classification of brain diseases [14], [15], [16], [17]. However, the challenge of diagnosing brain diseases with incomplete multi-modal data remains increasingly significant [18]. In brain imaging analysis, a common problem is the absence of certain modalities in some samples [19]. For example, patients may be unable to fully cooperate with all imaging procedures due to physical or economic factors. Additionally, noise or head motion during Diffusion Tensor Imaging (DTI) can hinder fiber tracking, leading to incomplete DTI structural networks. This missing condition can cause a modality to be missing in some samples, thus affecting the results of fusing multi-modal features to diagnose brain diseases. As shown in Fig. 1, different strategies for handling missing modalities lead to distinct decision boundaries. Dropping incomplete samples may result in underfitting due to reduced training data, while relying only on the limited modalities often weakens feature representation. In

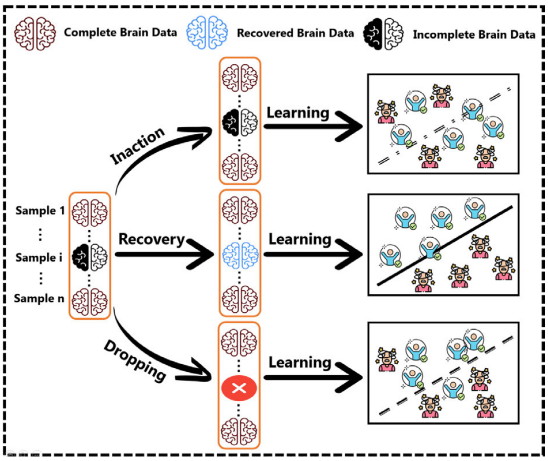


Fig. 1. The schematic illustration of how the processing of incomplete brain data affects disease diagnosis. The less information is missing, the more clearly defined the boundaries of the learned decision categories become.

contrast, strategies that reconstruct or approximate the missing modality can retain multi-modal information and support more accurate and robust decision boundary learning.

In the field of brain disease diagnosis with incomplete multi-modal data, various methods have been proposed to address the challenges posed by missing modalities. For instance, Zhu et al. [20] developed a low-rank representation method for epilepsy classification, achieving promising results with incomplete DTI and fMRI data. Similarly, Meng et al. [21] proposed the Multi-modal Modality-masked Diffusion Network (M2DN), which leverages progressive inpainting and modality masking to flexibly synthesize missing modalities. Despite these advances, these methods generally do not adequately capture the semantic relationships between neighbors with the same label. Moreover, they often overlook the latent information embedded in the missing modalities, limiting their effectiveness in comprehensively leveraging the multi-modal data.

To address these challenges, we propose a deep learning framework for incomplete multi-modal data, the Adjacent-Aware Distillation Recovery Network (ADRNet). This framework recovers missing modalities using the adjacent-aware distillation recovery strategy, which exploits semantic relationships between neighboring modalities with the same label and shared inter-modal semantics to enrich multi-modal representations. To enhance both robustness and efficiency, we introduce a modality generator that enables joint optimization, where the outputs of distillation recovery are refined through representation learning, which in turn enhances the effectiveness of the recovery process.

To evaluate the classification performance of our method for epilepsy and Alzheimer's disease, we use two datasets, Epilepsy and ADNI, containing both fMRI and DTI modalities. The Epilepsy dataset is collected from Jinling Hospital, while ADNI is sourced from the Alzheimer's Disease Neuroimaging Initiative [22]. Both datasets are preprocessed using consistent methods, with further details provided in the experiment sections. All experiments in this study are conducted on these two datasets. The code for our method is publicly available at: https://github.com/tweety1028/ADRNet.

The key contributions of this paper are summarized as follows:

- We propose ADRNet, a unified deep multi-modal learning framework specifically designed for incomplete brain disease diagnosis. Unlike existing approaches that separately perform modality recovery and representation learning, ADRNet jointly optimizes adjacent-aware distillation recovery and multi-modal representation learning in an end-to-end manner.
- For distillation recovery learning, we design a label-guided adjacent-aware recovery module with self-attention to capture same-label semantics and generate distribution-consistent features. These features are distilled into a modality generator to enhance generalization under severe data incompleteness.
- For multi-modal representation learning, the recovered and complete modalities are fused to learn discriminative multi-modal representations, which are directly optimized for the downstream diagnostic task, thereby improving classification accuracy and robustness.
- We construct a private dataset Epilepsy and an open-source dataset ADNI. Both datasets include fMRI and DTI modalities, and both contain three categories. The experimental results demonstrate the superior performance of our proposed module compared to previous state-of-the-art methods.

II. Preliminary

A. Problem Definition and Notations

For clarity, we first provide a formal definition of the problem under investigation. In this paper, bold uppercase letters (e.g., $\mathbf{Z}$) are used to denote matrices, while bold italic lowercase letters (e.g., $\boldsymbol{z}$) represent vectors. For convenience, we illustrate our method using fMRI and DTI modalities as examples. Given an incomplete multi-modal dataset $\mathbb{S}$, it consists of three parts: the complete set $\mathbb{S}^p = \{(x_i^p, y_i^p, l_i^p)|i \in [1, N_p]\}$, the fMRI-missing set $\mathbb{S}^d = \{(\star, y_i^d, l_i^d)|i \in [1, N_d]\}$, the DTI-missing set $\mathbb{S}^f = \{(x_i^f, \star, l_i^f)|i \in [1, N_f]\}$, where $\star$ denotes missing data. Here, $N_*, * \in \{p, d, f\}$ denotes the number of samples in each corresponding subset. The variables $x_i^* \in \mathbb{R}^{d_f}, y_i^* \in \mathbb{R}^{d_d}$, and l_i^* represent the fMRI data, DTI data, and the true label of the i-th sample, respectively. The single-label vector $l_i^* \in \{0, 1, 2, \ldots, C\}$ denotes the category to which the i-th sample belongs, where C is the total number of categories. The primary objective of our method is to predict the classification label l_i^* for each sample. A summary of the key notations is listed in Table I.

B. Related Work

1) Incomplete Multi-Modal Brain Imaging Analysis: Incomplete multi-modal brain imaging analysis is an emerging field focused on integrating neuroimaging data from various modalities to explore complex brain structures and relationships. Among these, fMRI and DTI are the most widely used modalities [23]. Recent studies have made notable progress in addressing these issues. Yang et al. [24] proposed a modal-mixup imputation and bilateral deep-supervision network for robust brain disorder diagnosis with incomplete

TABLE I
KEY NOTATIONS OF OUR FRAME

Notations	Descriptions
$* \in \{p, d, f\}$	subset marker, p: complete, d: fMRI-missing, f: DTI-missing
$\mathbb{S}^* = \{(x_i^*, y_i^*, l_i^*) \mid i \in [1, N_*]\}$	training set
$x_i^* \in \mathbb{R}^{d_f}$	fMRI data of the i-th sample
$y_i^* \in \mathbb{R}^{d_d}$	DTI data of the i-th sample
$l_i^* \in \{0, 1, 2, \ldots, C\}$	label of the i-th sample
$f_i^* \in \mathbb{R}^{d_f}$	latent fMRI features of the i-th sample
$d_i^* \in \mathbb{R}^{d_d}$	latent DTI features of the i-th sample
$h_i^* \in \mathbb{R}^{d_h}$	fusion features of the i-th sample
$X^a \in \mathbb{R}^{N_a \times d_f}$	anchor set of fMRI features
$Y^a \in \mathbb{R}^{N_a \times d_d}$	anchor set of DTI features
$\widetilde{x_i^d}$	recovered fMRI feature of the i-th sample
z_i^d	fMRI latent semantics of the i-th sample
ϕ_i	aggregation embedding of i-th sample
N_*	size of * subset
N	number of training samples
B	number of brain areas
T	number of time series
d_f	dimension of fMRI feature
d_d	dimension of DTI feature
d_h	dimension of fusion feature
r_f	missing rate of fMRI feature
r_d	missing rate of DTI feature
C	number of categories
N_a	size of the anchor set

modalities on the ADNI dataset. Liu et al. [25] proposed an autoencoder-based multi-view missing data completion framework (AEMVC) for addressing incomplete multi-modal data in AD diagnosis by learning robust representations through latent space mapping, graph regularization, and kernel methods. Wang et al. [26] presented ShaSpec, a method that models shared and specific features across modalities, enhancing data representation through auxiliary tasks and feature fusion for classification and segmentation. While these approaches demonstrate the potential of incomplete multi-modal analysis, they do not fully utilize neighbor-related semantics and inter-modal commonalities to guide the recovery of missing modalities.

2) Self-Attention Mechanism in Brain Imaging Analysis: In recent years, the self-attention mechanism [27] has become increasingly popular in brain imaging analysis for its effectiveness in modeling long-range dependencies and capturing global context. Unlike traditional convolutional approaches that focus on local receptive fields, self-attention models pairwise interactions across all input elements, making them ideal for complex brain imaging data such as fMRI and DTI. For instance, Zhang et al. [28] introduced mmFormer, a Transformer-based approach leveraging hybrid and modality correlated encoders for robust brain tumor segmentation from incomplete MRI data. Similarly, Oh et al. [29] integrated self-attention into unsupervised learning networks to enhance non-rigid registration by capturing both local and global structural features in brain MRIs. Huang et al. [30] developed a new deep learning-based segmentation method with a contextual transformer block and hybrid dilated convolution for fetal brain MRI segmentation.

Mathematically, the self-attention mechanism computes the output representation z_i as:

$$z_i = \text{Softmax}\left(\frac{\mathbf{q}_i \mathbf{K}^T}{\sqrt{d_k}}\right)\mathbf{V} \tag{1}$$

where $q_i \in \mathbb{R}^{1 \times d_k}$ denotes the query vector, and K, V represent the key and value matrices derived from all input features. d_k is the dimensionality of the key vectors. This formulation allows the model to focus on relevant regions of the input, facilitating effective feature extraction from high-dimensional brain imaging data.

III. METHOD

As illustrated in Fig. 2, we propose an enhanced incomplete multi-modal classification model, the **A**djacent-aware **D**istillation **R**ecovery **N**etwork (**ADRNet**), for a supervised single-label classification task, *i.e.* incomplete brain disease diagnosis. The framework consists of three key modules: the Adjacent-aware Recovery ("$\mathcal{A}$"), the Modality Generator ("$\mathcal{G}$"), and the Multi-modal Feature Encoder ("$\mathcal{E}$"). In this work, we define "adjacent" as referring to anchor samples that share the same label and exhibit high feature similarity with the target sample. The class-consistent semantic information distilled from these anchors, referred to as adjacent semantic, is used to guide the modality recovery process.

The workflow illustrated in Fig. 2 is described as follows: Firstly, the multi-modal data are processed through *MLP* to obtain multi-modal embeddings. Subsequently, the original features are subjected to "$\mathcal{A}$" to capture the adjacent semantics, and the learned recovery information is distilled into "$\mathcal{G}$" to enhance the generalization capability. Next, the Multi-modal Feature Encoder ($\mathcal{E}$) extracts potential semantic representations from both the original and recovered features, and merges them within the fusion module ($\mathcal{F}$) to generate the final latent multi-modal representations. It is worth noting that "$\mathcal{G}$" is supervised by "$\mathcal{A}$" during the training phase, and only "$\mathcal{G}$" is in effect during the prediction phase.

A. Feature Transformation

In this subsection, we first extract the feature embedding matrix. The N samples of incomplete brain image data, consisting of both fMRI and DTI modalities, are divided into B regions, with the fMRI data further segmented into T time series. After pre-processing, as described in Section IV-A.2, we obtain two tensors: $\hat{F} \in \mathbb{R}^{N \times B \times T}$, which contains dynamic functional information, and $\hat{D} \in \mathbb{R}^{N \times B \times B}$, which represents structural connectivity information. Additionally, we have a single-label matrix $L = \{l_i | l_i \in \{0, 1\}^{1 \times C}, i \in [1, N]\}$. To reduce computational complexity and facilitate subsequent processing, it is necessary to extract embeddings from the tensor data. Since the dimensions $B \times T$ and $B \times B$ can be excessively large, potentially leading to computational inefficiencies or overflow, we apply a Multilayer Perceptron (*MLP*) [31] to perform effective dimensionality reduction. The *MLP* comprises a hidden layer and a fully connected layer, designed to transform high-dimensional data into a more concise and easier-to-process representation, as shown below:

$$\begin{aligned} \tilde{F} &= MLP(\hat{F}; d_f), \\ \tilde{D} &= MLP(\hat{D}; d_d), \end{aligned} \tag{2}$$

where $\tilde{F} = \{x_i \mid x_i \in \mathbb{R}^{d_f}, i \in [1, N]\}$ and $\tilde{D} = \{y_i \mid y_i \in \mathbb{R}^{d_d}, i \in [1, N]\}$, with d_f and d_d representing the fMRI and DTI feature dimensions.

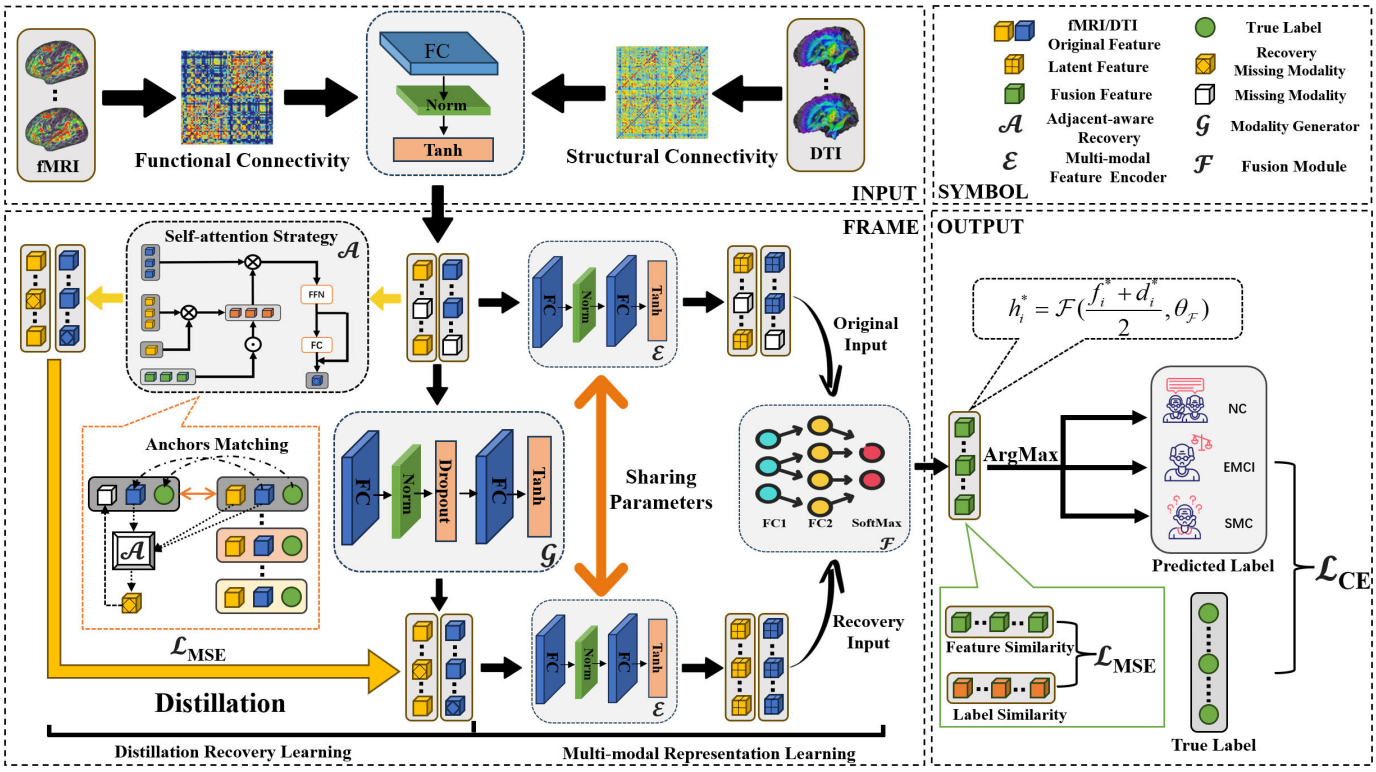


Fig. 2. Overview of the proposed Adjacent-aware Distillation Recovery Network (ADRNet). The framework consists of two main components: (1) a cross-modal recovery learning module, where the adjacent-aware recovery module "$\mathcal{A}$" captures neighboring semantics and distills them into the modality generator "$\mathcal{G}$" for missing modality reconstruction; and (2) a multi-modal feature learning module, which includes a feature encoder "$\mathcal{E}$" and a fusion module "$\mathcal{F}$" to extract and integrate features from both original and recovered modalities.

B. Distillation Recovery Learning

Samples with the same label inherently share semantic relevance, and semantic consistency exists across different modalities of the same sample. Missing modality can be effectively complemented using information from label-consistent anchor samples and other available modalities. Under this basis, we design a distillation recovery learning framework comprising: (1) Adjacent-aware Module ("$\mathcal{A}$"); (2) Modality Generator ("$\mathcal{G}$").

1) Adjacent-Aware Recovery ("$\mathcal{A}$"): Let's begin by introducing the core idea of this module. When fMRI data are missing, the missing values can be supplemented using data from label-consistent and semantically similar samples within the same modality, as well as from other modalities of the same sample. This assumes semantic correlations between neighboring regions and shared semantic information across different modalities of the same sample. By aggregating intra-modal information from samples with the same label, we can generate representations that are consistent in distribution. In general, similar data points tend to be close to one another. Therefore, it becomes essential to construct an anchor set from label-consistent and semantically similar samples, enabling the model to identify the most relevant anchors for each incomplete sample. To address the challenge of matching the missing data with the most similar anchors and capturing the shared semantics across modalities, we employ a self-attention strategy, which facilitates this process effectively. In this section, we present the module with an example drawn from the incomplete instances in the fMRI-missing set, $\mathbb{S}^d$. Before searching for the most similar neighbors within its adjacent region, it is necessary to construct an anchor set to narrow the search scope. This process can be uniformly formalized as follows:

$$\begin{aligned}\mathbf{X}^a &= \{(x_i^p, l_i^p) \mid (x_i^p, y_i^p, l_i^p) \in RandomSample(\mathbb{S}^p, N_a)\},\\ \mathbf{Y}^a &= \{(y_i^p, l_i^p) \mid (x_i^p, y_i^p, l_i^p) \in RandomSample(\mathbb{S}^p, N_a)\},\end{aligned} \tag{3}$$

where N_a is the size of the anchor set, and the function $RandomSample(\mathbb{S}^p, N_a)$ randomly selects N_a samples from $\mathbb{S}^p$. The anchor sets $\mathbf{X}^a \in \mathbb{R}^{N_a \times d_f}$ and $\mathbf{Y}^a \in \mathbb{R}^{N_a \times d_d}$ denote the anchors of fMRI or DTI from $\mathbb{S}^p$, respectively. Given these anchors, the adjacent-aware module "$\mathcal{A}$" leverages the complete modality of the target sample and the anchor sets to recover the missing modality, formulated as:

$$(\widetilde{x_i^d})_{\mathcal{A}} = \mathcal{A}(y_i^d, \mathbf{Y}^a, \mathbf{X}^a; \theta_{\mathcal{A}}), \tag{4}$$

where $\theta_{\mathcal{A}}$ is the trainable parameter, and $(\widetilde{x_i^d})_{\mathcal{A}}$ denotes the recovered fMRI feature of fMRI-Missing samples, generated by module "$\mathcal{A}$". As depicted in Figure 3, each anchor is assigned a semantic similarity based on how closely it resembles the DTI data y_i^d of the fMRI-Missing samples. Semantically relevant adjacent samples from the anchor set are then used for learning, and eq.(4) describes this process.

Based on Eq. (1), we introduce a learnable recovery strategy utilizing the self-attention mechanism to learn relationships between adjacent samples and facilitate adjacent-aware recovery. This approach can be expressed as:

$$\mathbf{K} = \mathbf{Y}^a\mathbf{W}_K, q_i^d = y_i^d\mathbf{W}_Q, \mathbf{V} = \mathbf{X}^a\mathbf{W}_V, \tag{5}$$

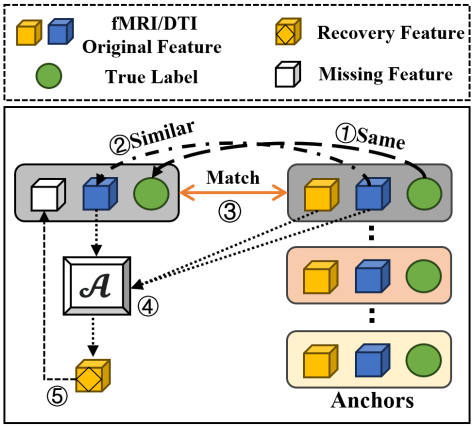


Fig. 3. The details about the process of matching missing samples to complete sample anchors. The process matches anchor points with the same label and similar complete modalities as the missing sample, then inputs the matched anchors and the sample's complete modality into Module "$\mathcal{A}$" to recover the missing modality.

where $\mathbf{W}_Q \in \mathbb{R}^{d_d \times d_k}$, $\mathbf{W}_K \in \mathbb{R}^{d_d \times d_k}$, $\mathbf{W}_V \in \mathbb{R}^{d_f \times d_k}$ are trainable parameters. To improve the quality of the selected nearest adjacent samples, we refine the learnable attention scores by introducing a label similarity score, denoted as s_i, which measures the overall semantic consistency between the i-th sample and the entire anchor set based on their label information, and is defined as:

$$s_i = \Sigma_{j=1}^{N_a} 2 \times \frac{1}{1 + exp(-(l_i^d \times l_j^{aT} + KL(\widetilde{l}_i^d, \widetilde{l}_j^a)))} - 1, \quad (6)$$

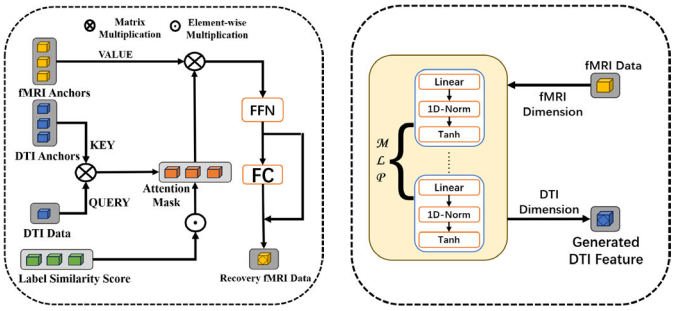
where $KL(\cdot,\cdot)$ is Kullback-Leibler Divergence [32], $\widetilde{l}_i^d$ and $\widetilde{l}_j^a$ are regularized. We employ KL divergence to measure the distributional difference between label representations, as it provides a principled way to quantify how much information is lost when approximating one label distribution with another. This is particularly important for capturing semantic relationships between samples with similar but not identical label patterns, enabling more nuanced similarity assessment beyond simple label matching. The label similarity score encodes inter-sample label relationships, enabling the refinement of self-attention scores to prioritize semantically relevant anchor samples. As illustrated in Figure 4a, the self-attention mechanism designed according to our method can be formulated as:

$$z_i^d = softmax\left(\frac{q_i^d \mathbf{K}^T}{\sqrt{d_k}} \odot s_i\right)\mathbf{V}, \quad (7)$$

where z_i^d represents the latent semantics of the fMRI-Missing samples, and $\odot$ represents the Hadamard product. Then we process the aggregated semantic information further to recover the missing modality as follows:

$$\hat{z}_i^d = LN(LN(z_i^d) + FFN(z_i^d; \theta_{FFN})), \quad (8)$$

where $LN(\cdot)$ denotes layer normalization and $FFN(\cdot;\theta_{FFN})$ refers to a feed-forward network. Eq. (8) combines the normalized feature sequences with the original input feature sequences, then transforms the spliced features using a feed-forward network. This process helps alleviate the gradient vanishing problem in deep networks and enables the learning of more complex feature representations. Finally, we design a



(a) Self-attention Strategy (b) Modality Generator

Fig. 4. Details of the distillation recovery learning module, containing the Transformer strategy and the multi-modal encoder.

fully-connected layer $FC(\cdot;\theta_{FC})$ to reconstruct missing fMRI modality:

$$(\widetilde{x_i^d})_{\mathcal{A}} = FC(\hat{z}_i^d; \theta_{FC}), \quad (9)$$

where θ_{FC} is the trainable parameter. It is important to note that this module will only be executed in the training phase for reasons that will be analyzed subsequently.

2) Modality Generator ("$\mathcal{G}$"): Using the adjacent-aware module "$\mathcal{A}$" directly as the final recovery module can present significant challenges. During the adjacent-aware process, certain modalities may be entirely missing, leaving the model unable to identify anchor points with relevant semantics. Another reason is that, considering the boundary condition where the data in the prediction phase may be complete, the recovered modality information could be less informative than the original data. In such cases, the adjacent semantics learned by the "$\mathcal{A}$" module may lack proper guidance, resulting in flawed recovery and negatively affecting the quality of the generated missing modality.

To deal with this concern, we introduce a linear modality generator "$\mathcal{G}$," which leverages the rich semantics captured by the "$\mathcal{A}$" module to guide the final recovery process. This essentially distills the knowledge learned by module "$\mathcal{A}$" into module "$\mathcal{G}$", guiding "$\mathcal{G}$" to capture the hidden semantics of the data and thereby enhancing its generalization capability. This approach uses available modalities to enrich the recovery process by integrating shared semantics across modalities. The process can be formulated as follows:

$$(\widetilde{x_i^d})_{\mathcal{G}} = \mathcal{G}(y_i^d; \theta_{\mathcal{G}}), \quad (10)$$

where $(\widetilde{x_i^d})_{\mathcal{G}}$ is the fMRI feature of $\mathbb{S}^d$ generated by the "$\mathcal{G}$" module, $\theta_{\mathcal{G}}$ are the trainable parameters, and $\mathcal{G}(\cdot;\theta_{\mathcal{G}})$ consists of multiple MLP modules. The module $\mathcal{G}$ utilizes the common semantics between modalities to achieve mutual recovery. While the generator leverages available modalities to recover missing modalities and maintain semantic consistency across modalities, it inevitably introduces a distributional inconsistency between the recovered modality and the original modality. As illustrated in Figure 4, the generated feature recovered only by the modality generator lacks effective guidance, which can hinder the overall quality of recovery. Therefore, the strategy of distilling the learned nearest-neighbor semantics from "$\mathcal{A}$" to "$\mathcal{G}$" naturally and reasonably solves this problem, making the whole distillation recovery learning more robust and generalizable.

C. Multi-Modal Representation Learning

After completing the distillation recovery learning process, the next step is multi-modal feature learning. As illustrated in Fig. 2, we introduce a multi-modal feature encoder, "$\mathcal{E}$", for each modality to map the data into a common latent space as follows:

$$f_i^* = \mathcal{E}_x(x_i^*; \theta_{\mathcal{E}_x}), d_i^* = \mathcal{E}_y(y_i^*; \theta_{\mathcal{E}_y}), \tag{11}$$

where $f_i^* \in \mathbb{R}^{d_f}$ and $d_i^* \in \mathbb{R}^{d_d}$ represent the latent features of fMRI and DTI, respectively, while x_i^* and y_i^* correspond to the complete data and the recovered features generated by "$\mathcal{G}$". The feature encoder, implemented as an MLP-based neural network, outputs relaxed modal features suitable for subsequent fusion. These multi-modal common representations are then fused, followed by a classification layer to predict the classification outcome, as expressed below:

$$h_i^* = \mathcal{F}\left(\frac{f_i^* + d_i^*}{2}, \theta_{\mathcal{F}}\right), \tag{12}$$

where the fusion module "$\mathcal{F}$" is implemented using a weighted averaging strategy, and h_i^* denotes the fused semantic representation. Finally, the predicted labels are obtained as follows:

$$\widehat{l}_i^* = Argmax(Softmax(h_i^*)). \tag{13}$$

Notably, by integrating the recovery modalities into the multi-modal representation learning process, the optimization of representation learning indirectly facilitates the distillation recovery learning. Therefore, our framework enables the joint optimization of both the distillation recovery module and multi-modal representation learning.

D. Loss Function

To achieve training convergence, we first distill the output of "$\mathcal{A}$" to "$\mathcal{G}$" using the following function:

$$\mathcal{L}_{distill} = \Sigma_{i=1}^{N} \parallel (\widetilde{x_i^*})_{\mathcal{G}} - (\widetilde{x_i^*})_{\mathcal{A}} \parallel_2^2 + \parallel (\widetilde{y_i^*})_{\mathcal{G}} - (\widetilde{y_i^*})_{\mathcal{A}} \parallel_2^2, \tag{14}$$

Subsequently, we calculate the classification loss by comparing the predicted labels $\widehat{l}_i^*$ with the true labels l_i^* using cross-entropy loss:

$$\mathcal{L}_{label} = \Sigma_{i=1}^{N} CE(\widehat{l}_i^*, l_i^*), \tag{15}$$

where $CE(\cdot)$ is the Cross-entropy loss.

To further enhance joint optimization, we introduce a pairwise matching loss as follows:

$$\mathcal{L}_{pair} = \Sigma_{i \neq j}^{N} (cos(h_i^*, h_j^*) - s_{ij})^2, \tag{16}$$

where s_{ij} is the label similarity score between the i-th sample and the j-th anchor sample. Finally, the total loss function is:

$$\mathcal{L} = \lambda_1 \mathcal{L}_{distill} + \lambda_2 \mathcal{L}_{label} + \mathcal{L}_{pair}, \tag{17}$$

where λ_1 and λ_2 are trade-off hyper-parameters balancing the contribution of each loss term. Algorithm 1 summarizes the proposed method.

Algorithm 1 The Algorithm of the Proposed Method

Input : fMRI data $\hat{F} \in \mathbb{R}^{N \times B \times T}$ and DTI data $\hat{D} \in \mathbb{R}^{N \times B \times B}$, true label $Y \in \mathbb{R}^N$, hyper-parameters λ_1, λ_2, anchor size N_a;
Output: Prediction $\widehat{L}$ of the test set;
Use fMRI and DTI data with Eq. (2) to construct the feature matrix $\widetilde{F}$ and $\widetilde{D}$;
Find anchors X^a and Y^a in $\mathbb{S}^p$ according to N_a with a random way;
for *epochs* **do**
  Calculate the label similarities s_i by Eq. (6);
  for *warm-up epochs* **do**
    Calculate the recovery data $(\widetilde{x_i^*})_{\mathcal{A}}$ by Eq. (7) - Eq. (9);
    Generate the recovery data $(\widetilde{x_i^*})_{\mathcal{G}}$ with the modality generator "$\mathcal{G}$" by Eq. (10);
    Calculate the adjacent loss $\mathcal{L}_{distill}$ by Eq. (14);
  Calculate the multi-modal feature $h_{f,i}^*$ and $h_{d,i}^*$ by Eq. (11);
  Calculate the fusion feature h_i^* by Eq. (12) and then calculate the predicted labels $\widehat{L}$ by Eq. (13);
  Calculate the label loss $\mathcal{L}_{label}$ and pair loss $\mathcal{L}_{pair}$ by Eq. (15) - Eq. (16);
  Finally calculate the final objective loss $\mathcal{L}$ by Eq. (17);
  Back-propagate loss $\mathcal{L}$ to update model parameters;
Return: Prediction label $\widehat{L}$ of the test set.

TABLE II
CHARACTERISTICS OF PARTICIPANTS OF TWO DATASETS

Dataset	Type	Quantity	Sex	Age
Epilepsy	NC	114	58/56	26.2
	FLE	103	53/50	24.1
	TLE	89	45/44	25.3
ADNI	NC	82	40/42	74.3
	SMC	76	50/29	77.2
	EMCI	69	37/32	74.8

IV. EXPERIMENTS

A. Materials

1) Data Collection: In this section, we conduct experiments on the Epilepsy dataset and the ADNI dataset. Detailed demographic information, including gender (Male/Female) and age (Mean ± Std), is provided in Table II. Detailed descriptions of the datasets are provided as follows:

a) Epilepsy: The resting-state fMRI and DTI data were collected from Jinling Hospital, Nanjing University School of Medicine. The Epilepsy dataset includes 306 subjects: 114 NC (58 males, 56 females; mean age 26.2 years), 103 FLE (53 males, 50 females; mean age 24.1 years), and 89 TLE (45 males, 44 females; mean age 25.3 years). Imaging data were acquired using a Siemens Trio 3T scanner. The DTI scanning parameters included an echo time of $93ms$, repetition time of $6100ms$, voxel size of $0.94 \times 0.94 \times 3mm^3$ and a flip angle of 90°. The rs-fMRI scanning parameters were an echo time of $30ms$, repetition time of $2000ms$, voxel size of $3.75 \times 3.75 \times 3.75\ mm^3$ and a flip angle of 90°.

b) Adni: A total of 226 samples were selected from ADNI2 and ADNI3 (http://adni.loni.usc.edu), each including both fMRI and DTI modalities. The dataset comprises 82 NC, 76 SMC, and 69 EMCI patients. The NC group includes 40 males and 42 females with a mean age of 74.3 years. The EMCI group consists of 37 males and 32 females with an average age of 74.8 years, while the SMC group includes 50 males and 29 females with an average age of 77.2 years.

2) Preprocessing: Specifically, all DTI data are preprocessed using PANDA [33]. The FSL toolbox [34] is employed to correct for distortions in the DTI data, and TrackVis [35] is utilized to generate fiber tract images. Anatomical regions are identified using the AAL criteria based on the subject's T1 images. Finally, the number of fibers is calculated as a measure of the structural connectivity between regions. For rs-fMRI data, preprocessing is carried out using SPM8 [36] as implemented in the DPARSF toolbox [37]. Initially, the serialized data are segmented and realigned to the EPI template to correct for motion artifacts. Detrending is applied to minimize the effects of head motion and to reduce interference from cerebrospinal fluid (CSF) and white matter. Following preprocessing, the rs-fMRI data are segmented into 90 regions of interest (ROIs) using the Automated Anatomical Labeling (AAL) atlas. In this process, the fMRI data for epilepsy patients are divided into 240 timepoints, whereas the ADNI dataset is divided into 197 timepoints.

B. Experimental Settings

We begin by outlining the experimental settings of the submodules in ADRNet, as described below:

- **Fully-Connected Layer.** The modality generator and the multi-modal representation learning process in Eqs. (10)-(12), are implemented using Fully-Connected (FC) layers. Each FC layer consists of a linear projection, batch normalization, and a $Tanh(\cdot)$ activation function.
- **Transformer Encoder.** We use a single transformer encoder layer, with a latent dimension of 1,024 for the self-attention mechanism (Eq. (7)) and 2,048 for the feed-forward network $FFN\ (.;\theta_{FFN})$.
- **Modality Generator.** This module comprises two FC layers, with a latent dimension of 2,048.
- **Multi-modal Feature Encoder.** Composed of two FC layers, this encoder processes fMRI and DTI modalities with latent dimensions of 2,048 and 1,024, respectively.
- **Fusion Module.** This module consists of a single FC layer followed by a $Softmax(\cdot)$ activation function.

The hyper-parameters in Eq. (17) are uniformly set for both the Epilepsy and ADNI datasets, with values $\lambda_1 = 0.35, \lambda_2 = 0.5$. The ADAM optimizer [38] is used with a learning rate of 0.01. The batch size is set to 128, while the size of anchor set N_a is adjusted to 100. We begin by optimizing the objective function in Eq. (14), followed by the optimization of Eq. (17). Specifically, the adjacent-aware recovery module is trained for 50 epochs. Subsequently, the overall framework, illustrated in Figure 2, is trained for 100 epochs on both the Epilepsy and ADNI datasets. The experiments are run on a system with a single NVIDIA RTX 4090 GPU and an Intel i9-13900KF CPU.

Since the collected dataset is complete, we simulate data incompleteness by randomly removing modalities from selected samples in the feature matrix. Specifically, we generate the missing sets $\mathbb{S}^d$ and $\mathbb{S}^f$ by randomly removing data from the original complete dataset $\mathbb{S}$ as follows:

$$\begin{aligned}\mathbb{S}^d &= RandomMiss(\mathbb{S}, r_d, N_d),\\ \mathbb{S}^f &= RandomMiss(\mathbb{S}, r_f, N_f),\end{aligned} \tag{18}$$

where the function $RandomMiss(\cdot)$ randomly selects N_d or N_f samples and removes their fMRI or DTI modalities according to the specified missing rates r_d and r_f. This approach enables the generation of artificial incomplete datasets that mimic real-world scenarios, providing a basis to assess the robustness of our method across different levels of data missingness.

C. Evaluation Metric and Baselines

For the evaluation metrics, we utilize accuracy (ACC), F1 score, and area under the curve (AUC) [39], [40], [41] to assess the performance of our method. To comprehensively evaluate ADRNet and benchmark methods under different missing data conditions, we introduce five levels of Missing Data Ratio (MDR) [42] for performance assessment. For instance, an MDR of 90% indicates that 90% of the data samples have missing modalities, with the missing proportions of the fMRI and DTI modalities being equal (*i.e.* 45% each). To ensure a comprehensive comparison, our approach is benchmarked against four state-of-the-art multi-modal brain imaging methods:

- *AEMVC* [25]: This method employs an autoencoder-based framework that completes missing MRI/PET modalities by aligning latent features using kernel space embedding, graph regularization, and HSIC constraints, while preserving structural and cross-modal correlations.
- *MMBNet* [24]: This method introduces a modal-mixup strategy to synthesize complete samples from incomplete data, combined with a bilateral deep-supervision network to regularize functional and structural features for robust classification independently.
- *mmformer* [28]: This method integrates CNN and Transformer modules, using modality-specific encoders and an inter-modal Transformer to extract local-global features and align cross-modality dependencies for segmentation with missing MRI modalities.
- *ShaSpec* [26]: This method models shared and modality-specific features through parallel encoders and residual fusion, enhanced by auxiliary domain classification and distribution alignment to handle missing modalities in multi-modal learning tasks.

Most of the experimental results are obtained by adapting the publicly available source codes of the respective methods to our framework and datasets.

D. Classification Performance

As demonstrated in Table III, our method consistently outperforms all baselines on the ADNI and Epilepsy datasets.

TABLE III
THE CLASSIFICATION PERFORMANCE (*Mean* ± *STD*) WITH VARIOUS MDR (%)

MDR	Method	Epilepsy			ADNI		
		ACC	F1	AUC	ACC	F1	AUC
10%	AEMVC [25]	81.70 ± 1.79	79.52 ± 1.56	80.56 ± 1.34	79.65 ± 2.03	78.93 ± 1.91	79.84 ± 1.36
	MMBNet [24]	91.50 ± 4.71	89.02 ± 4.46	90.23 ± 7.01	88.05 ± 6.36	87.32 ± 6.86	86.56 ± 8.96
	mmformer [28]	85.95 ± 5.61	85.27 ± 7.12	86.30 ± 5.25	84.74 ± 4.68	83.36 ± 7.23	83.57 ± 4.15
	ShaSpec [26]	89.54 ± 2.92	91.78 ± 2.96	88.97 ± 1.68	87.61 ± 3.55	87.89 ± 4.21	88.45 ± 3.04
	ADRNet (Ours)	**94.44** ± 1.89	**93.21** ± 1.21	**92.53** ± 1.34	**90.71** ± 1.17	**89.57** ± 1.27	**89.42** ± 1.32
30%	AEMVC	79.74 ± 1.34	77.26 ± 2.75	79.56 ± 1.73	78.32 ± 2.59	79.34 ± 1.83	78.23 ± 2.97
	MMBNet	90.20 ± 5.33	90.28 ± 5.48	89.87 ± 6.18	86.73 ± 8.85	84.85 ± 8.81	85.64 ± 5.07
	mmformer	82.35 ± 4.80	79.91 ± 6.21	78.09 ± 6.05	80.97 ± 3.32	82.24 ± 4.53	81.43 ± 8.10
	ShaSpec	87.58 ± 3.10	87.62 ± 4.55	86.53 ± 3.26	85.84 ± 3.02	84.14 ± 4.15	85.76 ± 3.01
	ADRNet (Ours)	**93.14** ± 2.21	**92.78** ± 2.44	**91.76** ± 1.84	**88.94** ± 1.57	**88.23** ± 2.48	**89.37** ± 1.32
50%	AEMVC	78.41 ± 2.23	77.53 ± 3.35	78.35 ± 2.97	74.62 ± 1.68	74.82 ± 3.26	73.33 ± 2.81
	MMBNet	88.97 ± 4.24	87.89 ± 4.37	90.14 ± 5.73	85.67 ± 7.52	84.11 ± 7.78	86.79 ± 6.84
	mmformer	80.25 ± 5.35	80.53 ± 3.44	79.46 ± 6.84	77.69 ± 3.87	75.37 ± 5.33	76.32 ± 6.13
	ShaSpec	86.27 ± 3.76	87.43 ± 2.69	86.76 ± 4.56	84.49 ± 2.64	84.98 ± 1.46	85.33 ± 3.54
	ADRNet (Ours)	**91.36** ± 1.84	**90.07** ± 1.60	**92.12** ± 1.28	**87.80** ± 1.47	**86.34** ± 1.76	**88.76** ± 2.80
70%	AEMVC	71.77 ± 2.62	71.23 ± 3.68	70.68 ± 1.78	68.41 ± 2.24	68.56 ± 2.03	67.78 ± 3.91
	MMBNet	83.45 ± 6.33	82.66 ± 6.24	84.44 ± 5.59	76.63 ± 4.01	76.13 ± 4.99	77.29 ± 7.76
	mmformer	74.97 ± 4.87	74.36 ± 5.72	76.59 ± 6.46	69.29 ± 5.35	69.73 ± 4.22	71.30 ± 3.11
	ShaSpec	81.25 ± 4.94	81.57 ± 5.49	82.46 ± 5.37	73.17 ± 2.96	73.23 ± 6.37	74.09 ± 4.84
	ADRNet (Ours)	**85.08** ± 1.18	**85.53** ± 2.94	**86.36** ± 1.27	**78.05** ± 1.46	**77.30** ± 2.68	**79.59** ± 2.77
90%	AEMVC	67.49 ± 2.35	67.35 ± 2.16	67.07 ± 2.32	61.10 ± 2.81	61.79 ± 1.48	64.03 ± 3.53
	MMBNet	81.37 ± 4.36	81.70 ± 4.34	79.45 ± 3.71	70.89 ± 7.64	70.09 ± 7.87	71.21 ± 9.48
	mmformer	70.69 ± 3.94	69.24 ± 4.31	72.87 ± 6.15	63.41 ± 7.57	62.12 ± 5.73	61.07 ± 6.59
	ShaSpec	78.97 ± 4.06	79.37 ± 3.82	78.46 ± 7.12	68.85 ± 4.84	68.71 ± 5.82	69.24 ± 6.65
	ADRNet (Ours)	**83.80** ± 1.64	**82.45** ± 3.53	**81.89** ± 2.51	**73.17** ± 1.38	**73.96** ± 0.96	**74.97** ± 1.10

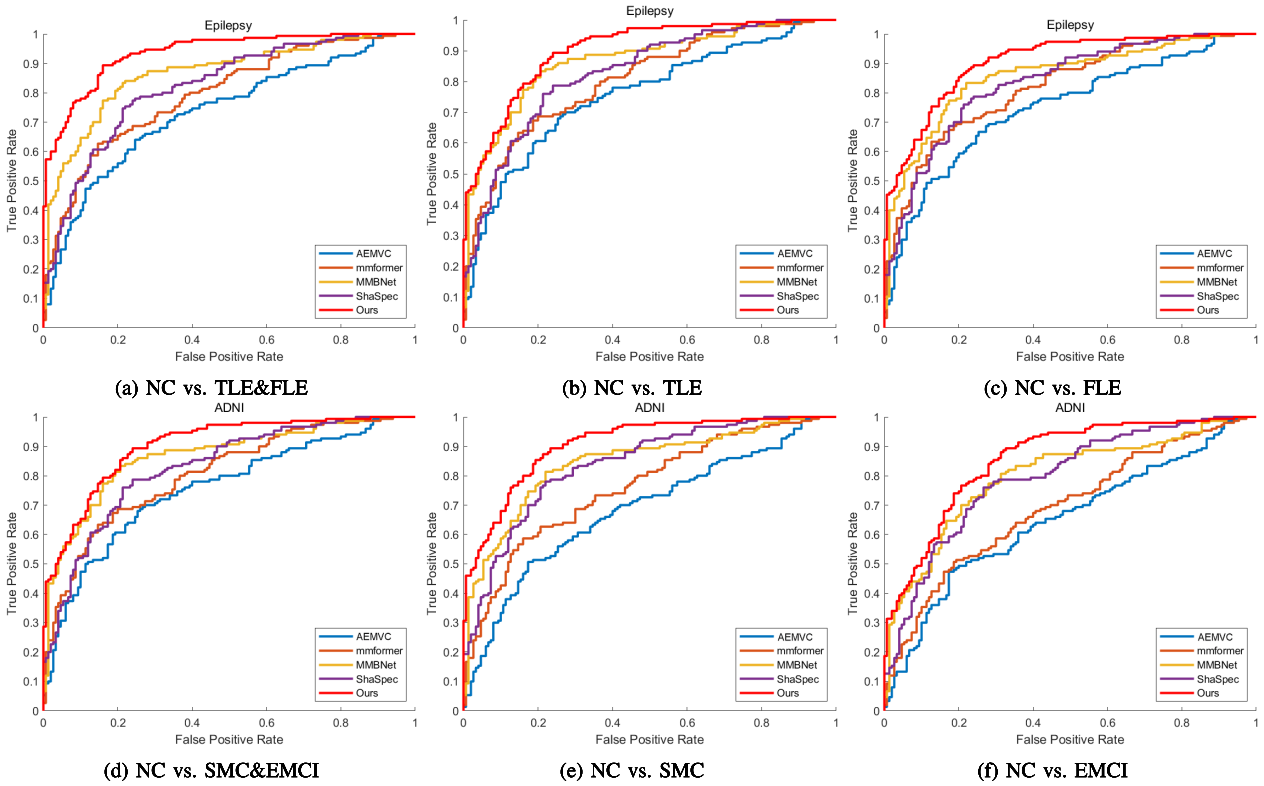


Fig. 5. ROC curves for the proposed method and comparison methods in brain disease diagnosis on the Epilepsy and ADNI datasets with MDR=90%.

The best results are marked in bold, and the second-best results are underlined. Notably, our method surpasses the suboptimal approach by 2.94%, 2.94%, 2.39%, 1.63%, and 2.43% in ACC on the Epilepsy dataset, and by 2.66%, 2.21%, 2.13%, 1.42%, and 2.28% in ACC on the ADNI dataset, under various MDR conditions. Moreover, the F1 Score and AUC achieved by our method consistently exceed those of the suboptimal models on both datasets, further confirming the enhanced performance capabilities of our approach.

To reflect the most challenging real-world scenarios, we further explored the ROC curves under a high MDR of 90%. The ROC curves (Fig. 5) for the NC vs. TLE&FLE, NC vs. TLE, and NC vs. FLE tasks on the Epilepsy dataset and the NC vs. SMC&EMCI, NC vs. SMC, and NC vs. EMCI tasks

TABLE IV
THE RECOVERY PERFORMANCE OF RECOVERED MODALITIES WITH MDR = 90%. (%)

Methods	Epilepsy			ADNI		
	RMS	MSE	SSIM	RMS	MSE	SSIM
AEMVC [25]	70.89	0.251	0.845	68.57	0.273	0.826
MMBNet [24]	<u>80.89</u>	<u>0.156</u>	<u>0.914</u>	76.83	0.198	0.900
mmformer [28]	73.29	0.203	0.875	71.46	0.247	0.859
ShaSpec [26]	78.71	0.179	0.893	<u>77.67</u>	<u>0.167</u>	<u>0.905</u>
ADRNet (Ours)	**86.87**	**0.105**	**0.943**	**83.43**	**0.124**	**0.934**

on the ADNI dataset further illustrate this advantage, with our method's curve dominating the top-left corner and achieving the highest AUC. These results validate the robustness, effectiveness, and superiority of our approach over baseline methods.

This consistent superiority can be attributed to the unique design of ADRNet in comparison to existing methods. While AEMVC and ShaSpec emphasize multi-view representation learning or decomposed feature fusion, they lack targeted strategies for handling missing modalities and do not explicitly enforce semantic alignment. MMBNet and mmformer, though capable of flexible fusion or cross-modal modeling, are limited when modality information is incomplete, as they do not incorporate explicit modality recovery guided by semantic context. In contrast, ADRNet introduces an adjacent-aware recovery module combined with label-guided distillation, which enables it to recover missing modalities more effectively and maintain semantic consistency across samples. This mechanism ensures that the learned representations remain discriminative even under varying degrees of incompleteness, ultimately leading to better classification performance.

Moreover, it is noteworthy that most methods, including ADRNet, achieve their peak performance when the MDR is 10%. This can be attributed to the relatively low level of data incompleteness, which preserves a sufficient number of high-quality complete samples. These intact samples provide reliable semantic context to guide the recovery process, allowing the model to learn inter-modal correlations more effectively. Consequently, the recovery module benefits from clearer supervision, leading to improved reconstruction fidelity and enhanced classification robustness.

E. Recovery Performance

To evaluate the model's capability in restoring missing modalities, we assessed recovery performance under an extremely high MDR of 90%. This setting simulates a worst-case clinical scenario in which a substantial proportion of modality data is unavailable. We employed three widely used quantitative metrics to evaluate the quality of recovered fMRI and DTI modalities:

- **Recovery Modality Similarity (RMS)** [43]: Measures the cosine similarity between the recovered and ground-truth modalities, reflecting semantic consistency.
- **Mean Squared Error (MSE)** [44]: Calculates the average squared difference at the voxel level. Lower values indicate better numerical accuracy.
- **Structural Similarity Index Measure (SSIM)** [45]: Measures image-level perceptual fidelity by comparing structural information between the recovered and ground-truth images. Extensively employed in medical image restoration tasks.

For clarity, the reported RMS, MSE, and SSIM values represent the average performance across both modalities (fMRI and DTI). As shown in Table IV, ADRNet consistently outperforms all baseline methods on both the Epilepsy and ADNI datasets, achieving the highest RMS and SSIM, along with the lowest MSE, even under MDR = 90%. These results highlight the robustness and effectiveness of ADRNet in recovering high-quality modalities under severe missingness.

The superior recovery performance of ADRNet under such extreme conditions can be attributed to two key architectural components: the adjacent-aware recovery module and the knowledge distillation mechanism. The adjacent-aware module effectively captures semantic relationships among same-label neighbors, allowing the model to utilize informative contextual signals when modality data is missing. In parallel, the knowledge distillation strategy facilitates semantic transfer from complete to incomplete modalities, further reinforcing the recovery process. Together, these modules enable ADRNet to minimize voxel-level reconstruction errors and preserve structural integrity, as reflected in the lowest MSE and highest SSIM scores.

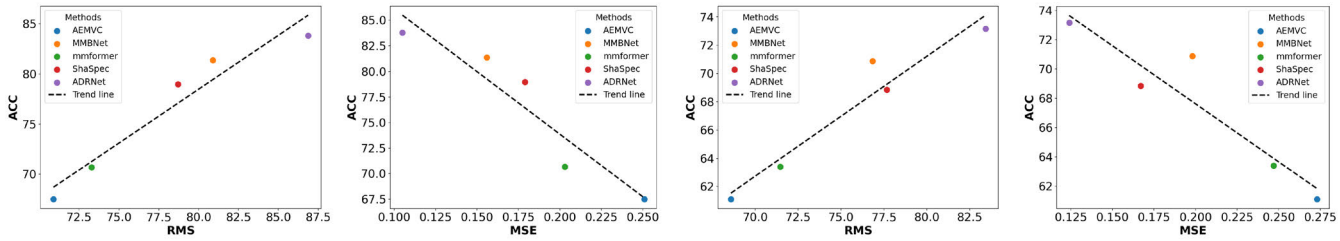
Importantly, this high-quality recovery translates into better downstream classification performance. By integrating the results from Tables III and IV, we plotted scatter plots of RMS vs. ACC and MSE vs. ACC under the 90% MDR condition. As shown in Fig. 6, there is a clear positive correlation between recovery quality and classification performance: models exhibiting higher recovery fidelity consistently achieve superior classification results. We also note that the second-best baseline differs between Tables III and IV for ADNI (MMBNet vs. ShaSpec), despite the small gap. This discrepancy reflects different design priorities: MMBNet emphasizes discriminative supervision for robust classification, whereas ShaSpec focuses on shared-specific feature modeling for modality reconstruction. Overall, ADRNet outperforms all baselines because its adjacent-aware and label-guided recovery enables high-fidelity reconstruction and superior diagnostic accuracy under extreme data incompleteness.

F. Ablation Study

To comprehensively evaluate the impact of key components in ADRNet, we performed an ablation study with the MDR fixed at 90%. Five model variants were designed for comparison: 1) **w/o Adjacent.** This variant excludes the adjacent-aware recovery module "$\mathcal{A}$," using original features to guide the modality generator "$\mathcal{G}$" with a reconstruction loss. 2) **w/o Score.** The label similarity score described in

TABLE V
THE PERFORMANCE (MEAN) OF ABLATION STUDY WITH MDR = 90%. (%)

Variant	Epilepsy			ADNI		
	ACC	F1	AUC	ACC	F1	AUC
w/o Adjacent	58.13	54.19	60.76	41.66	40.94	48.16
w/o Score	70.79	68.30	75.93	57.73	55.33	63.10
w/o Recovery	53.24	52.98	57.95	51.85	55.46	58.38
w/o Pair	60.27	61.53	69.18	61.61	59.24	67.16
w/o Anchor	83.03	82.12	80.98	72.87	73.84	74.57
ADRNet	**83.80**	**82.45**	**81.89**	**73.17**	**73.96**	**74.97**

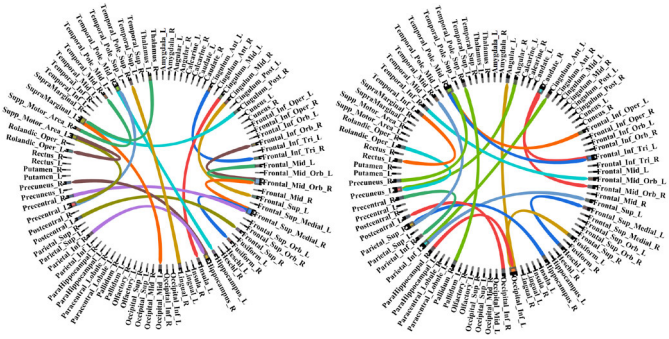


(a) RMS vs. ACC (Epilepsy) (b) MSE vs. ACC (Epilepsy) (c) RMS vs. ACC (ADNI) (d) MSE vs. ACC (ADNI)

Fig. 6. The correlation between recovery performance and classification performance with MDR=90%.

Eq. (6) is removed from the self-attention mechanism. 3) **w/o Recovery.** This variant removes the distillation recovery learning module, directly handling incomplete multi-modal data. 4) **w/o Pair.** The pair-wise matching loss is replaced with a classification loss. 5) **w/o Anchor.** The anchor selection strategy was modified from random sampling to balanced sampling, where each class contributes an equal proportion of anchor samples. The experimental results are presented in Table V, where the best results are marked in bold, and the second-best results are underlined. These results provide the following insights into the contributions of these components:

- **w/o Adjacent:** Removing the adjacent-aware recovery module "$\mathcal{A}$" leads to reduced performance. In scenarios where original features are missing for incomplete multi-modal data, reconstruction supervision cannot be applied. In contrast, "$\mathcal{A}$" effectively distills the adjacent semantic information into "$\mathcal{G}$", enabling a more robust recovery process.
- **w/o Score:** The results demonstrate that incorporating the label similarity score enhances the attention mechanism by filtering irrelevant adjacent data, improving the effectiveness of distillation recovery learning.
- **w/o Recovery:** Removing distillation recovery learning reduces the completeness of multi-modal semantic features, degrading classification performance. The greater the loss of information in the complete modality, the more significant the performance decline, highlighting the sensitivity of distillation recovery learning to multi-modal semantic completeness.
- **w/o Pair:** Removing the pair-wise matching loss causes a significant performance drop. The Hamming distance, based on fused modal features and label similarity, reflects matching relevance during training, making pair-wise matching loss more effective than traditional classification loss.
- **w/o Anchor:** The anchor selection strategy has limited impact on performance, as each dataset contains around 300 samples and the anchor set size is fixed at 100. Due to data splitting and missing rates, most available complete samples are already used as anchors. Additionally, ADRNet's label-aware matching reduces sensitivity to anchor choice.
- **ADRNet:** The third variant exhibits the most significant performance degradation, indicating that removing distillation recovery learning has a direct negative impact. In contrast, excluding the adjacent-aware module allows partial supervision from original features, reducing performance loss. Similarly, even without the label similarity score, the self-attention mechanism still generates effective attention weights, leading to less degradation.

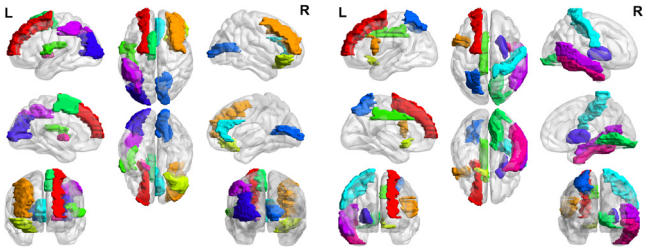


(a) Epilepsy (b) ADNI

Fig. 7. The top 20 most key brain connections in the Epilepsy and ADNI datasets.

G. Biomarker Detection

To investigate the effectiveness of the proposed method in identifying discriminative higher-order brain connections related to Alzheimer's disease and epilepsy, we identified the 20 most key brain connections from our experimental results using significant connectivity changes (SAC) [46]. These connections are visualized in Fig. 7. The identification is based on the fused features generated by ADRNet, which incorporate both complete and recovered modalities, enabling more complete and semantically aligned representations. In the ADNI dataset, discriminative connections are mainly located in the temporal lobe and middle temporal gyrus, superior temporal gyrus, hippocampus, and parahippocampal gyrus, which are



(a) Epilepsy (b) ADNI

Fig. 8. Spatial distribution of the top 10 most discriminative brain regions in the Epilepsy and ADNI datasets.

closely related to memory function. Structural and functional abnormalities in these regions are common in Alzheimer's disease, enabling the differentiation of NC from Alzheimer's patients. Additional key connections in the precentral and middle frontal gyri, associated with higher cognitive functions, highlight their importance in distinguishing NC, SMC, and EMCI. In the Epilepsy dataset, the top 20 most influential brain connections are mainly in the temporal lobe, including the hippocampus, which is related to memory and emotional regulation. Key connections in the frontal and parietal lobes highlight differences between NC, TLE, and FLE patients, playing a crucial role in distinguishing these groups based on memory, cognition, and motor function.

Furthermore, we also observe that some discriminative brain connections are located outside the key regions of the temporal lobe and hippocampus from Fig. 7. For example, in the Epilepsy dataset, discriminative connections include Heschl.L-Frontal.Sup.Medial.L and Pallidum.R-Temporal.Pole.Sup.R, etc. In the ADNI dataset, discriminative connections include Cingulum.Ant.L-Frontal.Inf.Tri.R and SupraMarginal.L-Occipital.Mid.L, etc. This phenomenon may be due to abnormal neural electrical activity caused by brain diseases, as well as compensatory mechanisms. For instance, one of the characteristics of Alzheimer's disease is the degenerative changes in neurons, leading to communication disruptions between brain regions [47]. Additionally, during epileptic seizures, abnormal neuronal discharges form seizure foci in local brain areas, which then spread through neural pathways to other regions [48]. Such irregularities may trigger compensatory responses in unaffected brain regions to mitigate functional deficits.

We also examine the top 10 most significant brain regions identified by ADRNet in both the ADNI and Epilepsy datasets, and visualize them as shown in Fig. 8. The discriminative power of each brain region is calculated by summing the regression coefficients of its associated connections, as derived from the SAC-based non-negative elastic net model. For epilepsy, the key brain regions primarily include Angular, Frontal, and Cingulum, etc., which are closely associated with higher cognitive functions and motor control. For Alzheimer's disease, the key brain regions primarily include the Precuneus, Temporal Mid, and Parietal Inf, which are closely associated with memory, language, and cognitive functions. Studies have highlighted the importance of these regions in diagnosing epilepsy [49] and Alzheimer's disease [50]. These experimental results demonstrate that our method can effectively provide biomarkers for epilepsy and Alzheimer's disease diagnosis.

From a quantitative analysis perspective, the effectiveness of our modality recovery module can be further verified through similarity metrics. As shown in Table IV, ADRNet achieves the highest recovery quality across all metrics on both datasets, including the lowest MSE, highest SSIM, and highest RMS, indicating that the recovered modalities are highly consistent with the original data. Unlike existing methods that either overlook missing modalities or lack explicit recovery strategies, ADRNet incorporates an adjacent-aware recovery mechanism guided by semantic distillation, which enables more accurate and semantically aligned reconstruction. This high-fidelity recovery plays a critical role in supporting reliable biomarker detection, as it ensures that key disease-related patterns are preserved rather than distorted during the imputation process. Consequently, the discriminative brain connections identified by ADRNet, such as those in the hippocampus and temporal lobes for Alzheimer's disease, as well as the frontal-temporal circuitry for epilepsy, are more likely to reflect true pathophysiological abnormalities rather than noise or artifacts. For example, the hippocampus and middle temporal gyrus are well-known to be involved in memory function and are commonly affected in early Alzheimer's disease, while epileptogenic activity frequently originates in the temporal lobe and spreads through frontal and parietal regions. The ability of ADRNet to recover such clinically meaningful connectivity patterns highlights its potential for both accurate diagnosis and biologically grounded interpretation [49], [51].

V. Conclusion

In this paper, we propose ADRNet, an incomplete multi-modal data fusion framework based on adjacent-aware distillation recovery learning for the diagnosis of epilepsy and Alzheimer's disease. Unlike conventional methods, ADRNet jointly performs adjacent-aware modality recovery and multi-modal representation learning in an end-to-end manner, effectively addressing missing modality challenges. This approach not only enables effective model training with incomplete multi-modal data but also effectively recovers missing modalities, leading to more precise predictions. We conduct extensive experiments on both our private Epilepsy dataset and the open-source ADNI dataset. The results demonstrate that our method outperforms existing incomplete multi-modal data fusion approaches in diagnosing epilepsy and Alzheimer's disease. Furthermore, our method can identify biomarkers, offering promising prospects for enhancing the diagnosis of brain disorders. Although the proposed framework achieves promising results, its evaluation is currently limited to small-scale datasets with two imaging modalities, and primarily assumes random missing patterns. In future work, we plan to incorporate large-scale pre-trained models to enhance modality recovery and extend the framework to more diverse or longitudinal neuroimaging datasets.

References

[1] E. Di Lullo and A. R. Kriegstein, "The use of brain organoids to investigate neural development and disease," *Nature Rev. Neurosci.*, vol. 18, no. 10, pp. 573–584, Oct. 2017.

[2] W. Zhu, L. Sun, J. Huang, L. Han, and D. Zhang, "Dual attention multi-instance deep learning for Alzheimer's disease diagnosis with structural MRI," *IEEE Trans. Med. Imag.*, vol. 40, no. 9, pp. 2354–2366, Sep. 2021.

[3] A. A. Asadi-Pooya, F. Brigo, S. Lattanzi, and I. Blumcke, "Adult epilepsy," *Lancet*, vol. 402, no. 10399, pp. 412–424, 2023.

[4] P. Scheltens et al., "Alzheimer's disease," *Lancet*, vol. 397, no. 10284, pp. 1577–1590, 2021.

[5] S. L. Warren and A. A. Moustafa, "Functional magnetic resonance imaging, deep learning, and Alzheimer's disease: A systematic review," *J. Neuroimaging*, vol. 33, no. 1, pp. 5–18, Jan. 2023.

[6] J. Yuan et al., "Machine learning applications on neuroimaging for diagnosis and prognosis of epilepsy: A review," *J. Neurosci. Methods*, vol. 368, Feb. 2022, Art. no. 109441.

[7] X. Chen et al., "Recent advances and clinical applications of deep learning in medical image analysis," *Med. Image Anal.*, vol. 79, Jul. 2022, Art. no. 102444.

[8] D. Gutierrez-Barragan, J. S. B. Ramirez, S. Panzeri, T. Xu, and A. Gozzi, "Evolutionarily conserved fMRI network dynamics in the mouse, macaque, and human brain," *Nature Commun.*, vol. 15, no. 1, p. 8518, Oct. 2024.

[9] Y. Chen, Y. Wang, Z. Song, Y. Fan, T. Gao, and X. Tang, "Abnormal white matter changes in Alzheimer's disease based on diffusion tensor imaging: A systematic review," *Ageing Res. Rev.*, vol. 87, Jun. 2023, Art. no. 101911.

[10] Z. Gu, K. W. Jamison, M. R. Sabuncu, and A. Kuceyeski, "Heritability and interindividual variability of regional structure-function coupling," *Nature Commun.*, vol. 12, no. 1, p. 4894, Aug. 2021.

[11] P. Fotiadis, L. Parkes, K. A. Davis, T. D. Satterthwaite, R. T. Shinohara, and D. S. Bassett, "Structure–function coupling in macroscale human brain networks," *Nature Rev. Neurosci.*, vol. 25, no. 10, pp. 688–704, Oct. 2024.

[12] W. Tan, L. Zhu, J. Li, Z. Zhang, and H. Zhang, "Partial multi-modal hashing via neighbor-aware completion learning," *IEEE Trans. Multimedia*, vol. 25, pp. 8499–8510, 2023.

[13] Q. Zhu et al., "Deep multi-modal discriminative and interpretability network for Alzheimer's disease diagnosis," *IEEE Trans. Med. Imag.*, vol. 42, no. 5, pp. 1472–1483, May 2023.

[14] Q. Zhu et al., "Spatio-temporal graph hubness propagation model for dynamic brain network classification," *IEEE Trans. Med. Imag.*, vol. 43, no. 6, pp. 2381–2394, Jun. 2024, doi: 10.1109/TMI.2024.3363014.

[15] F. Liu et al., "Multi-task joint learning network based on adaptive patch pruning for Alzheimer's disease diagnosis and clinical score prediction," *Biomed. Signal Process. Control*, vol. 95, Sep. 2024, Art. no. 106398.

[16] N. A. Zebari et al., "A deep learning fusion model for accurate classification of brain tumours in magnetic resonance images," *CAAI Trans. Intell. Technol.*, vol. 9, no. 4, pp. 790–804, Aug. 2024.

[17] A. D. Dhaygude et al., "Knowledge-based deep learning system for classifying Alzheimer's disease for multi-task learning," *CAAI Trans. Intell. Technol.*, vol. 9, no. 4, pp. 805–820, Aug. 2024.

[18] M. Liu, J. Zhang, P.-T. Yap, and D. Shen, "Diagnosis of Alzheimer's disease using view-aligned hypergraph learning with incomplete multi-modality data," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2016, pp. 308–316.

[19] D. Shen, G. Wu, and H. Suk, "Deep learning in medical image analysis," *Annu. Rev. Biomed. Eng.*, vol. 19, no. 1, pp. 221–248, 2017.

[20] Q. Zhu et al., "Incomplete multi-modal brain image fusion for epilepsy classification," *Inf. Sci.*, vol. 582, pp. 316–333, Jan. 2022.

[21] X. Meng, K. Sun, J. Xu, X. He, and D. Shen, "Multi-modal modality-masked diffusion network for brain MRI synthesis with random modality missing," *IEEE Trans. Med. Imag.*, vol. 43, no. 7, pp. 2587–2598, Jul. 2024, doi: 10.1109/TMI.2024.3368664.

[22] C. R. Jack et al., "The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods," *J. Magn. Reson. Imaging, Off. J. Int. Soc. Magn. Reson. Med.*, vol. 27, no. 4, pp. 685–691, 2008.

[23] J. Huang, L. Zhou, L. Wang, and D. Zhang, "Integrating functional and structural connectivities via diffusion-convolution-bilinear neural network," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2019, pp. 691–699.

[24] Y. Yang, H. Chen, Z. Chang, Y. Xiang, C. Ye, and T. Ma, "Incomplete learning of multi-modal connectome for brain disorder diagnosis via modal-mixup and deep supervision," in *Proc. Med. Imag. Deep Learn.*, 2024, pp. 1006–1018.

[25] Y. Liu et al., "Incomplete multi-modal representation learning for Alzheimer's disease diagnosis," *Med. Image Anal.*, vol. 69, Apr. 2021, Art. no. 101953.

[26] H. Wang, Y. Chen, C. Ma, J. Avery, L. Hull, and G. Carneiro, "Multi-modal learning with missing modality via shared-specific feature modelling," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2023, pp. 15878–15887.

[27] A. Vaswani et al., "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2025, pp. 5998–6008.

[28] Y. Zhang et al., "MmFormer: Multimodal medical transformer for incomplete multimodal learning of brain tumor segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2022, pp. 107–117.

[29] D. Oh, B. Kim, J. Lee, and Y.-G. Shin, "Unsupervised deep learning network with self-attention mechanism for non-rigid registration of 3D brain MR images," *J. Med. Imag. Health Informat.*, vol. 11, no. 3, pp. 736–751, Mar. 2021.

[30] X. Huang et al., "Deep learning-based multiclass brain tissue segmentation in fetal MRIs," *Sensors*, vol. 23, no. 2, p. 655, Jan. 2023.

[31] M. W. Gardner and S. R. Dorling, "Artificial neural networks (the multilayer perceptron)—A review of applications in the atmospheric sciences," *Atmos. Environ.*, vol. 32, no. 14, pp. 2627–2636, Aug. 1998.

[32] S. Kullback and R. A. Leibler, "On information and sufficiency," *Ann. Math. Statist.*, vol. 22, no. 1, pp. 79–86, 1951.

[33] Z. Cui, S. Zhong, P. Xu, Y. He, and G. Gong, "PANDA: A pipeline toolbox for analyzing brain diffusion images," *Frontiers Human Neurosci.*, vol. 7, p. 42, Jun. 2013.

[34] M. Jenkinson, C. F. Beckmann, T. E. J. Behrens, M. W. Woolrich, and S. M. Smith, "FSL," *NeuroImage*, vol. 62, no. 2, pp. 782–790, 2011.

[35] R. Wang, T. Benner, A. G. Sorensen, and V. J. Wedeen, "Diffusion toolkit: A software package for diffusion imaging data processing and tractography," in *Proc. Intl. Soc. Mag. Reson. Med.*, vol. 15, 2007, p. 3720.

[36] J. Ashburner et al., "SPM8 manual," Funct. Imag. Lab., Wellcome Trust Centre Neuroimaging, Univ. College London, London, U.K., Tech. Rep., Feb. 2012.

[37] C. Yan and Y. Zang, "DPARSF: A MATLAB toolbox for 'Pipeline' data analysis of resting-state fMRI," *Frontiers Syst. Neurosci.*, vol. 4, p. 1377, May 2010.

[38] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.

[39] M. Mitra and B. B. Chaudhuri, "Information retrieval from documents: A survey," *Inf. Retr.*, vol. 2, nos. 2–3, pp. 141–163, May 2000.

[40] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognit. Lett.*, vol. 27, no. 8, pp. 861–874, Jun. 2006.

[41] Y.-J. Horng, S.-M. Chen, Y.-C. Chang, and C.-H. Lee, "A new method for fuzzy information retrieval based on fuzzy hierarchical clustering and fuzzy inference techniques," *IEEE Trans. Fuzzy Syst.*, vol. 13, no. 2, pp. 216–228, Apr. 2005.

[42] Y. Dong and C.-Y.-J. Peng, "Principled missing data methods for researchers," *SpringerPlus*, vol. 2, no. 1, p. 222, Dec. 2013.

[43] P. Xia, L. Zhang, and F. Li, "Learning similarity with cosine similarity ensemble," *Inf. Sci.*, vol. 307, pp. 39–52, Jun. 2015.

[44] D. Chicco, M. J. Warrens, and G. Jurman, "The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation," *PeerJ Comput. Sci.*, vol. 7, p. 623, Jul. 2021.

[45] A. Hore and D. Ziou, "Image quality metrics: PSNR vs. SSIM," in *Proc. 20th Int. Conf. Pattern Recognit.*, Aug. 2010, pp. 2366–2369.

[46] Q. Zhu, J. Huang, and X. Xu, "Non-negative discriminative brain functional connectivity for identifying schizophrenia on resting-state fMRI," *Biomed. Eng. OnLine*, vol. 17, no. 1, pp. 1–15, Dec. 2018.

[47] J. Zhang, Y. Zhang, J. Wang, Y. Xia, J. Zhang, and L. Chen, "Recent advances in Alzheimer's disease: Mechanisms, clinical trials and new drug development strategies," *Signal Transduction Targeted Therapy*, vol. 9, no. 1, p. 211, Aug. 2024.

[48] Z. Chen, M. J. Brodie, D. Ding, and P. Kwan, "Epidemiology of epilepsy and seizures," *Frontiers Epidemiol.*, vol. 3, p. 1273163, Jun. 2023.

[49] S. G. Mueller, K. D. Laxer, J. Barakos, I. Cheong, P. Garcia, and M. W. Weiner, "Widespread neocortical abnormalities in temporal lobe epilepsy with and without mesial sclerosis," *NeuroImage*, vol. 46, no. 2, pp. 353–359, Jun. 2009.

[50] C. L. Grady, "Altered brain functional connectivity and impaired short-term memory in Alzheimer's disease," *Brain*, vol. 124, no. 4, pp. 739–756, Apr. 2001.

[51] J. Zhou et al., "Divergent network connectivity changes in behavioural variant frontotemporal dementia and Alzheimer's disease," *Brain*, vol. 133, no. 5, pp. 1352–1367, 2010.