

DCL-SE: Dynamic Curriculum Learning for Spatiotemporal Encoding of Brain Imaging

Meihua Zhou, Xinyu Tong, Jiarui Zhao, Min Cheng, Li Yang, Lei Tian, and Nan Wan

Abstract—High-dimensional neuroimaging analyses for clinical diagnosis are often constrained by compromises in spatiotemporal fidelity and the limited adaptability of large-scale, general-purpose models. To address these challenges, we introduce Dynamic Curriculum Learning for Spatiotemporal Encoding (DCL-SE), an end-to-end framework centered on data-driven spatiotemporal encoding (DaSE). We leverage Approximate Rank Pooling (ARP) to efficiently encode three-dimensional volumetric brain data into information-rich, two-dimensional dynamic representations, and then employ a dynamic curriculum learning strategy—guided by a Dynamic Group Mechanism (DGM)—to progressively train the decoder, refining feature extraction from global anatomical structures to fine pathological details. Evaluated across six publicly available datasets—including Alzheimer’s disease and brain tumor classification, cerebral artery segmentation, and brain age prediction—DCL-SE consistently outperforms existing methods in accuracy, robustness, and interpretability. These findings underscore the critical importance of compact, task-specific architectures in the era of large-scale pretrained networks.

Index Terms—Neuroimaging; Dynamic Curriculum Learning; Spatiotemporal Encoding; Dynamic Group Mechanism; Interpretability.

I. INTRODUCTION

NEUROLOGICAL disorders such as Alzheimer’s disease (AD) and brain tumors remain critical, unsolved challenges in clinical medicine, where early and accurate diagnosis can profoundly influence patient outcomes [1]–[3]. Modern neuroimaging modalities—especially Magnetic Resonance Imaging (MRI) and Computed Tomography (CT)—enable detailed, noninvasive visualization of brain structure [4]–[6]. However, translating such high-dimensional imaging data into robust, actionable diagnostic decisions remains challenging due to subtle clinical signals, spatial heterogeneity, and limited annotated datasets [7].

A common strategy to alleviate computational burdens in neuroimaging is to project 3D brain volumes into 2D slices for analysis [8]–[10]. While computationally efficient,

Manuscript received April 19, 2021; revised August 16, 2021. (Corresponding authors: Lei Tian and Nan Wan)

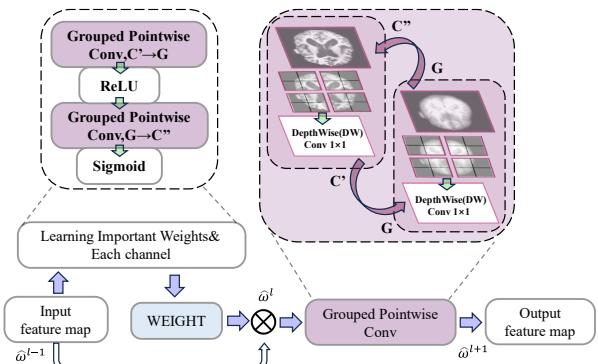
Meihua Zhou, Xinyu Tong and Jiarui Zhao are graduate students at the University of Chinese Academy of Sciences, Beijing 100049, China. (e-mail: mhzhou0412@gmail.com; tonxyc@gmail.com; zhaojiarui@binn.cas.cn)

Min Cheng is with the Information Department, Wuhu Hospital, Beijing Anding Hospital, Capital Medical University, Wuhu, Anhui 241000, China. (e-mail: minchenginfo@163.com)

Lei Tian is a Professor and Deputy Chief Physician with Beijing Tongren Hospital, Capital Medical University, and the Beijing Institute of Ophthalmology, Beijing 100730, China. He is also a postgraduate advisor. (e-mail: tianlei0131@163.com)

Li Yang and Nan Wan is an Associate Professor with Wannan Medical university, Wuhu, Anhui 241002, China. (e-mail: wannan@wnmc.edu.cn; yangli@wnmc.edu.cn)

this approach incurs a critical trade-off: the loss of spatial continuity and essential anatomical relationships within the brain. Such simplification disrupts global spatial coherence, complicating the capture of critical pathological patterns, especially when lesions or atrophic regions are sparse or irregularly distributed. Moreover, transferring information from 3D to 2D exacerbates the challenge by losing both spatial and semantic context, which are crucial for accurate disease characterization [11], [12]. This issue is further compounded by the limitations of traditional learning strategies, such as Curriculum Learning (CL). These methods often rely on static difficulty levels, which fail to adapt to the complexities inherent in neuroimaging data. Conventional CL methods use manually defined stages that do not dynamically adjust to varying feature complexities, thus hindering the model’s ability to progressively learn from simple to complex features. As both spatial relationships and pathologies evolve throughout the learning process, this static approach leads to models failing to disentangle fine-grained features, resulting in suboptimal performance. Consequently, there is a clear need for a new approach—one that not only bridges the cross-modal gap between 3D and 2D but also incorporates an adaptive, dynamic learning process that progressively captures and refines diagnostically relevant features (See the related work section below for further details.).



Notes: C&G:Come (Initial channel), Go (Each channel using the Channel Attention Mechanism)

Fig. 1. Dynamic Grouping Mechanism (DGM), an important component of DCL-SE and the carrier of Dynamic Curriculum Learning (DCL), centralizes DCL by generating, weighting, and adjusting dynamic features. These features are then processed through grouped convolution to extract critical channel-wise information.

Recent advances in large-scale neural network models have promoted the development of universal, transferable architectures, including vision and multimodal frameworks such as MedSAM [15], BioGPT [16], and MedCLIP [17], demonstrating promising cross-domain generalization and zero-shot

learning capabilities [18]. Nevertheless, the practical utility of these large models in neuroimaging remains constrained by data scarcity, stringent privacy requirements, and significant computational costs. In our systematic evaluation of eleven state-of-the-art large models (GPT-4 [19], Gemini [20], Claude [21], Pixtral [22], ERNIE [23], Flux [24], GLM [25], HunYuan-lite [26], Moonshot-V1 [27], among others) on Alzheimer’s disease MRI tasks via direct zero-shot inference, we find these off-the-shelf architectures consistently underperform specialized domain-specific frameworks—even with advanced prompt engineering. This observation reveals a critical gap between theoretical generalization advances and practical application constraints, motivating tailored lightweight solutions designed explicitly for clinical neuroimaging tasks.

To address these pressing challenges, we introduce Dynamic Curriculum Learning for Spatiotemporal Encoding (DCL-SE), an end-to-end lightweight framework specifically developed for robust and interpretable brain imaging analysis. Central to DCL-SE is the concept of Data-based Spatiotemporal Encoding (DaSE), which comprises two sequential stages: an encoding stage and a decoding stage. In the encoding stage, DaSE employs Approximate Rank Pooling (ARP) to efficiently distill three-dimensional volumetric brain data into compact yet information-rich two-dimensional dynamic representations, thus preserving crucial spatial and anatomical progression contexts essential for accurate clinical diagnosis. However, directly training on these ARP projections may lead to entangled representations and obscure subtle pathological cues. To address this issue, the decoding stage of DaSE adopts a dynamic curriculum learning strategy guided by a Dynamic Group Mechanism (DGM), which progressively disentangles and refines diagnostically salient features from the encoded representations. As illustrated in Figure 1, DGM adaptively recalibrates feature importance, mimicking the hierarchical reasoning process of clinical experts—initially focusing on global anatomical structures and subsequently refining attention toward subtle pathological details.

Beyond disease classification, we demonstrate the versatility of DCL-SE across multiple neuroimaging tasks, including brain tissue segmentation and age prediction, accompanied by comprehensive interpretability analyses via t-SNE projections, heatmaps, and saliency mapping. Extensive experiments on diverse public datasets confirm that DCL-SE consistently surpasses existing conventional architectures and recent large-scale pretrained models in accuracy, efficiency, and interpretability.

The main contributions of this work are summarized as follows:

- 1) We develop the DCL-SE framework, a lightweight adaptive model designed to dynamically prioritize feature extraction according to complexity and clinical significance.
- 2) Proposal of DaSE, a data-driven spatiotemporal encoding method composed of two sequential stages: encoding via ARP to transform volumetric brain data into compact dynamic representations, and decoding via Dynamic Curriculum Learning guided by the DGM.
- 3) Comprehensive validation of DCL-SE’s robustness and interpretability across representative neuroimaging

tasks—including brain disease classification, cerebral artery segmentation, and brain age prediction—providing extensive visual evidence and analyses supporting its applicability.

4) A systematic empirical evaluation revealing the significant performance limitations of large general-purpose models in practical clinical neuroimaging scenarios characterized by limited samples and stringent privacy constraints.

II. RELATED WORK

A. 2D, 3D, and Hybrid Approaches for Brain Image Analysis

Brain image analysis has advanced along three major lines: 2D slice-based methods, 3D volumetric models, and hybrid 2D-3D approaches [28]–[31]. 2D methods are widely adopted due to computational efficiency but fail to capture global anatomical context, leading to limited performance in complex tasks [32]. 3D models offer improved spatial representation and have achieved success in segmentation and disease detection [33], [34], but their high computational and data requirements restrict clinical utility. Hybrid approaches—such as multi-view fusion and 2.5D methods that aggregate orthogonal 2D slices—aim to balance computational efficiency with spatial awareness by approximating 3D context through coordinated 2D representations [35], [36]. Some methods also incorporate interpolation to reconstruct pseudo-3D volumes from sparse slices. However, these strategies often rely on static fusion rules or shallow aggregation schemes, which may lead to loss of spatial fidelity, introduce interpolation artifacts, or fail to capture heterogeneity in small-sample neuroimaging datasets. Therefore, designing models that efficiently preserve diagnostically critical spatial information remains a key challenge in brain image analysis..

B. Transfer Learning and Foundation Models in Medical Imaging

Transfer learning has become a cornerstone of medical image analysis, particularly under the constraints of limited labeled data and stringent privacy requirements [37]–[39]. Conventional approaches are most effective when applied within the same modality and dimensionality—such as 2D-to-2D or 3D-to-3D tasks—where feature representations are naturally aligned and semantic gaps are minimized. In contrast, cross-modal transfer learning, such as 3D-to-2D adaptation or transferring between imaging modalities (e.g., CT to MRI), remains challenging due to substantial differences in spatial context, resolution, and underlying data distributions. Direct projection or feature transfer from 3D volumes to 2D slices often results in significant information loss, limiting the model’s capacity to generalize to clinically relevant tasks [40].

To mitigate these limitations, a variety of advanced strategies have been proposed. Spiral-spinning “2.75D” representations, for instance, attempt to encode volumetric structure within 2D feature spaces, aiming to retain spatial information while reducing computational cost [41]. Meanwhile, self-supervised learning and hybrid initialization techniques have been introduced to enhance feature robustness and address overfitting in low-resource settings [42], [43]. Nevertheless, excessive similarity between source and target datasets can

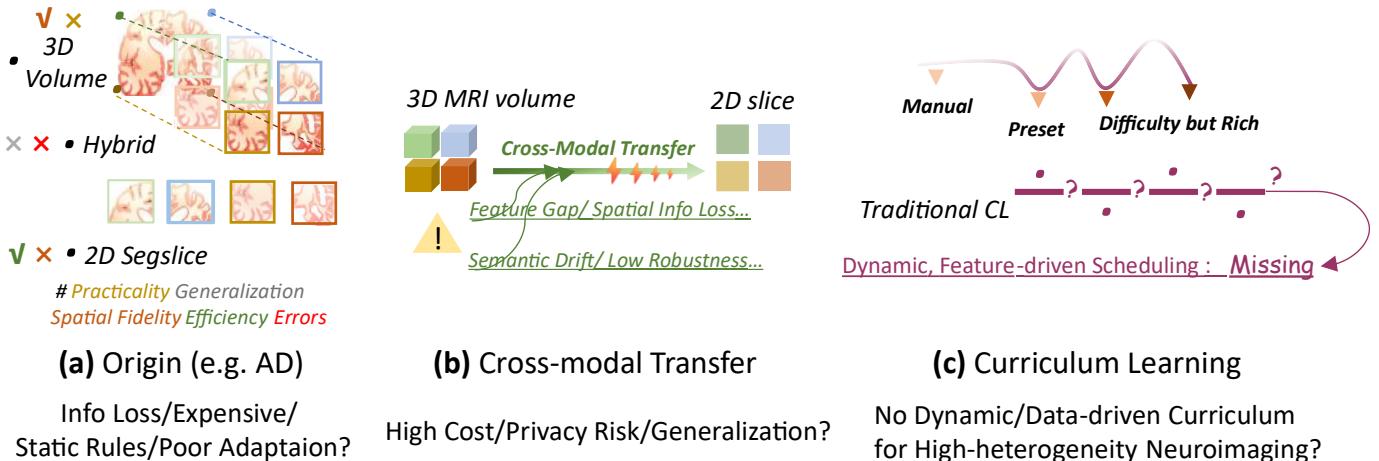


Fig. 2. Illustration of existing multimodal image data processing and learning strategies. (a) Static 2D, 3D, and hybrid methods are widely used in brain imaging analysis but each has limitations: 2D methods are highly efficient yet lack spatial information [8], [11]. 3D methods offer strong spatial expression but are constrained by computing power and sample size, limiting clinical practicality [13], [31]. hybrid methods integrate 2D and 3D via static fusion rules or post-processing but are prone to interpolation errors and poor generalizability [7], [14]. (b) Cross-modal transfer, including 3D to 2D and CT to MRI conversions, suffers from core issues of spatial and semantic information loss during transfer [37], [38], [40]. While large/foundational models enable such transfer, they face bottlenecks of high computational demands and privacy risks, with their practical generalization and information retention remaining unresolved [16], [17]. (c) Curriculum Learning (CL) has notable limitations: most applications rely on static, manually defined difficulty classification, lacking feature-driven dynamic scheduling [4], [10], [29]. In heterogeneous brain imaging scenarios specifically, the absence of automatic complexity measurement and adaptation mechanisms hinders gradual improvement of diagnostic complexity and generalization [9], [30].

cause “inbreeding effects,” where iterative reuse of datasets degrades the benefits of transfer learning and impairs model generalization.

More recently, the emergence of foundation models—including MedSAM [15], pre-trained backbones for medical imaging. Trained on large-scale, heterogeneous datasets, these models offer strong potential for adaptation to downstream clinical tasks via fine-tuning. Vision-language and multimodal models, such as BioGPT [16], MedCLIP [17], and DINOv2 [44], further extend this paradigm by incorporating cross-domain knowledge. However, the practical deployment of such models is limited by substantial computational requirements, the scarcity of large annotated clinical datasets, and persistent privacy concerns. Critically, their efficacy in cross-modal and cross-dimensional transfer scenarios—such as 3D-to-2D adaptation or integration across different imaging modalities—remains an open problem, particularly regarding information retention, semantic alignment, and computational efficiency.

C. Curriculum Learning

Curriculum learning (CL) organizes training from easy to hard examples, and has shown effectiveness in computer vision by improving convergence and generalization [45]–[48]. In medical imaging, existing CL strategies are rare and mostly use static or heuristic difficulty criteria, such as image blurring or staged sample complexity [49]. However, these methods generally lack principled, data-driven metrics for task or feature complexity, and do not adaptively adjust curricula during training.

This limitation is especially apparent in brain imaging, where high heterogeneity and diverse tasks make static curricula insufficient. To date, automatic quantification of sample

complexity and dynamic scheduling remain largely unexplored in medical image analysis, particularly for complex scenarios such as subject-level brain image classification, segmentation, and age prediction.

III. METHODOLOGY

A. Framework Overview

Accurate subject-level analysis of brain images is fundamentally challenged by limited labeled data, high anatomical heterogeneity, and the need to balance computational efficiency with preservation of 3D spatial context. Conventional 2D or 3D models typically either sacrifice global anatomical information or incur prohibitive computational costs, while static learning pipelines do not reflect the progressive and hierarchical nature of clinical diagnostic reasoning.

To address these limitations, (See Figure 3b) we propose Dynamic Curriculum Learning for Spatiotemporal Encoding (DCL-SE), a unified framework designed for efficient and interpretable analysis of heterogeneous, data-limited brain imaging scenarios. (See Figure 3a) The core of DCL-SE is the Data-based Spatiotemporal Encoding (DaSE) paradigm, which serves as the main architectural pipeline throughout the framework.

DaSE is composed of two tightly connected stages. In the first stage, Approximate Rank Pooling (ARP) is employed to encode the ordered spatial and temporal progression of 3D brain MRI data into a compact two-dimensional dynamic representation. This encoding step preserves diagnostically relevant spatial structure and progression cues with minimal information loss. In the second stage, the ARP-encoded dynamic image is decoded through a dynamic curriculum learning (DCL) strategy, which is implemented using a Dynamic

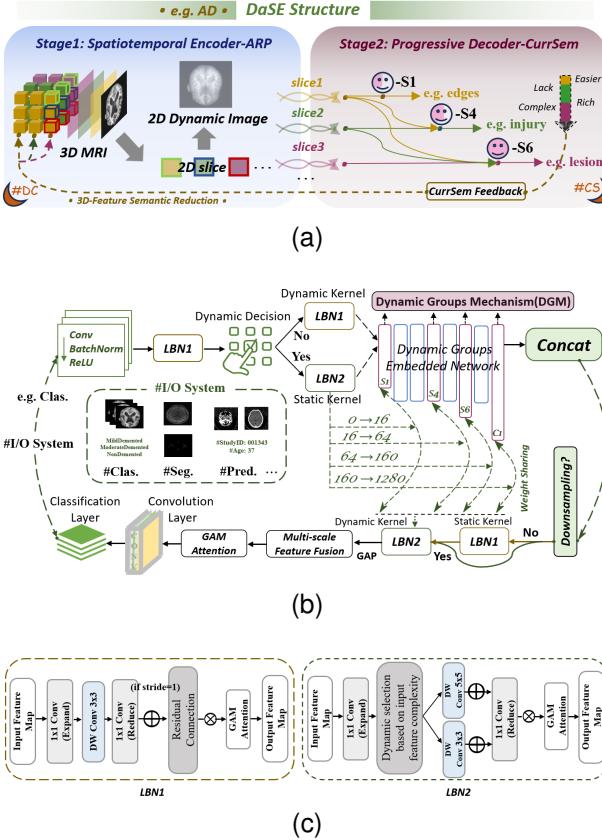


Fig. 3. (a) Data-based Spatiotemporal Encoding (DaSE) structure, which consists of two stages: ARP-based Spatiotemporal Encoding and Curriculum Semantic-based Progressive Decoding. For example, in Alzheimer's disease (AD), the former is a Data Conversion (DC) process that first converts the 3D MRI data into 2D slices via ARP to generate 2D dynamic images, which are used as inputs for the latter Curriculum Semantic (CurrSem) phase. At the CurrSem (CS) process, these 2D slices are then trained to decode multidimensionally through the Curriculum Learning strategy. Finally, the trained features are fed back to the 3D information via CurrSem to greatly reduce information loss between medical multimodal data under the condition of effective training of the 3D data features. (b) DCL-SE architecture, featuring a distinctive "D" shape that represents its Dynamic Adjustment capabilities. The overall model design provides theoretical support for the decoding stage of DaSE. Specifically, the dynamic grouping mechanism (DGM) integrates the DCL strategy to perform progressive hierarchical feature extraction and, through enhanced depth separable convolution and linearization processing, improves training efficiency while reducing model size and computational complexity. (c) DCL-SE adaptively selects modules (LinearBottleNeck1 or LinearBottleNeck2) based on network positions (e.g., convolution grouping, downsampling layer placement).

Group Mechanism (DGM). This decoding stage progressively extracts, disentangles, and refines features from coarse anatomical context to fine pathological details, with DGM adaptively recalibrating feature importance at each step based on data complexity.

The tightly introduce encoder-decoder structure of DaSE allows DCL-SE to bridge the gap between computational efficiency and spatial fidelity. By first compressing 3D volumetric data and then decoding it in a staged and adaptive manner, the framework enables standard 2D neural networks to robustly capture and utilize the rich diagnostic information present in three-dimensional brain images. This approach is widely applicable to a variety of neuroimaging tasks, including disease classification, tissue segmentation, and regression-based predictions such as brain age.

B. Data-based Spatiotemporal Encoding

This section explores the scientific rationale underlying our approach, with a focus on the unified encoding-decoding framework that forms the foundation of our design. At the heart of this methodology is the Data-based Spatiotemporal Encoding (DaSE) paradigm, which is essential for preserving the integrity of both spatial and temporal information throughout the entire process. DaSE acts as the cornerstone of the framework, offering an efficient mechanism to encode 3D brain MRI data into 2D representations, all while retaining critical anatomical and pathological features. In the following two sections, we provide a detailed breakdown of the overall architecture and technical components of our framework, including Approximate Rank Pooling (ARP), Dynamic Curriculum Learning (DCL), and the Dynamic Group Mechanism (DGM).

A persistent challenge in brain MRI analysis is how to maintain the integrity of three-dimensional anatomical and pathological information while ensuring that the model remains practical and interpretable. In clinical settings, most efficient neural network frameworks and established diagnostic workflows rely on two-dimensional representations. However, direct 2D analysis inevitably loses a significant amount of contextual detail that defines disease progression and spatial relationships within the brain. Additionally, even when data is processed with minimal semantic loss, the training process often fails to achieve precise matching, resulting in low training efficiency despite improved data processing interventions. Consequently, the model's outcomes fall short of expectations.

To bridge this gap, our DaSE structure (See Figure 3a) is built around a data-driven spatiotemporal encoding-decoding strategy that closely reflects the underlying scientific questions in medical imaging (See Algorithm 1). We view it as a critical stage that must be designed to minimize the loss of meaningful structure and dynamics.

Our method begins with an encoding process: Approximate Rank Pooling (ARP) projects the ordered stack of 3D MRI slices into a single 2D dynamic image, designed to retain both the spatial structure and the progression of anatomical changes across the volume. This ensures that diagnostically relevant temporal and spatial patterns are not discarded, but instead made accessible to subsequent 2D processing.

However, the complexity of real brain data means that these compressed dynamic images still contain highly entangled features. Important pathological cues, subtle tissue boundaries, and global context may overlap in ways that challenge direct interpretation by standard convolutional networks. Recognizing this, we introduce a decoding process based on Dynamic Curriculum Learning (DCL), supported by a Dynamic Group Mechanism (DGM). This stepwise decoding does not simply “unpack” the compressed image, but instead guides the model to progressively disentangle and interpret features from coarse, easily recognizable structures to fine-grained pathological signatures, much as a clinician would approach a complex case.

By explicitly combining the encoding and decoding stages, our framework not only reduces conversion losses in multimodal data, promotes robust hierarchical learning and model

Algorithm 1 Data-based Spatiotemporal Encoding

Input: Ordered 3D MRI slices $\{\mathbf{I}_1, \dots, \mathbf{I}_T\}$
Output: Task-specific prediction (e.g., brain image classification, brain image segmentation, brain age prediction)

- 1: // Step 1: Spatiotemporal Encoding (ARP)
- 2: **for** $t = 1$ to T **do**
- 3: Extract feature vector $\psi_t \leftarrow f_{\text{enc}}(\mathbf{I}_t)$
- 4: **end for**
- 5: // Step 1.2: Aggregate slice features with ARP to obtain the dynamic 2D descriptor
- 6: $\mathbf{D}^* \leftarrow \text{ARP}(\{\psi_1, \dots, \psi_T\})$
- 7: // Step 2: Progressive Decoding (DCL with DGM)
- 8: Initialize decoded feature $\mathbf{D}_0 \leftarrow \mathbf{D}^*$
- 9: **for** stage $i = 1$ to N **do**
- 10: Stage-wise feature extraction and complexity evaluation
- 11: $\mathbf{D}_i \leftarrow \text{DCLStage}(\mathbf{D}_{i-1})$
- 12: Compute complexity $\lambda(\mathbf{D}_i)$
- 13: **if** $\lambda(\mathbf{D}_i) > \tau_i$ **then**
- 14: Apply dynamic group mechanism for recalibration
- 15: $\mathbf{D}_i \leftarrow \text{DGM}(\mathbf{D}_i)$
- 16: Proceed to next stage
- 17: **else**
- 18: Continue current stage refinement
- 19: **end if**
- 20: **end for**
- 21: // Step 3: Final Prediction
- 22: $\hat{y} \leftarrow \text{Predict}(\mathbf{D}_N)$
- 23: **return** \hat{y}

transparency, but also flexibly retains and transmits the semantic information of multidimensional features after training through Curriculum Semantic Feedback (CurrSem Feedback), thereby maximizing the positive feedback regulation effect throughout the entire data processing and training process. This method, which we refer to as Data-driven Spatio-temporal Encoding (DaSE), enables actual 2D neural networks to fully leverage the available information depth in multi-dimensional brain MRI, ultimately supporting accurate, interpretable, and clinically meaningful analysis.

C. Dynamic Image Construction and Cross-Modal Transfer via Approximate Rank Pooling

Analysis of medical images often involves bridging substantial gaps between different data organizations and modalities. In conventional medical imaging research, “cross-modal” generally refers to the transfer of knowledge between different imaging mechanisms such as CT and MRI, or between various MRI sequences (e.g., T1- and T2-weighted images). However, a significant and often overlooked modality gap also exists between 2D MRI—typically stored as individual DICOM slices—and 3D volumetric MRI, such as NIFTI (.nii) data, which contain richer spatial context and different organizational structures. Beyond modality, there is also a dimensionality gap: models designed for 2D images are not

directly compatible with 3D data due to mismatches in spatial continuity and feature distribution.

Our approach addresses both the cross-dimensional and cross-modal transfer challenges inherent in brain imaging analysis. The core scientific problem is how to exploit robust, well-generalized feature representations learned from abundant 2D annotated data for efficient and accurate analysis of volumetric 3D MRI, while preserving critical spatial and sequential (structural) information. Traditional pooling strategies, such as mean or max pooling across slices, are order-invariant and tend to obscure essential anatomical progressions and dynamic relationships, limiting their effectiveness for downstream volumetric tasks.

To overcome these challenges, we propose to summarize the spatial evolution of 3D MRI data using approximate rank pooling (ARP), a principled, order-sensitive aggregation technique originally developed for temporal video modeling [50], [51]. ARP enables the conversion of an ordered sequence of slices into a single “dynamic image”—a compact descriptor that preserves both appearance and spatial progression information. This makes it possible to leverage standard 2D CNN architectures and pretrained weights for volumetric medical data, thus bridging both the dimensional and modality gap efficiently.

Given a 3D MRI volume as an ordered sequence of T slices $\{\mathbf{I}_1, \dots, \mathbf{I}_T\}$, each slice is mapped to a feature vector $\psi_t \in \mathbb{R}^d$. The task is to construct a summary descriptor \mathbf{d}^* that captures the structural progression along the anatomical axis. Rather than simply averaging features, ARP models the sequence as a ranking problem. The idea is to learn a function that assigns higher scores to later slices, capturing the directional change along the volume.

The optimization problem for rank pooling is defined as:

$$\min_{\mathbf{d}} \frac{\lambda}{2} \|\mathbf{d}\|^2 + \frac{2}{T(T-1)} \sum_{q>t} \max\{0, 1 - S(q|\mathbf{d}) + S(t|\mathbf{d})\} \quad (1)$$

where $S(t|\mathbf{d}) = \langle \mathbf{d}, \mathbf{V}_t \rangle$ and $\mathbf{V}_t = \frac{1}{t} \sum_{i=1}^t \psi_i$ is the average of features up to slice t . This formulation, inspired by the RankSVM approach, encourages the summary vector \mathbf{d} to reflect the ordered progression through the volume.

Direct optimization can be computationally intensive and not amenable to end-to-end training. Approximate rank pooling provides an efficient solution by taking a single gradient step from the origin, resulting in:

$$\mathbf{d}^* \propto \sum_{q>t} (\mathbf{V}_q - \mathbf{V}_t) \quad (2)$$

This can be further expanded as:

$$\mathbf{d}^* \propto \sum_{q>t} \left[\frac{1}{q} \sum_{i=1}^q \psi_i - \frac{1}{t} \sum_{j=1}^t \psi_j \right] \quad (3)$$

which is equivalent to a weighted sum:

$$\mathbf{d}^* = \sum_{t=1}^T \alpha_t \psi_t \quad (4)$$

where the coefficients α_t are analytically defined as:

$$\alpha_t = 2(T - t + 1) - (T + 1)(H_T - H_{t-1}) \quad (5)$$

and $H_t = \sum_{i=1}^t 1/i$ is the t -th harmonic number ($H_0 = 0$).

Here, \mathbf{d}^* serves as a 2D dynamic descriptor summarizing the 3D volume; \mathbf{V}_q and \mathbf{V}_t are the mean feature vectors up to slices q and t ; and ψ_i represents the feature at index i . This order-sensitive aggregation preserves both the spatial structure and sequential relationships of the original data, enabling direct compatibility with conventional 2D convolutional architectures. The resulting dimensionality alignment allows the weights pretrained on 2D data to be transferred efficiently to volumetric tasks, achieving both computational speed and accurate structural modeling.

The process is detailed in the following Algorithm 2 :

Algorithm 2 Approximate Rank Pooling

Input: Ordered slice features $\{\psi_1, \dots, \psi_T\}$

Output: Dynamic descriptor \mathbf{d}^*

```

1: Initialize harmonic numbers:  $H_0 \leftarrow 0$ 
2: for  $t = 1$  to  $T$  do
3:    $H_t \leftarrow H_{t-1} + 1/t$       // Compute harmonic numbers
4: end for
5: Initialize descriptor:  $\mathbf{d}^* \leftarrow \mathbf{0}$ 
6: for  $t = 1$  to  $T$  do
7:    $\alpha_t \leftarrow 2(T - t + 1) - (T + 1)(H_T - H_{t-1})$ 
8:    $\mathbf{d}^* \leftarrow \mathbf{d}^* + \alpha_t \psi_t$           // Weighted aggregation
9: end for
10: return  $\mathbf{d}^*$ 

```

The ARP operator thus generates a compact dynamic representation from volumetric data that is compatible with standard 2D convolutional architectures and preserves essential spatial and sequential information for subsequent analysis. This compressed dynamic image serves as the input to our staged learning process, where the network progressively decodes the embedded spatiotemporal features through dynamic curriculum learning, as detailed in Section III-D.

D. Dynamic Curriculum Learning and Dynamic Group Mechanism

In medical brain imaging, accurate diagnosis relies on a progressive understanding of anatomical structures and pathological features, from broad tissue types to subtle disease-specific cues. Motivated by this, we introduce Dynamic Curriculum Learning (DCL), which mimics the clinical reasoning process by gradually increasing the difficulty of the model's learning objectives. In DCL, each model stage (e.g., S1, S4, S6, Conv1) is associated with a different level of receptive field and anatomical abstraction. For instance, early stages (S1, S4) primarily capture global tissue structures such as gray/white matter and ventricles, while later stages (S6, Conv1) progressively capture fine boundaries, lesions, and focal pathology. The transitions between stages are governed by the complexity of learned feature maps, ensuring that the model only advances once it has sufficiently captured information at the current scale.

This approach is central to our proposed DCL-SE framework, which aims to decode the compressed spatiotemporal information from ARP projections in a stepwise manner that mimics clinical diagnostic logic. By progressively guiding the network from global anatomical context to local pathological details, DCL-SE directly addresses the scientific challenge of disentangling and reconstructing multi-scale 3D features from a compact 2D representation.

The complexity of a feature map at stage i is quantitatively defined as

$$\lambda(X_i) = \sum_{c=1}^C \left| \nabla X_i^{(c)} \right| \quad (6)$$

where $X_i \in \mathbb{R}^{C \times H \times W}$, C is the number of channels, and $\nabla X_i^{(c)}$ denotes the spatial gradient in channel c . This metric reflects spatial richness and structural heterogeneity, which, in brain MRI, correspond to anatomical boundaries or regions of abnormal intensity. When $\lambda(X_i)$ exceeds a threshold τ_k (determined empirically or by the distribution of training complexities), the model transitions to the next curriculum stage. This mechanism enforces an adaptive, data-driven curriculum, rather than a static training order.

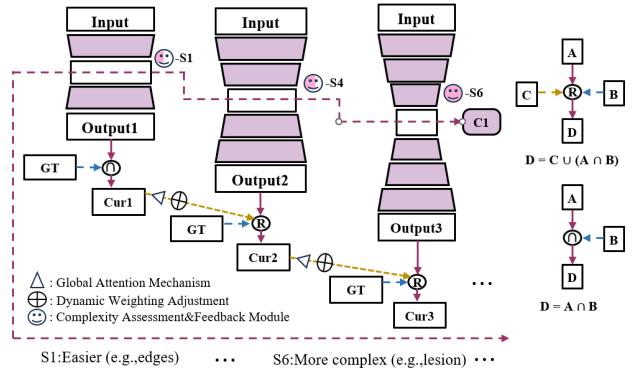


Fig. 4. Dynamic Curriculum Learning (DCL) runs through the whole process of S1,S4,S6,C1, the feature learning process is from simple to complex, and the complexity is evaluated and feedback is given at each stage. Feature fusion on (Cur.) intersection and merger sets, weighted merging of features from different stages to form a multidimensional important and comprehensive information representation through global attention approach.

To enable stage-wise adaptive learning, we introduce the Dynamic Group Mechanism (DGM), which recalibrates feature importance at each stage. DGM is defined by the following computational principle:

$$\begin{cases} \hat{w}^l = LIW(\text{Sum}(\hat{w}^{l-1})) \times \hat{w}^{l-1} \\ \hat{w}^{l+1} = GPCconv(\hat{w}^l) \end{cases} \quad (7)$$

Here, $LIW(\cdot)$ (Learned Importance Weighting) denotes an adaptive weighting computed from the aggregated feature response of the previous stage, operationalized as a grouped convolution followed by a non-linearity and a sigmoid normalization, as in:

$$\begin{cases} W = \sigma(\text{Conv}_G(\text{ReLU}(\text{Conv}_G(X)))) \\ X' = X \odot W \\ Y = \text{Conv}_G(X') \end{cases} \quad (8)$$

where Conv_G is the grouped convolution, W is the adaptive weight map highlighting diagnostically relevant features, X' is the reweighted feature, and Y is the stage output. The DGM thereby ensures that the features with high clinical relevance (such as edges, boundaries, or lesions) are adaptively prioritized at each curriculum stage.

Figure 4 illustrates how DCL runs across S1, S4, S6, and Conv1, with the feature complexity guiding the stage transitions, and feature fusion across curriculum stages yielding a comprehensive information representation. At each step, DGM adapts channel and spatial attention, resulting in efficient feature learning and enhanced diagnostic focus.

For efficiency, DGM employs grouped convolution, which reduces the parameter count compared to standard convolution. Specifically, for input channels C_{in} , output channels C_{out} , kernel size K , and G groups, the parameter count is

$$\text{Params}_{\text{GroupConv}} = \frac{C_{in} \cdot C_{out} \cdot K^2}{G} \quad (9)$$

Setting $G = 4$ for example, the parameter count is reduced by a factor of 4, making DGM suitable for data-limited medical imaging scenarios.

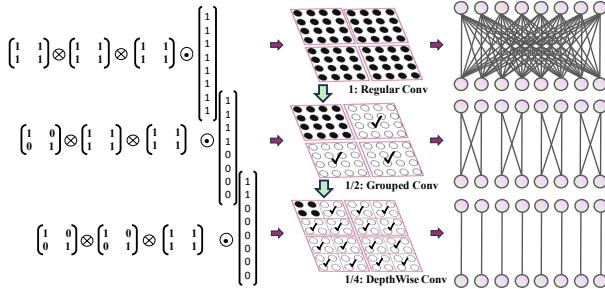


Fig. 5. The generation process of GPConv from the core part of DGM, where operator 1 denotes elementwise product, operate 2 denotes a Kronecker product from left to right.

In essence, the interplay between DCL and DGM constitutes a hierarchical decoding process, in which the model sequentially reconstructs and interprets the multi-scale spatial structure and temporal progression information embedded by ARP in the 2D projection. Through this process, diagnostically critical features are progressively recovered and emphasized during training, effectively overcoming the entanglement and information loss often encountered in direct 2D modeling.

It is crucial to note that the Dynamic Group Mechanism (DGM) operationalises the adaptive nature of DCL. At each curriculum stage, the DGM quantitatively recalibrates channel- and spatial-level attention based on the evolving complexity of features. This ensures that the model's focus remains dynamically aligned with the most salient anatomical or pathological cues throughout the learning process. Such dynamic recalibration is pivotal in achieving the full potential of staged curriculum learning in heterogeneous brain imaging data. The stepwise implementation of DGM is detailed in Algorithm 3, and the difference between standard and grouped convolution is visualized in Figure 5.

Figure 3b illustrates the core model design for the decoding phase of DaSE, highlighting the dynamic curriculum learning

Algorithm 3 Dynamic Group Mechanism

Input: Feature map $X \in \mathbb{R}^{C \times H \times W}$

Output: Transformed feature map Y

```

1: Parameters:
2:  $C$ : Number of input channels
3:  $r$ : Reduction factor
4:  $G$ : Number of groups
5: Step 1: Adaptive Channel Weighting
6:  $C' \leftarrow \max(1, \lfloor C/r \rfloor)$  // Channel reduction
7:  $W \leftarrow \text{GroupedConv}(X, C', G)$  // Feature extraction
8:  $W \leftarrow \text{ReLU}(W)$  // Non-linearity
9:  $W \leftarrow \text{GroupedConv}(W, C, G)$  // Channel expansion
10:  $W \leftarrow \sigma(W)$  // Weight normalization
11: // Dynamic importance scaling
12: Step 2: Dynamic Feature Recalibration
13:  $X' \leftarrow X \odot W$  // Adaptive modulation
14: Step 3: Efficient Feature Fusion
15:  $X'' \leftarrow \text{Pointwise Groups Conv}(X', C, G)$ 
16: // Spatial feature refinement
17:  $Y \leftarrow \text{GroupedPointwiseConv}(X'', C, G)$ 
18: // Final feature aggregation
19: return  $Y$ 

```

strategy developed specifically for spatiotemporal information extraction from encoded representations.

IV. EXPERIMENTS

In this section, we first describe the datasets, evaluation metrics, and implementation details. Then, we present the comparison with latest state-of-the-art methods, and discuss the ablation results.

A. Dataset and Performance Metrics

To comprehensively evaluate the effectiveness and generalizability of the proposed DCL-SE framework, we conducted experiments on a range of public datasets covering classification, segmentation, and regression tasks in brain imaging (A detailed presentation of the dataset can be found in Supplementary Figures 1-7). While classification serves as the main evaluation scenario, segmentation and brain age prediction tasks are included to demonstrate the extensibility of DCL-SE to broader multi-task applications in neuroimaging.

For classification, four datasets were used. The first is a 2D MRI-based Alzheimer's disease dataset from Kaggle, comprising a total of 40,584 images (33,820 for training/validation, 6,764 for testing). The second and third datasets are multimodal brain tumor classification datasets from Kaggle: BT-MRI (3,278 training/validation images, 655 test images) and BT-CT (4,034 training/validation images, 805 test images). The fourth dataset is a 3D NIFTI-format Alzheimer's disease dataset from the ADNI (Alzheimer's Disease Neuroimaging Initiative) database, used for binary classification of 3D volumes.

To further demonstrate the multi-task capability of DCL-SE, we evaluate vessel segmentation and brain age prediction. For segmentation, the CAS2023 dataset from the Cerebral Artery

Segmentation Challenge 2023 is used. This dataset contains 100 annotated 3D TOF-MRA (Time-of-Flight Magnetic Resonance Angiography) scans, with voxel dimensions ranging from (208, 320, 96) to (784, 784, 255). A total of 14,349 slices are available, of which 1,400 containing vessel regions were selected for analysis. In segmentation labels, 0 denotes background and 1 denotes cerebral arteries.

For brain age prediction, we utilize the SPR Head CT Age Prediction Challenge dataset, organized by the Radiology and Diagnostic Imaging Society of São Paulo (SPR). This large-scale public dataset consists of 525,337 head CT scans in DICOM format (total size 276.93 GB), with accompanying patient age information provided in CSV files. The challenge is designed to develop models capable of accurately predicting the patient's age group from head CT images.

Classification performance is measured using accuracy (Acc), area under the receiver operating characteristic curve (AUC), F1-score (F1), precision (Pre), recall (Rec), average precision (AP), number of model parameters (#Params), and floating point operations per second (GFLOPS). For segmentation, the Dice Similarity Coefficient (DSC, also referred to as DICE) is adopted as the main evaluation metric. For brain age prediction, mean absolute error (MAE) is used.

The evaluation metrics are defined as follows:

- Accuracy (Acc):

$$\text{Acc} = \frac{TP + TN}{TP + TN + FP + FN} \quad (10)$$

where TP , TN , FP , and FN denote true positives, true negatives, false positives, and false negatives, respectively.

- Precision (Pre):

$$\text{Pre} = \frac{TP}{TP + FP} \quad (11)$$

- Recall (Rec):

$$\text{Rec} = \frac{TP}{TP + FN} \quad (12)$$

- F1-score (F1):

$$\text{F1} = \frac{2 \cdot \text{Pre} \cdot \text{Rec}}{\text{Pre} + \text{Rec}} \quad (13)$$

- Dice Similarity Coefficient (DSC):

$$\text{DSC} = \frac{2|P \cap G|}{|P| + |G|} \quad (14)$$

where P and G are the sets of predicted and ground truth positive voxels, respectively.

- Mean Absolute Error (MAE):

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i| \quad (15)$$

where y_i and \hat{y}_i denote the true and predicted ages, and N is the number of samples.

These diverse tasks collectively highlight not only the accuracy and efficiency of DCL-SE in standard classification, but also its ability to generalize and adapt to complex multi-task scenarios in neuroimaging analysis.

B. Implementation Details

In our experiments, we primarily used GeForce RTX 3090 and 4090 GPUs equipped with 24GB of memory for brain image classification tasks. As a multi-task extension, We also used NVIDIA Tesla A100 and P100 GPUs for training brain tumor segmentation and comparative experiments on brain age prediction.

For classification tasks, we use the AdamW optimiser with weight decay (0.01) is used in conjunction with a cosine annealing learning rate strategy (initially set to 0.001 and reset every 50 epochs) in order to avoid local optima. 2D data is trained for 300 epochs (batch size 64), while 3D data uses 3-fold cross-validation (200 epochs per fold).

For segmentation tasks, we made targeted adjustments to the classification task training framework to accommodate the characteristics of pixel-level recognition: The AdamW optimiser (weight decay 0.0001, initial learning rate 0.0001) is used in conjunction with a cosine annealing restart (cycle of 10 epochs) and a ReduceLROnPlateau strategy (50% decay after 5 epochs without improvement). The loss function is a weighted combination of $0.5 \times$ Lovász-Softmax loss (category balance), $0.3 \times$ Lovász-Hinge loss (region overlap) and $0.2 \times$ boundary loss (edge accuracy).

The prediction tasks, we use an end-to-end training approach with the AdamW optimiser (learning rate 5e-5, weight decay 1e-5) is optimised using L1 loss (mean absolute error (MAE)) and a OneCycleLR scheduler with 5 epoch warm-up, followed by cosine annealing to 1e-6 to balance convergence speed and parameter tuning.

C. Ablation Studies

The DCL-SE framework is built on the Data-based Spatiotemporal Encoding (DaSE) paradigm, which integrates ARP-based encoding and dynamic curriculum-driven decoding. To clarify the contribution of each component within this tightly-coupled framework, we conduct ablation experiments on the 3D Alzheimer's disease dataset, focusing on three key aspects: spatiotemporal encoding (ARP), staged curriculum decoding (DCL-SE), and dynamic feature recalibration (DGM).

We first evaluate the role of Approximate Rank Pooling (ARP) in DaSE's encoding stage. Unlike order-invariant pooling (e.g., mean or max), ARP preserves anatomical progression and global structural cues, which are essential for subject-level diagnosis. As shown in Table I and Supplementary Tables 1-5, replacing ARP with standard pooling significantly reduces accuracy and F1-score (e.g., F1 drops from 96.6% to 89.2% for max pooling, and to 84.7% for global average pooling). The AUC with ARP reaches 97.1%, substantially outperforming alternatives. This confirms that order-sensitive aggregation is indispensable for encoding clinically relevant spatiotemporal information.

For the decoding stage, we analyze the impact of the dynamic curriculum mechanism. Removing the curriculum strategy causes the F1-score to drop from 96.6% to 88.1%, and AUC from 97.1% to 90.0% (Table II), indicating that progressive, staged learning is critical for disentangling complex

TABLE I
PERFORMANCE COMPARISON OF DIFFERENT POOLING METHODS ON AD 3D DATASET USING DCL-SE MODEL

Method	Acc	Auc	F1	Pre	Recall	Ap
Max Pooling	88.5%	92.3%	89.2%	85.2%	93.5%	83.0%
Mean Pooling	90.1%	89.2%	90.0%	93.1%	87.1%	87.6%
Global Average Pooling	85.2%	89.5%	84.7%	89.2%	80.6%	81.8%
Stochastic Pooling	88.5%	88.3%	88.1%	92.8%	83.8%	86.0%
Spatial Pyramid Pooling	88.5%	91.9%	88.5%	90.0%	87.1%	84.9%
Approximate Rank Pooling	96.4%	97.1%	96.6%	96.6%	96.6%	95.2%

TABLE II
THE PERFORMANCE OF EACH MODEL OF ABLATION EXPERIMENT

Stage				Paradigm				Evaluation				
S1	S4	S6	C1	Dynamic	CL	DCL	F1	Acc	Auc	Ap	#params	GFLOPS
✓	✓	✓	✓	✗	✓	✗	93.6%	93.0%	96.4%	89.4%	7.43M	1.09G
✗	✓	✓	✓	✓	✗	✗	90.9%	89.4%	91.1%	83.3%	8.060M	1.126G
✓	✗	✓	✓	✓	✗	✗	90.0%	89.4%	93.3%	86.2%	8.059M	1.127G
✓	✓	✗	✓	✓	✗	✗	88.1%	87.7%	90.0%	84.7%	8.051M	1.127G
✓	✓	✓	✗	✓	✗	✗	89.6%	89.4%	93.2%	87.4%	7.446M	1.097G
✓	✓	✓	✓	✓	✓	✓	96.6%	96.4%	97.1%	95.2%	8.060M	1.127G

TABLE III
MODEL PERFORMANCE COMPARED ON ALZHEIMER'S DISEASE(AD) AND BRAIN TUMOR 2D DATASET

Method	AD-MRI/Acc	BT-MRI/Acc	BT-CT/Acc	#params	GFLOPS
SwinT	98.00%	97.23%	98.21%	87.70M	15.16G
ConvNeXt	99.36%	98.78%	98.14%	88.54M	15.35G
NiFuse(2D-SOTA)	99.88%	97.02%	98.78%	13.30M	20.23G
Avg.+VGG11+Att [53]	99.80%	99.85%	92.68%	121.87M	30.13G
DAMNet(2D,3D-SOTA) [54]	99.90%	99.69%	98.01%	7.4M	1.1G
DCL-SE	99.94%	99.85%	97.89%	8.0M	1.1G

TABLE IV
MODEL PERFORMANCE COMPARED ON ALZHEIMER'S DISEASE(AD) 3D DATASET. (#BE:BEST EPOCH)

3D-NII/Method	Acc	Auc	F1	Pre	Recall	Ap	#params	GFLOPS
AlexNet [53]	87%	90%	86%	89%	83%	82%	61.10M	0.71G
MobileNetV2 [53]	88%	89%	87%	89%	85%	83%	3.50M	0.33G
ConvNeXt	57.8%	62.3%	64.7%	57.8%	73.3%	56.4%	88.54M	15.35G
NiFuse(2D-SOTA)	73.6%	72.5%	73.6%	77.7%	70.0%	70.2%	13.30M	20.23G
Avg.+VGG11+Att [53]	88.0%	89.0%	88.0%	85.0%	91.0%	82.0%	121.87M	30.13G
DAMNet(2D,3D-SOTA) [54]	93.0%	96.4%	93.6%	90.6%	96.7%	89.4%	7.4M	1.1G
Fold1-(#BE=32)	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	8.0M	1.1G
Fold2-(#BE=47)	94.1%	95.4%	95.2%	100.0%	90.9%	96.7%	8.0M	1.1G
Fold3-(#BE=51)	95.2%	98.1%	95.6%	91.6%	100.0%	91.6%	8.0M	1.1G
DCL-SE	96.4%	97.1%	96.6%	96.6%	95.2%	95.2%	8.0M	1.1G

features and aligning model focus with the hierarchical nature of neuroanatomical progression.

The Dynamic Group Mechanism (DGM) further enhances decoding by adaptively recalibrating channel and spatial attention at each curriculum stage. Removing DGM from any stage results in consistent drops in all metrics, showing that static attention is inadequate for capturing diagnostically subtle features. Partial removal of curriculum stages (e.g., omitting S4 or S6) also degrades performance (F1 and AUC decrease by 6–8%), indicating that multi-scale feature fusion is essential for robust discrimination.

D. Comparative Studies

To rigorously assess the effectiveness and generalizability of DCL-SE, we benchmarked the framework on four representative neuroimaging datasets, covering both 2D (AD-MRI, BT-MRI, BT-CT) and 3D (AD-NII) classification tasks. For fair comparison, state-of-the-art 2D and 3D methods—including SwinT [52], Avg.+VGG11+Att [53], DAMNet [54], NiFuse [55] and ConvNeXt [56] and classic CNNs—were used as baselines.

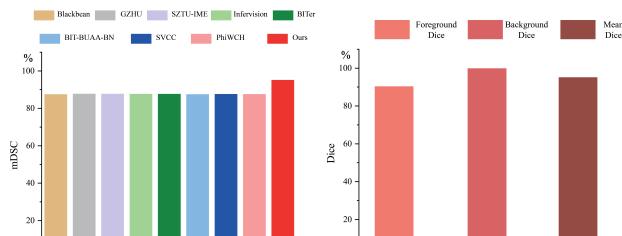
Experimental results (Table III and Table IV) demonstrate that DCL-SE consistently matches or outperforms all baselines across both 2D and 3D settings. Notably, on the AD-NII volumetric dataset, DCL-SE achieves 96.4% accuracy, 97.1% AUC, and a 96.6% F1-score, surpassing advanced 2D-centric and hybrid models, which suffer significant performance drops in 3D contexts. This advantage is directly attributed to the DaSE paradigm: Stage 1 (ARP encoding) preserves spatial progression and anatomical context, while Stage 2 (curriculum-driven decoding with DGM) incrementally disentangles and emphasizes diagnostically relevant features. The synergy between order-sensitive encoding and dynamic, complexity-aware decoding is essential for robust volumetric brain analysis.

The comparative effect of DaSE is further quantified in Table V: only DCL-SE fully implements both ARP encoding and curriculum-based decoding, leading to the best F1 and AUC scores (96.6%, 97.1%) compared to DAMNet (which lacks dynamic curriculum) and ResNet18 (which lacks both). This validates that both structural encoding and staged, adaptive feature refinement are critical to effective multimodal medical image modeling.

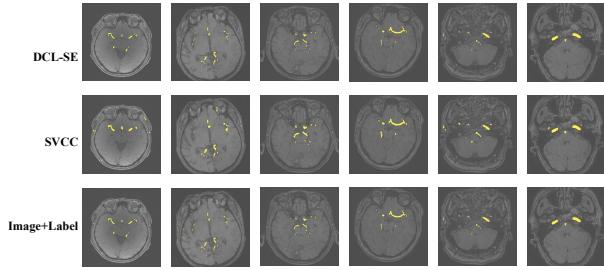
TABLE V

COMPARISON AND ANALYSIS WITH SOTA WORKS ON AD USING THE DASE PROCESSING METHOD.

Models	DaSE Encapsulation Idea		Evaluation	
	#DC	#CS	F1	Auc
			(%)↑	(%)↑
ResNet18	✗	✗	84.0	84.0
DAMNet	✓	✗	93.6	96.4
DCL-SE	✓	✓	96.6	97.1



(a)



(b)

Fig. 6. (a) Vascular segmentation on the CAS2023 3D-MRA (.nii.gz) dataset: The left panel presents mean Dice similarity coefficients (mDSC) for eight methods—Blackbean, GZHU, SZTU-IME, Infervision, BITer, BIT-BUAA-BN, SVCC, and PhiWCH—compared with DCL-SE. The right panel reports DCL-SE's Dice scores for vessel tissue (DSC 1), background (DSC 0), and their mean. All experiments used .nii.gz volumes from the CodaLab Competition (<https://codalab.lisn.upmc.fr/competitions/9804>); full results are in Supplementary Table 6. (b) Qualitative comparison on CAS2023 3D-MRA: columns show different axial slices, with rows for DCL-SE (top), SVCC (middle), and ground-truth annotations (bottom, yellow overlay).

E. Expansion Studies

To evaluate the multi-task generalizability of DCL-SE, we conducted experiments on cerebral artery segmentation and brain age prediction, extending beyond standard classification tasks.

For brain vessel segmentation, DCL-SE was benchmarked on the CAS2023 dataset, achieving the highest mean Dice similarity coefficient (mDSC) among all competing methods (Fig.6a), and demonstrating more anatomically precise boundaries than strong baselines (Fig.6b). More details can be found in Supplementary Figures 8-12. For brain age prediction, DCL-SE outperformed six widely-used open-source models on the SPR Head CT Age Prediction Challenge dataset, achieving the lowest mean absolute error (MAE) despite large-scale data heterogeneity. Notably, the MAE of DCL-SE was 9.47 lower than that of FiA-Net, which previously exhibited strong performance on this task (see Supplementary Table 7 for full results).

To further contextualize the strengths and limitations of

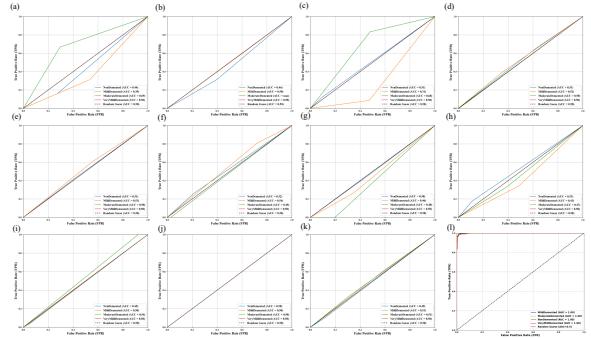


Fig. 7. Receiver-operating characteristic (ROC) curves for eleven off-the-shelf, non-fine-tuned large language models—ERNL-Speed-128K (a), claude-3-5-haiku (b), claude-3-haiku-20240307 (c), flux (d), gemini-2.0-flash-thinking-exp (e), glm-4-flash (f), gpt4 (g), gpt4 (h), hunyuan-lite (i), moonshot-v1-8k-vision-preview (j) and pixtral-12b-2409 (k)—alongside our proposed compact model DCL-SE (l). The dashed diagonal line represents random-guess performance (AUC = 0.5).

specialized models like DCL-SE in the era of foundation models, we additionally conducted a systematic comparison between 11 representative large-scale pre-trained models and lightweight task-specific networks on the 2D Alzheimer's disease classification benchmark. Importantly, all large models were evaluated in a zero-shot manner, without task-specific fine-tuning, reflecting the realities of resource-limited clinical environments. As illustrated in Fig. 7 and detailed in Supplementary Tables 8–10, small models such as DCL-SE not only maintained consistently high accuracy and AUC (with Val_Acc and Test_Acc reaching up to 97.81% and 97.71%, and even 100% AUC in some cases), but also outperformed large models, whose predictions exhibited greater variability and, in some instances, approached random chance. This pronounced performance contrast on a relatively simple four-class classification task underscores the continuing necessity of compact, specialized architectures in foundational clinical applications, even as large models proliferate. These findings highlight the potential for complementarity between large-scale and small-scale models in practical medical AI systems, suggesting that lightweight models remain indispensable for efficient and robust deployment in real-world clinical scenarios.

F. Interpretability

A major challenge in deep learning-based neuroimaging is model interpretability, particularly for complex transformations such as 3D-to-2D compression and multi-stage feature decoding. In DCL-SE, interpretability arises directly from the explicit architectural design of DaSE, which consists of two tightly coupled stages: (1) ARP-based encoding, and (2) dynamic curriculum decoding via DGM. We conduct both quantitative and visual analyses to clarify how each stage preserves and structures diagnostically relevant information.

The first stage, ARP encoding, projects 3D brain MRI volumes into 2D dynamic images while retaining essential spatial and progression cues. As demonstrated in Fig.8a, the ARP-generated representations maintain both global and local structural features, effectively filtering out noise. Principal component analysis confirms that the major axes of anatomical

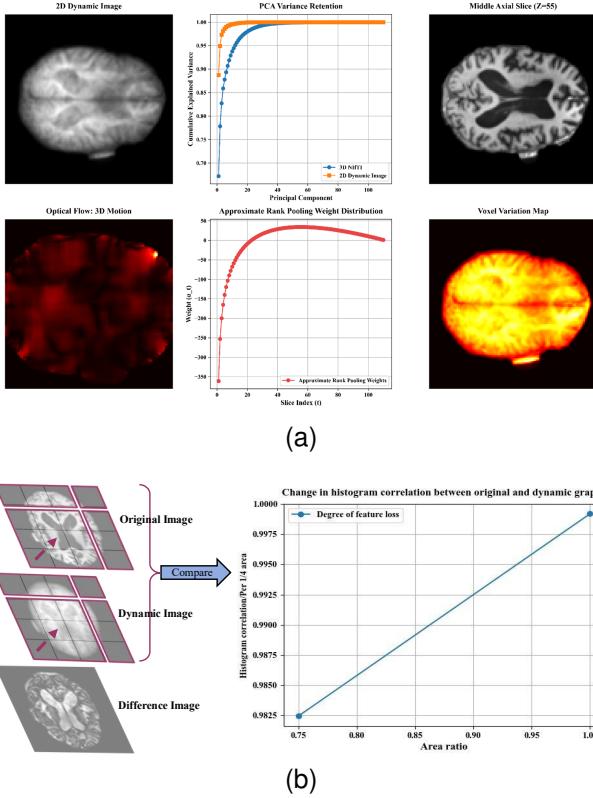


Fig. 8. Interpretability of feature loss in image format conversion. (a) Dynamic information analysis and visualization of 2D-3D conversion; (b) Histogram correlation-based feature loss degree for data format conversion.

variation are preserved in the ARP outputs, with minimal loss limited to higher-order components. Histogram correlation analysis (Fig. 8b) further quantifies this, indicating high fidelity between the original volumes and their dynamic representations across all modalities. For quantitative assessment, all data are normalized to the $[0, 1]$ range, ensuring consistent cross-modal comparison.

The second stage, dynamic curriculum decoding, enables progressive disentanglement of features and hierarchical learning—mirroring the clinical reasoning process from coarse anatomy to fine pathology. t-SNE visualization of feature embeddings at successive curriculum stages (Fig. 9) reveals how model representations evolve: early stages show mixed or overlapping clusters, while final stages form clearly separated, clinically meaningful categories. This progressive organization demonstrates that the curriculum-driven decoding (supported by DGM) systematically increases feature separability and transparency throughout the learning process.

In segmentation tasks, GradCAM-based visualizations (Fig. 10) provide additional interpretability: attention maps evolve from diffuse, uncertain focus at early decoding stages to precise anatomical localization at later stages. This step-wise sharpening illustrates how DCL-SE’s staged decoding mechanism achieves both fine discrimination and anatomical consistency without sacrificing context.

These analyses confirm that DCL-SE’s DaSE mechanism—through ARP encoding and curriculum-driven, dynamically recalibrated decoding—achieves interpretable, information-preserving transformation and progressive feature

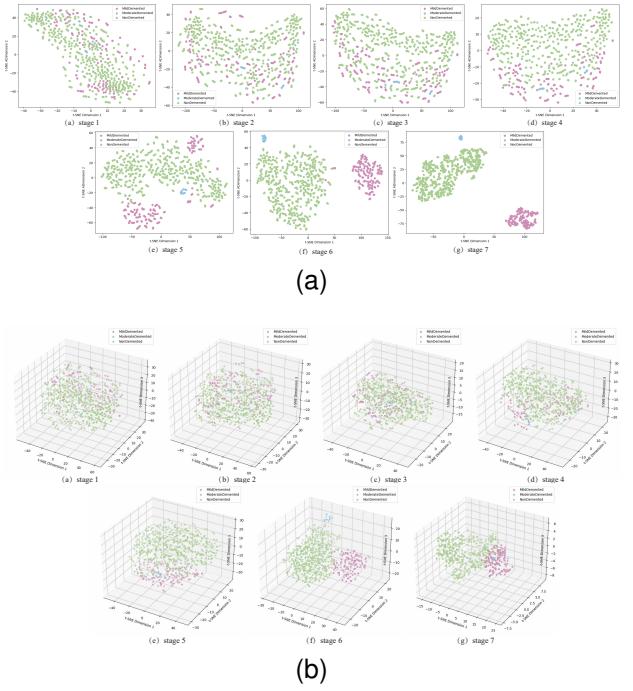


Fig. 9. Visualization of model interpretability across DaSE stages in the classification task: progressive learning from stage 1 to stage 7 transforms Alzheimer’s disease classification from dispersed clusters to three clearly separable categories. (a) t-SNE 2D visualization for 2D images; (b) t-SNE 3D visualization for 2D images.

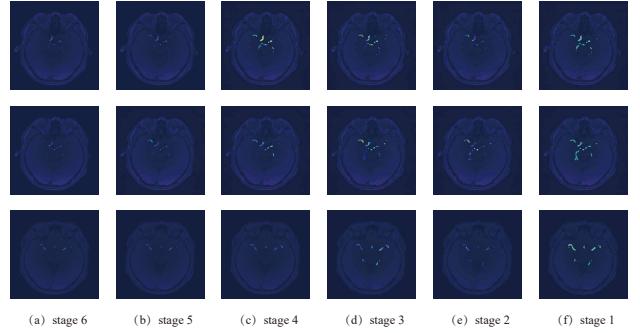


Fig. 10. Visualization of the interpretability of each stage of the model based on the DaSE mechanism in segmentation task. Through progressive learning using the DaSE mechanism, from stage 6 to stage 1, the model’s segmentation of cerebral arteries evolves from a blurred state to a clearly visible arterial division.

refinement. The method’s internal representations align with clinical reasoning, ensuring both robust performance and practical transparency for subject-level neuroimaging analysis.

V. CONCLUSION

This work presents DCL-SE, a unified framework for subject-level neuroimaging analysis that integrates Data-based Spatiotemporal Encoding (DaSE), dynamic curriculum learning (DCL), and the Dynamic Group Mechanism (DGM). The core paradigm—DaSE—organizes the process into two tightly coupled stages: ARP-based encoding, which preserves essential spatial and progression information from 3D volumetric data, and dynamic curriculum decoding, which enables progressive, adaptive feature refinement via DGM. This design aligns model learning with the hierarchical structure of

clinical reasoning, resulting in robust, interpretable extraction of diagnostically salient features.

Comprehensive experiments across multiple neuroimaging benchmarks—including classification, segmentation, and regression tasks—demonstrate that DCL-SE consistently achieves state-of-the-art accuracy, efficiency, and generalizability. Comparative analyses confirm that DCL-SE maintains high performance and interpretability in the face of data heterogeneity, and outperforms both traditional compact models and large-scale foundation models, particularly under real-world clinical constraints.

The staged feature disentanglement and adaptive focus provided by DCL and DGM are central to DCL-SE’s effectiveness, ensuring transparency and reliability in complex clinical scenarios. By bridging the gap between 3D volumetric data and efficient 2D neural architectures, DaSE enables practical cross-modal transfer and progressive training without sacrificing spatial or temporal fidelity.

While this study focuses on neuroimaging, the underlying principles of DCL-SE are broadly applicable. Future work will explore its extension to other imaging domains, integration with external clinical knowledge, and scalable deployment across diverse clinical environments. Collectively, these advances lay the foundation for next-generation, adaptive, and interpretable medical AI systems.

REFERENCES

- [1] C. Iadecola, “Neurovascular regulation in the normal brain and in Alzheimer’s disease,” *Nature Reviews Neuroscience*, vol. 5, no. 5, pp. 347–360, 2004.
- [2] E. Jun, S. Jeong, D.-W. Heo, and H.-I. Suk, “Medical Transformer: Universal Encoder for 3-D Brain MRI Analysis,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 35, no. 12, pp. 17779–17789, 2024, doi: 10.1109/TNNLS.2023.3308712.
- [3] A. Martinez-Usatorre and M. De Palma, “Alzheimer’s drug turns macrophages against cancer,” *Nature Cancer*, vol. 2, no. 11, pp. 1119–1121, 2021.
- [4] M. Hatami, F. Yaghmaee, and R. Ebrahimpour, “Investigating the potential of reinforcement learning and deep learning in improving Alzheimer’s disease classification,” *Neurocomputing*, vol. 597, Art. no. 128119, 2024.
- [5] C. Wang, Y. Li, Y. Tsuboshita, T. Sakurai, T. Goto, H. Yamaguchi, Y. Yamashita, A. Sekiguchi, H. Tachimori, and Alzheimer’s Disease Neuroimaging Initiative, “A high-generalizability machine learning framework for predicting the progression of Alzheimer’s disease using limited data,” *NPJ Digital Medicine*, vol. 5, no. 1, Art. no. 43, 2022.
- [6] M. Menagadevi, S. Devaraj, N. Madian, and D. Thiagarajan, “Machine and deep learning approaches for Alzheimer disease detection using magnetic resonance images: An updated review,” *Measurement*, Art. no. 114100, 2024.
- [7] S. Roy, S. Sadhu, and S. K. Bandyopadhyay, “A useful approach towards 3D representation of brain abnormality from its 2D MRI slides with a volumetric exclamation,” in *Proc. 2015 Third Int. Conf. Computer, Communication, Control and Information Technology (C3IT)*, 2015, pp. 1–6, doi: 10.1109/C3IT.2015.7060227.
- [8] T. A. Soomro, L. Zheng, A. J. Afifi, A. Ali, S. Soomro, M. Yin, and J. Gao, “Image Segmentation for MR Brain Tumor Detection Using Machine Learning: A Review,” *IEEE Reviews in Biomedical Engineering*, vol. 16, pp. 70–90, 2023, doi: 10.1109/RBME.2022.3185292.
- [9] H. Liu, Z. Ni, D. Nie, D. Shen, J. Wang, and Z. Tang, “Multimodal Brain Tumor Segmentation Boosted by Monomodal Normal Brain Images,” *IEEE Transactions on Image Processing*, vol. 33, pp. 1199–1210, 2024, doi: 10.1109/TIP.2024.3359815.
- [10] S. Zhang, B. Ren, Z. Yu, H. Yang, X. Han, X. Chen, Y. Zhou, D. Shen, and X.-Y. Zhang, “TW-Net: Transformer Weighted Network for Neonatal Brain MRI Segmentation,” *IEEE Journal of Biomedical and Health Informatics*, vol. 27, no. 2, pp. 1072–1083, 2023, doi: 10.1109/JBHI.2022.3225475.
- [11] B. Woo and M. Lee, “Comparison of tissue segmentation performance between 2D U-Net and 3D U-Net on brain MR Images,” in *Proc. 2021 Int. Conf. Electronics, Information, and Communication (ICEIC)*, 2021, pp. 1–4, doi: 10.1109/ICEIC51217.2021.9369797.
- [12] D. Konar, S. Bhattacharyya, T. K. Gandhi, B. K. Panigrahi, and R. Jiang, “3-D Quantum-Inspired Self-Supervised Tensor Network for Volumetric Segmentation of Medical Images,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 35, no. 8, pp. 10312–10325, 2024, doi: 10.1109/TNNLS.2023.3240238.
- [13] W. GTIFA, I. Khoja, and A. SAKLY, “Multimodal MRI Brain Tumor Segmentation Using 3D and 3D/2D Methods: A Study on the MICCAI BRATS Dataset,” in *Proc. 2024 IEEE Int. Conf. Advanced Systems and Emergent Technologies (IC_ASET)*, 2024, pp. 1–6, doi: 10.1109/IC_ASET61847.2024.10596251.
- [14] W. Gtifa, M. Fradi, A. Sakly, and M. Machhout, “Comparative Study of Enhanced Methods for Intelligent Segmentation of Brain Tissue in 3D/2D MRI Brain Images Through Multiple Thresholding,” in *Proc. 2023 IEEE Int. Conf. Artificial Intelligence & Green Energy (ICAIGE)*, 2023, pp. 1–5, doi: 10.1109/ICAIGE58321.2023.10346268.
- [15] J. Ma, Y. He, F. Li, L. Han, C. You, and B. Wang, “Segment anything in medical images,” *Nature Communications*, vol. 15, Art. no. 654, 2024, doi: 10.1038/s41467-024-44824-z.
- [16] R. Luo, L. Sun, Y. Xia, T. Qin, S. Zhang, H. Poon, and T.-Y. Liu, “BioGPT: Generative Pre-trained Transformer for Biomedical Text Generation and Mining,” *Briefings in Bioinformatics*, vol. 23, no. 6, Nov. 2022. [Online]. Available: <https://www.microsoft.com/en-us/research/publication/biogpt-generative-pre-trained-transformer-for-biomedical-text-generation-and-mining/>.
- [17] Z. Wang, Z. Wu, D. Agarwal, and J. Sun, “MedCLIP: Contrastive Learning from Unpaired Medical Images and Text,” in *Proc. Conf. Empir. Methods Nat. Lang. Process.*, 2022, pp. 3876–3887, doi: 10.18653/v1/2022.emnlp-main.256.
- [18] H. Wang, S. Guo, J. Ye, Z. Deng, J. Cheng, T. Li, J. Chen, Y. Su, Z. Huang, Y. Shen, B. Fu, S. Zhang, and J. He, “SAM-Med3D: A Vision Foundation Model for General-Purpose Segmentation on Volumetric Medical Images,” *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–14, 2025, doi: 10.1109/TNNLS.2025.3586694.
- [19] OpenAI, “GPT-4 Technical Report,” arXiv:2303.08774 [cs.CL], 2023.
- [20] Google DeepMind, “Gemini: A Family of Highly Capable Multimodal Models,” arXiv:2312.11805 [cs.CV], 2023.
- [21] Anthropic, “Citations,” Anthropic API documentation, [Online]. Available: <https://docs.anthropic.com/>, accessed July 2025.
- [22] Mistral AI, “Pixtral Large,” [Online]. Available: <https://mistral.ai/news/pixtral-large>, 2024.
- [23] Y. Sun, S. Wang, Y. Li, S. Feng, X. Chen, H. Zhang, X. Tian, D. Zhu, H. Tian, and H. Wu, “ERNIE: Enhanced Representation through Knowledge Integration,” arXiv:1904.09223 [cs.CL], 2019.
- [24] Black Forest Labs, “Flux (text-to-image model),” Wikipedia page, May 29, 2025. [Online]. Available: <https://en.wikipedia.org/wiki/Flux>.
- [25] Omniverse AI Models, “GLM-4-Flash: The ideal choice for handling simple tasks, with 128K context,” AI model description, 2025. [Online]. Available: <https://omniverse.com/im/discover/model/glm-4-flash?hl=en-US>.
- [26] LobeChat, “Hunyuan Lite: Preview version of next-generation Hunyuan large language model featuring MoE and 256K context,” AI model description, 2025. [Online]. Available: <https://lobechat.com/discover/model/hunyuan-lite?hl=en-US>.
- [27] Moonshot AI, “moonshot-v1-8k-vision-preview: 8K vision model capable of image understanding and text output,” Software documentation, 2025. [Online]. Available: <https://platform.moonshot.ai/>.
- [28] Z. Huang, W. Li, Y. Wu, L. Yang, Y. Dong, Y. Yang, H. Zheng, D. Liang, M. Wang, and Z. Hu, “Accurate Whole-Brain Image Enhancement for Low-Dose Integrated PET/MR Imaging Through Spatial Brain Transformation,” *IEEE Journal of Biomedical and Health Informatics*, vol. 28, no. 9, pp. 5280–5289, 2024, doi: 10.1109/JBHI.2024.3407116.
- [29] C. Lin, Y. Wang, T. Wang, and D. Ni, “Segmentation and Recovery of Pathological MR Brain Images Using Transformed Low-Rank and Structured Sparse Decomposition,” in *Proc. 2019 IEEE 16th Int. Symp. Biomedical Imaging (ISBI 2019)*, 2019, pp. 1878–1881, doi: 10.1109/ISBI.2019.8759441.
- [30] Z. Tang, S. Ahmad, P.-T. Yap, and D. Shen, “Multi-Atlas Segmentation of MR Tumor Brain Images Using Low-Rank Based Image Recovery,” *IEEE Transactions on Medical Imaging*, vol. 37, no. 10, pp. 2224–2235, 2018, doi: 10.1109/TMI.2018.2824243.
- [31] D. Nie, L. Wang, E. Adeli, C. Lao, W. Lin, and D. Shen, “3-D Fully Convolutional Networks for Multimodal Isointense Infant Brain Image

- Segmentation,” *IEEE Transactions on Cybernetics*, vol. 49, no. 3, pp. 1123–1136, 2019, doi: 10.1109/TCYB.2018.2797905.
- [32] K. M. Chew, C. Y. Yong, R. Sudirman, and S. T. C. Wei, “Human brain modeling tumor detection in 2D and 3D representation using microwave signal analysis,” in *Proc. 2018 IEEE Symposium on Computer Applications & Industrial Electronics (ISCAIE)*, 2018, pp. 310–316, doi: 10.1109/ISCAIE.2018.8405490.
- [33] Y. Wu, G. Wu, J. Lin, Y. Wang, and J. Yu, “Role Exchange-Based Self-Training Semi-Supervision Framework for Complex Medical Image Segmentation,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 36, no. 5, pp. 8372–8386, 2025, doi: 10.1109/TNNLS.2024.3432877.
- [34] Y. He, R. Ge, X. Qi, Y. Chen, J. Wu, J.-L. Coatrieux, G. Yang, and S. Li, “Learning Better Registration to Learn Better Few-Shot Medical Image Segmentation: Authenticity, Diversity, and Robustness,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 35, no. 2, pp. 2588–2601, 2024, doi: 10.1109/TNNLS.2022.3190452.
- [35] K. Vidhya, M. V. Patil, and R. S. Hegadi, “A Systematic Approach for Constructing 3D MRI Brain Image over 2D Images,” in *Recent Trends in Image Processing and Pattern Recognition: Second Int. Conf., RTIP2R 2018, Solapur, India, Dec. 21–22, 2018, Revised Selected Papers, Part II* 2, 2019, pp. 154–162.
- [36] M. Riyazudeen and M. Sathik, “Converting 2D magnetic resource imagining brain tumors to 3D structure using depth map machine learning techniques,” *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 27, no. 1, pp. 513–520, 2022.
- [37] Z. Peng, W. Zhang, N. Han, X. Fang, P. Kang, and L. Teng, “Active Transfer Learning,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, no. 4, pp. 1022–1036, 2020, doi: 10.1109/TCSVT.2019.2900467.
- [38] F. Zhuang, Z. Qi, K. Duan, D. Xi, Y. Zhu, H. Zhu, H. Xiong, and Q. He, “A comprehensive survey on transfer learning,” *Proceedings of the IEEE*, vol. 109, no. 1, pp. 43–76, 2020.
- [39] A. Hosna, E. Merry, J. Gyalmo, Z. Alom, Z. Aung, and M. A. Azim, “Transfer learning: a friendly introduction,” *Journal of Big Data*, vol. 9, no. 1, Art. no. 102, 2022.
- [40] P. Kora, C. P. Ooi, O. Faust, U. Raghavendra, A. Gudigar, W. Y. Chan, K. Meenakshi, K. Swaraja, P. Plawiak, and U. R. Acharya, “Transfer learning techniques for medical image analysis: A review,” *Biocybernetics and Biomedical Engineering*, vol. 42, no. 1, pp. 79–107, 2022.
- [41] X. Wang, R. Su, W. Xie, W. Wang, Y. Xu, R. Mann, J. Han, and T. Tan, “2.75 D: Boosting learning by representing 3D Medical imaging to 2D features for small data,” *Biomedical Signal Processing and Control*, vol. 84, Art. no. 104858, 2023.
- [42] Y. Wen, L. Chen, C. Zhou, Y. Deng, H. Zeng, S. Xi, and R. Guo, “On the effective transfer learning strategy for medical image analysis in deep learning,” in *Proc. 2020 IEEE Int. Conf. Bioinformatics and Biomedicine (BIBM)*, 2020, pp. 827–834.
- [43] X. Wei, Y. Qiu, X. Xu, J. Xu, J. Mei, and J. Zhang, “ECINFusion: A Novel Explicit Channel-Wise Interaction Network for Unified Multi-Modal Medical Image Fusion,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 35, no. 5, pp. 4011–4025, 2025, doi: 10.1109/TCSVT.2024.3516705.
- [44] M. Oquab, T. Dariseti, T. Moutakanni, H. V. Vo, M. Szafraniec, V. Khalidov, P. Fernandez, D. Haziza, F. Massa, A. El-Nouby, M. Assran, N. Ballas, W. Galuba, R. Howes, P.-Y. Huang, S.-W. Li, I. Misra, M. Rabbat, V. Sharma, G. Synnaeve, H. Xu, H. Jegou, J. Mairal, P. Labatut, A. Joulin, and P. Bojanowski, “DINOv2: Learning Robust Visual Features without Supervision,” *Transactions on Machine Learning Research*, 2024. [Online]. Available: <https://openreview.net/forum?id=a68SUt6zFt>.
- [45] Y. Bengio, J. Louradour, R. Collobert, and J. Weston, “Curriculum learning,” in *Proc. 26th Annual Int. Conf. Machine Learning*, 2009, pp. 41–48.
- [46] J. Wang, Y. Tang, Y. Xiao, J. T. Zhou, Z. Fang, and F. Yang, “GREnet: Gradually REcurrent Network With Curriculum Learning for 2-D Medical Image Segmentation,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 35, no. 7, pp. 10018–10032, 2024, doi: 10.1109/TNNLS.2023.3238381.
- [47] J. Zhang, J. Song, L. Gao, Y. Liu, and H. T. Shen, “Progressive Meta-Learning With Curriculum,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 32, no. 9, pp. 5916–5930, 2022, doi: 10.1109/TCSVT.2022.3164190.
- [48] T. Matiisen, A. Oliver, T. Cohen, and J. Schulman, “Teacher–student curriculum learning,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 31, no. 9, pp. 3732–3740, 2019.
- [49] M. Burduja and R. T. Ionescu, “Unsupervised Medical Image Alignment With Curriculum Learning,” in *Proc. 2021 IEEE Int. Conf. Image Processing (ICIP)*, 2021, pp. 3787–3791, doi: 10.1109/ICIP42928.2021.9506067.
- [50] H. Bilen, B. Fernando, E. Gavves, A. Vedaldi, and S. Gould, “Dynamic image networks for action recognition,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2016, pp. 3034–3042.
- [51] H. Bilen, B. Fernando, E. Gavves, and A. Vedaldi, “Action recognition with dynamic image networks,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 12, pp. 2799–2813, 2017.
- [52] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo, “Swin transformer: Hierarchical vision transformer using shifted windows,” in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2021, pp. 10012–10022.
- [53] X. Xing, G. Liang, H. Blanton, M. U. Rafique, C. Wang, A.-L. Lin, and N. Jacobs, “Dynamic Image for 3D MRI Image Alzheimer’s Disease Classification,” in *Computer Vision – ECCV 2020 Workshops*, Cham: Springer International Publishing, 2020, pp. 355–364.
- [54] M. Zhou, T. Zheng, Z. Wu, N. Wan, and M. Cheng, “DAMNet: Dynamic mobile architectures for Alzheimer’s disease,” *Computers in Biology and Medicine*, vol. 185, Art. no. 109517, 2025.
- [55] X. Huo, G. Sun, S. Tian, Y. Wang, L. Yu, J. Long, W. Zhang, and A. Li, “HiFuse: Hierarchical multi-scale feature fusion network for medical image classification,” *Biomedical Signal Processing and Control*, vol. 87, Art. no. 105534, 2024.
- [56] A. Todi, N. Narula, M. Sharma, and U. Gupta, “ConvNext: A Contemporary Architecture for Convolutional Neural Networks for Image Classification,” in *Proc. 2023 3rd Int. Conf. Innovative Sustainable Computational Technologies (CISCT)*, 2023, pp. 1–6, doi: 10.1109/CISCT57197.2023.10351320.



Meihua Zhou received her Bachelor of Science degree in Information Management and Information Systems from Wannan Medical university in 2023. From 2024 to 2025, she studied embodied intelligence and multimodal computing in the Department of Computer Science at Tsinghua University, while also acquiring medical knowledge at the Ophthalmology Research Institute of Beijing Tongren Hospital. She is committed to interdisciplinary approaches for solving real-world problems. Currently, she is a graduate student at the University of Chinese Academy of Sciences. Her research interests include, but are not limited to, intelligent medicine, embodied intelligence, and interdisciplinary applications.



Xinyu Tong is a graduate student at University of Chinese Academy of Sciences, studying Artificial Intelligence and Biomedical Information Processing. Prior to this he obtained a Bachelor’s Degree in Computer Science and his scientific dreams revolved around research in image processing and pattern recognition, using Multimodal-Big data-Efficient algorithms for intelligent medical image analysis and diagnostic interventions. Now his research interests include image processing, pattern recognition, intelligent Ecology and Drug information processing.



Jiarui Zhao received her Bachelor's degree in Energy and Electronic Devices from Chang'an University in 2023. She is currently pursuing her Master's degree at the University of the Chinese Academy of Sciences. Her research focuses on the fields of micro/nano energy and sensing technology, with particular interests in control systems and human-computer interaction. She is dedicated to exploring advanced technologies for energy harvesting, high-precision sensing, and the integration of intelligent control in next-generation interactive systems.



Nan Wan received her Bachelor of Engineering in Computer Science and Technology from Anhui Normal University in 2003 and her Master's degree in Educational Technology from Anhui Normal University in 2011. She is now an associate professor at Wannan Medical university, with a main research focus on machine learning and intelligent medicine.



Min Cheng received her Bachelor of Science degree in Information Management and Information Systems from Wannan Medical University in 2010. From 2010 to 2011, I served as an electronic medical record implementation engineer at Winning HEALTH Group Company. From 2011 to 2013, he worked as a smart healthcare project manager at Neusoft Group Company. Since 2013, he has been a senior engineer at the Information Center of Wuhu Hospital, Beijing Anding Hospital, Capital Medical University. His research interests include, but not limited to, medical software engineering methodology, intelligent medicine. At present, he is promoting the planning work for the informatization and intelligence construction of the National Medical Center, and am committed to promoting the deep integration of new-generation information technology and medical services.



Li Yang received his Bachelor of Engineering in Computer Science and Technology from Anhui University in 2006 and his Master of Engineering in Computer Technology from Anhui University in 2012. In 2019, as a young backbone teacher at Wannan Medical University, he participated in an exchange program at the University of Hong Kong. From 2023 to 2024, he is working as a visiting scholar at the University of Science and Technology of China, focusing on research related to artificial intelligence. Currently, Li Yang is an associate professor at Wannan Medical University. His research interests include but are not limited to artificial intelligence, medical image processing, and the psychology of college students.



Lei Tian is an Associate Professor, Deputy Chief Physician, and Master's Supervisor. He was selected as a Beijing Science and Technology Rising Star, a "Qingmiao" Talent by the Beijing Hospital Authority, a "Distinguished Talent" of Beijing Tongren Hospital, and a "Star of Tomorrow" in the China Optometry Talent Development Program. He serves as a committee member of the Ocular Surface and Dry Eye Section of the Chinese Medical Doctor Association, the National Orthokeratology Safety Monitoring Committee, and the Ophthalmology Branch of the Chinese Medical Equipment Association. He is also the Vice Chairman of the Youth Committee of the Ophthalmology Branch of the Beijing Medical Association. His research interests include ophthalmology, intelligent medicine and medical informatics.