

临床数据建模中的多域自适应问题研究进展

陈秀, 张馨匀, 程煜婷, 陈伟, 黄正行, 刘振宇, 张远鹏

引用本文

陈秀, 张馨匀, 程煜婷, 陈伟, 黄正行, 刘振宇, 张远鹏. 临床数据建模中的多域自适应问题研究进展[J]. 计算机科学, 2025, 52(9): 25-36.

CHEN Xiu, ZHANG Xinyun, CHENG Yuting, CHEN Wei, HUANG Zhengxing, LIU Zhenyu, ZHANG Yuanpeng. Research Progress on Multi-domain Adaptation Problems in Clinical Data Modeling[J]. Computer Science, 2025, 52(9): 25-36.

相似文献推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

智能医学工程专题序言

Perface of Special Issue of Intelligent Medical Engineering

计算机科学, 2025, 52(9): 1-3. <https://doi.org/10.11896/jsjcx.qy20250901>

基于渐进式自训练开集域适应的辐射源个体识别

Specific Emitter Identification Based on Progressive Self-training Open Set Domain Adaptation

计算机科学, 2025, 52(7): 279-286. <https://doi.org/10.11896/jsjcx.240600073>

基于相关熵的多视角低秩矩阵分解和多视角数据聚类中的约束图学习

Correntropy Based Multi-view Low-rank Matrix Factorization and Constraint Graph Learning for Multi-view Data Clustering

计算机科学, 2025, 52(6A): 240900131-10. <https://doi.org/10.11896/jsjcx.240900131>

基于多尺度注意力和不确定性损失的两阶段左心房疤痕分割

Two-stage Left Atrial Scar Segmentation Based on Multi-scale Attention and Uncertainty Loss

计算机科学, 2025, 52(6): 264-273. <https://doi.org/10.11896/jsjcx.241200197>

基于多原型重放和对齐的类增量无源域适应

Class-incremental Source-free Domain Adaptation Based on Multi-prototype Replay and Alignment

计算机科学, 2025, 52(3): 206-213. <https://doi.org/10.11896/jsjcx.240100166>

临床数据建模中的多域自适应问题研究进展

陈秀¹ 张馨匀¹ 程煜婷¹ 陈伟¹ 黄正行³ 刘振宇⁴ 张远鹏^{1,2}

1 南通大学医学院医学信息学系 江苏 南通 226001

2 香港理工大学医疗科技与资讯学系 香港 999077

3 浙江大学计算机科学与技术学院 杭州 310058

4 中国科学院自动化研究所 北京 100190

(2431310054@stmail.ntu.edu.cn)

摘要 随着人工智能与医疗健康的深度融合,临床数据正经历从“辅助决策”到“驱动决策”的范式转变。临床数据包括患者症状、诊断影像、治疗记录等结构化与非结构化信息,为医疗决策提供重要支撑。然而,由于“领域偏移”现象的普遍存在,临床AI模型训练评估依赖的数据独立同分布假设(i. i. d.)失效,模型的跨域泛化能力被严重制约。域适应和域泛化技术可有效提升模型跨域表现。前者利用无标注目标域数据调整模型,使其适配新环境;后者基于源域数据学习域不变特征,实现无目标域数据下的泛化。针对两类技术在临床数据建模中的应用进展,按浅层、深层方法分类,展示其在不同数据类型中的应用场景,并总结了当前各类方法在泛化性能、数据依赖性与可解释性等方面的表现差异。

关键词: 临床数据; 领域偏移; 域适应; 域泛化; 数据依赖

中图分类号 R319

Research Progress on Multi-domain Adaptation Problems in Clinical Data Modeling

CHEN Xiu¹, ZHANG Xinyun¹, CHENG Yuting¹, CHEN Wei¹, HUANG Zhengxing³, LIU Zhenyu⁴ and ZHANG Yuanpeng^{1,2}

1 Department of Medical Informatics, School of Medicine, Nantong University, Nantong, Jiangsu 226001, China

2 Department of Health Technology and Informatics, The Hong Kong Polytechnic University, Hong Kong 999077, China

3 College of Computer Science and Technology, Zhejiang University, Hangzhou 310058, China

4 Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China

Abstract With the deep integration of artificial intelligence and healthcare, clinical data is undergoing a paradigm shift from “aiding decision-making” to “driving decision-making”. Clinical data encompasses both structured and unstructured information such as patient symptoms, diagnostic images, and treatment records, providing crucial support for medical decision-making. However, due to the prevalent “domain shift” phenomenon, the independent and identically distributed (i. i. d.) assumption, which clinical AI models rely on for training and evaluation, is invalidated, severely restricting the models’ cross-domain generalization ability. Domain adaptation and domain generalization techniques can effectively enhance the cross-domain performance of models. The former adjusts models use unlabeled target domain data to adapt them to new environments, the latter learns domain-invariant features based on source domain data to achieve generalization without target domain data. Regarding the application progress of these two types of techniques in clinical data modeling, this paper classifies them into shallow and deep methods, demonstrates their application scenarios across different data types, and summarizes the current performance differences of various methods in terms of generalization performance, data dependency, and interpretability.

Keywords Clinical data, Domain shift, Domain adaptation, Domain generalization, Data dependence

随着人工智能与医疗健康的深度融合,临床数据正经历从“辅助决策”到“驱动决策”的范式转变^[1]。在“辅助决策”阶段,临床数据作为医生决策的辅助信息,支持医生基于经验、指南和部分数据进行判断;在“驱动决策”阶段,临床数据则成为决策的核心驱动力,通过多模态建模、深度学习等技术,

实现对患者患病情况的精准预测以及生成个性化的治疗方案。临床数据不仅涵盖医疗过程中产生的结构化或非结构化信息,如患者症状、体征、诊断、治疗及预后,还包括诊断数据(如影像、病理)、干预数据(如手术记录、用药记录)以及结局数据(如生存期、并发症)。近年来,基于深度学习的多模态建

到稿日期:2025-06-14 返修日期:2025-08-01

基金项目:江苏高校“青蓝工程”

This work was supported by the “Qinglan Project” of Jiangsu Higher Education Institutions.

通信作者:张远鹏(y. p. zhang@ieee. org)

模方法取得显著进展,但无论是传统的机器学习还是深度学习,模型的训练和评估都建立在数据独立同分布(i. i. d.)的假设基础之上。这一假设在现实世界的临床应用中常常被“领域偏移”现象打破,导致模型泛化性能下降^[2-3]。

领域偏移的根源主要包括:模型过度拟合特定中心的噪声与伪特征;标注主观性引起的标签偏移;隐私政策限制下数据孤岛阻碍特征对齐。这些问题共同制约了临床 AI 模型跨机构、跨设备和跨人群的泛化能力。

域适应(Domain Adaptation, DA)和域泛化(Domain Generalization, DG)技术可以在有限目标数据甚至无目标数据参与的情况下提升模型的跨域表现,是构建鲁棒临床决策系统的关键路径。域适应允许访问无标注目标域数据以对模型进行调整,而域泛化则要求模型在无法访问目标域数据的前提下,仅基于源域数据学习域不变特征以实现跨域泛化。两者虽场景设置不同,但目标均为缓解领域间的数据分布偏移问题。

本文旨在回顾近年来域适应和域泛化技术在临床数据建模中的研究进展,将它们按照浅层和深层方法划分后进一步分类,以探讨各类策略在不同数据类型中的应用实例,并对比不同方法在应对数据异质性、隐私隔离和跨机构推广能力上的优势与不足。

1 相关概念界定

DA 和 DG 是当前应对数据分布偏移问题的两大主流技术,两者的核心目标均为提升模型在目标域上的泛化能力,区别在于 DG 不能接触目标域数据。在深入探讨具体方法前,需要明确“域”的基本定义、两种策略的适用场景以及理论基础。

1.1 域的定义

在机器学习中,“域”(Domain)通常被定义为特定的数据分布环境,其形式化表示为 $D = \{X, Y\}$ 。其中: X 是输入空间(特征空间),表示数据的所有可能输入特征; Y 为输出空间(标签空间),表示数据的所有可能标签。域之间的差异可能来自于数据采集设备、地理位置、人群组成或者标签标准等。对于临床数据而言,域的划分可能包括不同医院、不同模态、

不同的数据类型等。

设源域为 $\mathcal{D}_S = \{(x_i^S, y_i^S)\}_{i=1}^{n_S}$, 目标域为 $\mathcal{D}_T = \{(x_j^T, y_j^T)\}_{j=1}^{n_T}$, 两者在边缘概率分布 $P(X)$ 或条件概率分布 $P(Y|X)$ 上存在差异,即构成“领域偏移”问题。

1.2 域适应

域适应的核心思想是通过调整特征空间或学习映射函数,使得模型能够将从源域中学到的知识有效地迁移到目标域,从而提高目标域的表现。在域适应任务中,假设存在源域 \mathcal{D}_S 和目标域 \mathcal{D}_T , X_i 和 X_j 分别为源域和目标域的输入特征, Y_i 和 Y_j 为对应的标签。域适应的目标是通过学习一个映射,使得在源域上训练的模型能够在目标域上获得较好的泛化能力:

$$\min_{\theta} \mathcal{L}_S(f_{\theta}(x^S), y^S) + \lambda \mathcal{L}_T(f_{\theta}(x^T)) \quad (1)$$

其中, \mathcal{L}_S 和 \mathcal{L}_T 分别为源域和目标域的损失函数, λ 是权衡源域和目标域损失之间的超参数。在实际的临床应用中,域适应策略特别适用于存在部分目标域标注数据或者可以获取无标注目标数据(如联邦学习中的客户端数据)的情境。

1.3 域泛化

域泛化强调模型在训练阶段完全不接触目标域数据,而是依赖多个源域中的通用特征,学习具有迁移能力的特征。在此设定下,模型需要在面对未见分布时仍然保持稳健性,其目标可形式化为:

$$\min_{\theta} \sum_{k=1}^K \mathcal{L}_{D_k}(f_{\theta}(x^{(k)}), y^{(k)}) \quad (2)$$

其中, k 表示第 k 个源域,目标是在多源学习中构建泛化性强的表示空间,具备“领域不可知”能力。这一策略特别适用于目标域无法提前获取,但模型需要具备跨域能力的场景,如多中心的影像分割系统。

1.4 两者比较

表 1 列出了域适应和域泛化技术的差异,域适应强调“迁移”,而域泛化更侧重于“泛化能力的内生建构”,但它们的共同目标都是解决数据的异质性问题,提升模型的泛化能力。在实际的临床应用中,两者并非对立,而是根据数据可用性、部署需求等进行策略选择的互补方案。

表 1 DA 与 DG 的差异

Table 1 Differences between DA and DG

对比项	DA	DG
是否使用目标域数据	是	否
目标	适配特定目标域	泛化未知目标域
训练方法	迁移调整	多源融合+域不变建模
适用情境	有目标域数据可用	目标域数据未知
优点	适应更精准	应用场景更广泛
缺点	泛化能力相对较弱	模型训练难度高
临床应用实例	多中心/多设备抑郁症自动识别 ^[4] 多中心/多设备脑肿瘤自动分割的 ^[5] 多中心/多设备低剂量图像重建 ^[6]	多中心医学影像分割 ^[7] 隐私保护下的眼底疾病诊断辅助 ^[8] 隐私保护下的多中心胸部 X 光影像多标签疾病分类 ^[9]

2 域适应

2.1 浅层域适应

浅层域适应主要采用两种学习策略:实例加权和特征匹配。图 1 展示了浅层域适应方法的分类及其对应流程。在临床应用场景中,实例加权是目前浅层方法中更为常见的实现形式,其核心思想是通过评估源域实例与目标域的

相关性,对源域实例进行重新加权,使加权调整后的源域数据分布更贴近目标域分布。基于特征匹配的域适应方法则通过构建新的特征表示空间,显式地减小源域和目标域特征在该空间中的分布差异。这类方法可进一步细分为子空间学习与分布对齐两类。由表 2 可以看出,前者主要关注于设计合适的特征变换策略,将原始特征映射至新的子空间,以缩小源域与目标域之间的表征差异;而后者

则侧重于通过显式度量和最小化不同域之间的分布差异来实现域对齐。

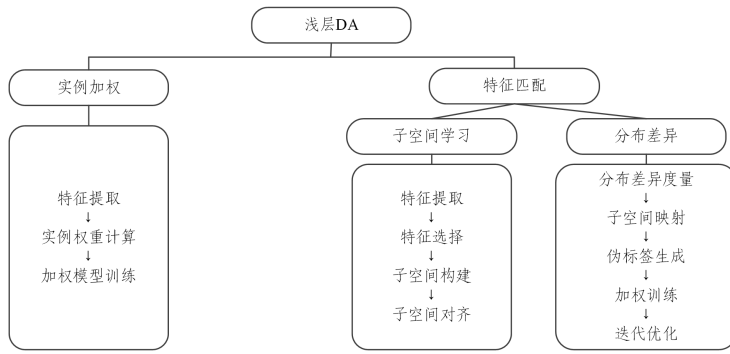


图1 浅层 DA 的主要策略和流程

Fig.1 Key strategies and processes of shallow DA

表2 浅层 DA 的核心策略比较

Table 2 Comparison of core strategies in shallow DA

作者	核心策略	核心问题	优势	局限	适用场景
Wachinger 等	实例加权+多特征融合	多中心 MRI 数据分布差异	将“形状特征”引入阿尔茨海默病诊断	依赖少量标记数据;实现难度高;可能丢失关键差异	阿尔茨海默病诊断
Cheplygina 等	实例加权+纹理特征提取	不同扫描仪或协议的 CT 数据分布差异	高斯尺度空间特征对 CT 设备或协议差异更鲁棒	距离加权计算耗时;极端分布差异的处理能力有限	患者级标签易获取; ROI 级标注成本高
Li 等	改进的子空间对齐	数据不足+分布差异	实现难度低;小样本适应性强	依赖目标域标记;未扩展到其他模态	阿尔茨海默病诊断
Jiang 等	RKHS 子空间学习+SLDA 正则化	EEG 的非平稳性	非平稳性适应好;可迁移性强	参数复杂度高;计算开销大	运动想象分类
Peng 等	特征适应+标签传播+图学习	EEG 的非平稳性和个体差异	可解释性强;联合优化;解决传统固定图结构导致的标签传播偏差	参数复杂度高;计算开销大;依赖标签质量	基于脑电图的跨被试情绪识别
Cui 等	动态分布对齐+双子空间映射	EEG 的个体差异和时段差异	双子空间保留域特性;高效处理实时场景	可能丢失 EEG 的时频细节;源域敏感性高	基于 EEG 的跨被试驾驶员精神状态检测

2.1.1 实例加权

实例加权方法通过对源域中与目标域分布更相似的样本赋予更高权重,而对与目标域差异较大的源域样本降低权重,来提升模型在目标域上的泛化能力。例如,Wachinger 等^[10]提出一种基于实例加权的监督域适应方法,首次将形状特征(基于 BrainPrint 的拉普拉斯谱特征)引入阿尔茨海默病诊断,使用磁共振成像(Magnetic Resonance Imaging, MRI)数据,融合了体积、厚度、形状等多模态特征,实现了阿尔茨海默病的三分类诊断。该方法通过核密度估计构建在临床变量上的联合概率分布,进而计算每个源域样本的概率权重,并集成至弹性网多类逻辑回归模型中,以此实现跨域对齐。结果显示,即使在目标数据量极少的情况下,实例加权仍然优于基线。而针对多中心慢性阻塞性肺病分类,Cheplygina 等^[11]使用的 GSS(Gaussian Scale Space)纹理特征对 CT(Computed Tomography)设备差异更具鲁棒性。该方法只需患者级标签,通过引入基于分类器、距离和对数的 3 种加权策略,调整训练样本的权重,以此来增强模型对目标数据集的适应性。

以上两者均强调“鲁棒特征提取”的重要性,认为特征质量是跨域性能的基础,需要针对数据模态设计专属特征。此外,在有监督场景中可采用实例加权,弱监督或半监督场景中需结合数据结构挖掘^[12]。

2.1.2 特征匹配

子空间学习主要关注设计出合适的特征变换策略,将原始特征映射至新的子空间中,以缩小源域与目标域之间的

表征差异。例如,Li 等^[13]通过 SVD(Singular Value Decomposition)构建源域与目标域的子空间,学习映射矩阵 M 使目标域数据对齐到源域子空间,以此减少域间整体分布差异。用 SVD 替代 PCA(Principal Component Analysis),避免了小样本下协方差矩阵计算不稳定的问题,且计算简单,不需要复杂的参数调优,适用于罕见病等小样本的情况。Jiang 等^[14]将数据映射到 RKHS(Reproducing Kernel Hilbert Space)空间,并引入 SLDA(Sparse Linear Discriminant Analysis)正则化项达到增大源域类间距离,缩短类内距离的目的,弥补了最大均值差异(Maximum Mean Discrepancy, MMD)只关注分布均值的缺陷,不仅增强了源域判别性,还减少了域间差异,对脑电图(Electroencephalogram, EEG)等高噪声、低信噪比的数据更具鲁棒性。在人脸、文本等多个数据集上进行了实验,验证了该方法具有可迁移性。以上两种方法均基于子空间学习,将源域与目标域数据映射到共享子空间,减少域间分布差异,而非直接训练一个通用模型。

基于分布差异的方法侧重于通过显式度量和最小化不同域之间的分布差异来实现域对齐。其中,最常见的策略是使用 MMD 来衡量源域和目标域的边缘概率分布差异和(或)条件概率分布差异。例如,Peng 等^[15]提出的基于联合分布匹配和图自适应标签传播的域适应模型,通过最小化源域和目标域的概率分布差异,动态更新图相似性矩阵,从而基于 EEG 信号准确估计目标样本的情绪状态。图相似度矩阵随着数据的对齐过程动态更新,解决了传统固定图结构导致的标签传播偏

差。Cui 等^[16]则结合子空间学习策略和分布匹配策略,提出一种带有双子空间映射的动态分布对齐方法。该方法将目标和源域分别映射到不同的子空间,避免了单空间映射导致的信息丢失,而动态权重在分布差异大的场景中,可自适应地增强关键分布的对齐强度。两者均从脑电图信号的非平稳性和高噪声特性出发,包含特征映射与分布对齐两个步骤,通过特征映射将数据投影到共享空间,再通过分布对齐缩小域间差异。

2.2 深度域适应

相较于浅层域适应方法,深度域适应在临床应用中更为广泛。如图 2 所示,深度域适应方法主要从对抗学习、分布差异和重构 3 个方面展开。基于对抗学习的方法主要通过对抗训练学习域不变的特征表示。基于领域分布差异的方法侧重于缩小源域和目标域在特征分布上的差异,从而提升模型在目标域上的泛化能力。此类方法一般通过深度神经网络学习一个跨域共享的特征表示,使得源域训练的模型能够适应目标域的任务需求。基于重构的方法则通过重构源域或目标域数据,引导模型学习域不变的特征表示,这类方法的主要优势在于能够在保持原始数据信息的同时,增强特征空间的可迁移性。

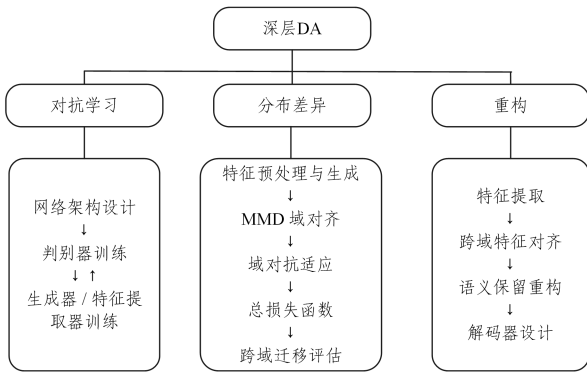


图 2 深层 DA 的主要策略和流程

Fig. 2 Key strategies and processes of deep DA

2.2.1 对抗学习

领域对抗神经网络 (Domain-Adversarial Neural Network, DANN)^[17]是一种广泛应用于医学领域的深度域适应框架,它包含 3 个主要结构:特征提取器、标签预测器和域判别器。其中,特征提取器作为前馈神经网络,用于提取样本数据的特征,旨在学习一种域不变的特征表示;标签预测器根据特征提取器提取的特征,对样本进行分类预测;域判别器用于区分提取的特征所属的领域。特征提取器和域判别器之间通过梯度反转层连接,以实现对抗性训练。

在 DANN 框架的基础上,许多研究者提出了不同的应用和改进方案。Jie 等^[18]提出了一种基于 DANN 的早期胃癌诊断模型,利用源域特征提取器在公共数据集上训练模型,并为每个目标域特征提取器添加自适应源——目标域映射分类器,可动态调整目标域类别数量,解决了源域和目标域类别数量不同的域适应问题。DANN 的对抗学习思路也被拓展到基因组相关数据的域适应任务中。例如,Ren 等^[19]针对单细胞测序数据的域偏移问题,提出了多对抗结合数据增强的域适应网络,通过引入多个域判别器来分别对齐不同层级的特征分布,实现了单细胞类型的跨域分配。该模型构建了含

170 万细胞的参考图谱,满足多种场景注释。多对抗器消除了不同测序平台的批次差异,SMOTE 过采样稀有细胞,解决了传统方法会忽视小样本的问题。与浅层域适应的分布对齐一样,DANN 结合了特征映射与分布对齐两大步骤,不同的是,这里的两种方式通过对抗学习来缩小域间差异,且依赖于深度神经网络提取特征,利用端到端训练优化模型。

另一种对抗学习策略是基于生成对抗网络 (Generative Adversarial Network, GAN) 来解决领域偏移的问题。Somers 等^[20]提出了一种基于 GAN 的改进方法,先合成膀胱环境训练深度估计网络,然后通过带门控残差块的对抗域适应,稳定训练并减少域偏移。门控机制能使对抗训练更加稳定,避免模式崩溃。Qin 等^[21]提出了一种基于 CycleGAN 的语义保持对抗域适应网络,通过对抗重构翻译图像,并利用特征一致性约束和条件标签损失确保翻译后的图像类别不变,从根本上解决了翻译后图像类别失真的问题。以上两种方式均以对抗学习为核心,利用生成器与判别器来对齐源域和目标域特征,其本质是学习域不变表示。

2.2.2 分布差异

无论是浅层还是深层方法,基于 MMD 的特征对齐方法是一类经典且广泛应用的代表。例如,Bao 等^[22]提出的两级域适应神经网络,第一级使用 MMD 来减少源域和目标域之间深度特征的分布差异,第二级使用 DANN 来强制深度特征更接近其相应的类中心,以此兼顾全局与局部分布对齐,而拓扑图特征保留了 EEG 空间拓扑信息,相比于传统特征更能捕捉差异。Liu 等^[23]采用注意力引导的细粒度对齐结合残差校准,解决了部分域适应的空间与语义负迁移。其中,区域注意力判别块生成细粒度注意力值,聚焦病变区域,减少无关信息的干扰,残差特征校准块用类加权 MMD 降低无关源样本权重。

与显式度量分布差异的方法不同,近年来也有临床研究通过对抗机制隐式对齐源域与目标域的特征分布。例如,Hong 等^[24]提出的基于对抗学习的动态联合域适应网络,通过引入全局和局部判别器分别处理边缘和条件概率分布差异,并通过动态对抗因子自适应地调整两者的权重,以适应不同的跨域差异。

2.2.3 重构

在临床应用中,自编码器和变分自编码器是重构方法中常用的两种网络架构。自编码器专注于学习数据的有效表示,而变分自编码器在自编码器的基础上引入了概率建模和随机元素,使其在提取特征的同时具备生成新数据的能力。例如,在 Ofer 等^[25]提出的监督自编码器框架中,编码器先将 EEG 信号压缩到潜在空间,解码器对信号进行重构,通过重构损失去除噪声,并且引入任务损失和会话损失,无需生成新样本,可直接优化特征表示。Cui 等^[26]提出了一种基于端到端变分自编码器的特征对齐方法,通过变分自编码器提取源域和目标域的特征,并结合 Kullback-Leibler 散度和图像重建损失对齐两域的全局与局部特征分布,同时结合对抗学习与原型对比损失进一步提升特征适应性。最终,该方法结合 CycleGAN 实现图像和特征空间的双重对齐,从而提升跨模态心脏 MRI 分割的准确性。

另一类方法将图像表示显式地划分为内容与风格两部

分,通过表示解耦或循环重构确保结构信息保持一致。例如, Jiang 等^[27]提出通过解耦表示学习将医学图像分解为内容编码(解剖结构)和样式编码(模态特征)。其核心在于设计了双重重构机制:内容重构损失强制跨模态翻译时保持解剖结构一致性,样式重构损失通过特征分布匹配实现模态对齐,具有

很强的语义保留和双向适应能力。更加详细的技术特性可以参考表 3。

以上方法均基于深度学习的表示学习,通过特征映射将数据投影到共享空间,结合损失函数实现分布对齐,同时强调保留任务相关语义信息,避免因对齐分布而丢失关键信息。

表 3 深层 DA 的核心策略比较

Table 3 Comparison of core strategies in deep DA

作者	核心策略	核心问题	优势	局限	适用场景
Jie 等	对抗性域适应神经网络	内窥镜图像的分类差异	灵活处理类别差异;迁移能力强;端到端设计高效	依赖源域数据质量;对抗训练不稳定	早期胃癌(EGC)预测
Ren 等	多对抗+数据增强	大规模单细胞数据的批次效应与类别不平衡	支持多场景注释;批次效应处理彻底;稀有类型适应性强	计算成本高;依赖生物学先验	单细胞 RNA-seq 数据的细胞类型注释
Somers 等	合成-真实域迁移	对抗训练不稳定	对抗训练稳定;解决数据稀缺的痛点	依赖模型重建精度;对手术工具的深度估计不准确	膀胱镜图像的单目深度估计
Qin 等	语义保持对抗网络	疾病相关语义信息丢失	解决翻译后图像类别失真问题;无需成对数据	模型复杂度;对罕见疾病不适用	胸部 X 光片疾病识别
Bao 等	两级域适应网络	EEG 非平稳性	兼顾全局分布对齐与类中心聚集;拓扑图特征更能捕捉差异	超参数敏感;自建数据集样本量较小	基于 EEG 的情感识别
Liu 等	注意力引导的细粒度对齐+残差校准	部分域适应存在空间负迁移和语义负迁移	减少无关信息干扰;有效过滤源域无关类别	依赖源域与目标域标签空间的包含关系;计算成本高	基于胸部 X 光片的肺炎自动诊断
Hong 等	动态联合域适应	跨会话 EEG 分布差异	无需手动调整权重,适应不同跨会话差异;端到端架构兼容多种 EEG 任务	可能引入负迁移;对 EEG 预处理较敏感	基于 EEG 的运动想象分类
Ofer 等	有监督自编码器	跨会话 EEG 非平稳性	无需新会话数据;可解释性强;兼容性好	依赖源域质量;仅验证运动想象任务;训练时间长	基于 EEG 的跨会话运动想象脑机接口;源域会话数量>20
Cui 等	变分自编码器(显式对齐+隐式对齐+CycleGAN+知识蒸馏)	源域与目标域分布差异大;分割边界模糊	多尺度对齐有效;同时解决监督语义误差和推理语义误差	模型复杂;对稀有类别敏感;依赖数据规模	心脏 MRI 跨模态分割;目标域样本量>400
Jiang 等	解耦表示+跨模态翻译	跨模态图像风格差异;丢失语义信息	语义保留;双向适应;适合个性化场景	计算量大;对形状不规则器官不友好	腹部器官跨模态分割;源域样本>2000

3 域泛化

3.1 浅层域泛化

如图 3 所示,本文将浅层的域泛化方法分为基于数据操作、基于表示学习和基于学习策略 3 类^[28]。但由于临床数据之间关系复杂,而浅层模型的表达力十分有限,浅层域泛化方法在临床数据建模中一般结合其他方法共同作用,较少作为主要策略单独使用。

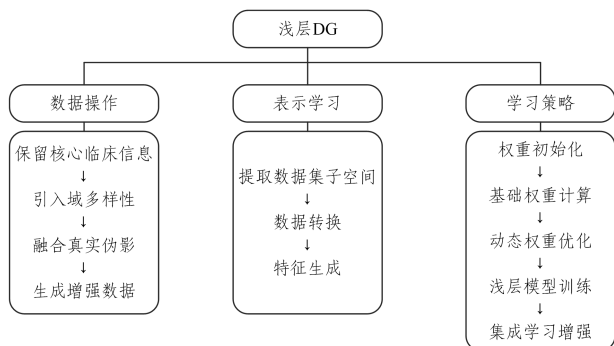


图 3 浅层 DG 的主要策略和流程

Fig. 3 Key strategies and processes of shallow DG

3.1.1 数据操作

在浅层域泛化方法中,数据操作是一种常见而有效的

策略。在临床建模中,成像设备、扫描协议、数据分布乃至语义标注标准的差异均可能导致显著的领域偏移。Guo 等^[29]系统地评估了扫描仪间的差异性对模型性能的影响,表明在训练数据中引入来自目标设备的图像或施加相应风格扰动,能显著提升模型在异构设备下的适应表现。

传统的数据增强可能破坏医学图像的解剖学上下文,导致模型依赖无关特征,且难以应对类别不平衡问题。基于此, Athalye 等^[30]针对胎儿超声图像的视图分类任务,提出域信息引导的剪切-粘贴数据增强策略,通过胸腔分割模型提取真实图像的关键区域(胸腔),混合关键区域生成新的训练数据。该方法的域信息引导策略使模型更加关注关键区域,且策略基于真实的解剖结构,生成的混合图像减少了人工造影。表 4 总结了各方法的局限和适用场景,该方法的局限是需要依靠真实图像,仍然存在隐私风险。相反, Kaisti 等^[31]构建的基于域随机化的合成心电图增强方案无需真实数据参与,无隐私风险。该方法通过缩放系数 C 控制波形、RR 间期(R-R Interval)、噪声的随机化程度,生成多样化的合成信号。

3.1.2 表示学习

在浅层域泛化的表示学习方法中,特征变换方法的核心是通过数学变换,将不同域的原始特征映射到一个共享的特征空间,在此空间中,不同域的特征分布更接近,任务相关的

本质特征被保留。以 Ayodele 等^[32]的研究为例,其在异构头皮脑电图数据上的癫痫检测任务中引入迁移成分分析(Transfer Component Analysis, TCA),将不同 EEG 数据集的特征映射到共享子空间,再通过坐标变换、插值等操作生成统一格式的光栅特征,然后让循环卷积神经网络在这个统一特征空间中学习通用模式,实现对目标域的泛化。该方法整合了不同数据集,通过增加训练数据的数量和多样性来减少偏差,有助于模型学习更具泛化性的特征,同时也可能面临协变量偏移问题,即使采用域泛化方法也难以完全消除。

表 4 浅层 DG 的核心策略比较

Table 4 Comparison of core strategies in shallow DG

作者	核心策略	核心问题	优势	局限	适用场景
Athalye 等	上下文保留的剪切-粘贴策略 (离线生成+在线生成)	传统数据增强破坏医学图像的解剖学上下文,且难以应对类别不平衡	保留医学真实性;灵活平衡数据分布;减少假阳性;可与传统增强结合使用	依赖胸腔分割模型的准确性;实现复杂度高;整体性能与传统增强接近;仍依赖真实图像	胎儿超声图像的视图分类
Kaisti 等	合成 ECG+域随机化	真实 ECG 标注成本高、隐私限制严格、泛化性差	合成数据无需采集或标注;数据分布、类别比例可精确控制;符合隐私要求	可能忽略罕见病理特征;扩展到复杂疾病分类难度较高	ECG 的 R 波检测
Ayodele 等	TCA 基于再生核希尔伯特空间的子空间提取、坐标插值	自动癫痫发作检测器在实际临床数据测试时泛化能力差	增加训练数据的数量和多样性;减少数据集偏差	面临协变量转移问题	癫痫发作检测;训练数据不足
He 等	随机样本加权	协变量偏移	无需强假设;适应性强;性能稳定;随机权重自动生成,保证足够异质性	有限样本下的随机性风险	多中心或跨机构的疾病风险预测

3.2 深层域泛化

3.2.1 数据操作

深层域泛化方法中的数据增强方法可分为混合增强与基于模型的增强两大路径,分别从数据与模型两个维度提升模型对未知域的适应能力。

混合增强是指将传统的数据扰动操作与更复杂的结构建模方法相结合,构建多层次、多模态的扰动机制。与浅层增强不同,深层模型中的旋转、裁剪、缩放等传统操作常作为子模块集成到整体框架中,配合风格转换、伪样本生成等机制共同完成训练分布的拓展。例如, Liu 等^[34]为解决视网膜血管分割中的域偏移问题,将结构增强和风格增强相结合,通过空间殖民算法生成类似视网膜血管的分支结构,结合改进的 Pix2Pix 模型(多尺度判别器)生成伪视网膜图像,扩展血管形态多样性,并使用 Pix2Mix 进行随机光度增强,引入不确定性扰动来丰富图像风格风格。Soleimani 等^[35]提出测试时数据增强策略,仅在测试时对输入 ECG 信号进行随机缺失、掩码、加噪声、移位等多种增强,生成多个增强样本,十分适合临床实时场景,性能也优于元学习和传统深度学习方法。但测试时需生成多个增强样本并融合预测,导致推理速度降低,不适合实时性要求极高的场景,且该策略只针对 ECG 信号设计,增强方法难以直接迁移到其他医学信号。Ye^[36]为解决医学图像分割中标注数据少和域偏移两大问题,首次融合了半监督学习和域泛化,提出结合全局增强和局部区域增强,并通过增强尺度平衡机制融合两者来扩展域分布,既覆盖了整体图像风格差异,又保留了器官局部特征的一致性,在标注数据仅 10%~20% 时仍能保持高性能。

基于模型的增强机制即通过干预模型内部结构,实现更加灵活的样本合成与分布重构。这类策略通常依托于深度生

3.1.3 学习策略

浅层域泛化中的学习策略方法通过设计合理的训练机制,引导模型从源数据中学习具备更强迁移能力的表示,增强其在不同分布下的适应性。

典型的路径是样本重加权机制,其核心在于动态调整训练样本的重要性,引导模型更关注具有代表性或挑战性的样本,从而提升泛化能力。He 等^[33]通过模拟测试阶段可能出现的分布变化,引入扰动权重机制来增强模型对潜在边界样本的敏感性。该方法结构简洁且性能稳定,克服了传统方法对数据结构的强依赖。

成模型(如 GAN、变分自编码器、扩散模型)^[37-38],能够从训练语义出发在特征空间中生成具有目标域特征的样本。例如, Yu 等^[39]提出语义引导的扩散模型来解决结肠息肉检测与诊断中的域偏移问题。通过整合内镜报告文本等多模态临床标注,构建“粗粒度+细粒度”的组合提示机制,粗粒度提示捕捉息肉位置、大小等空间结构,细粒度提示编码纹理、病理类型等语义细节,从低频率结构特征到高频语义细节逐步生成伪样本,确保结构完整性与临床相关性,并通过提示重放机制保证模型在学习新标注时保留对原有语义的生成能力,适用于小样本和形态不规则的息肉等特殊场景。

另外,数据孤岛问题在医疗场景中广泛存在,部分研究开始将联邦机制嵌入增强策略中,以实现数据隐私保护与分布扩展的协同优化。例如, Che 等^[40]在联邦学习框架下,通过实例级对抗学习生成语义一致但风格新颖的图像,最大化原始与生成数据的特征差异,同时通过像素级和任务级语义约束保证语义一致性。实例级对抗学习无需依赖模糊的域标签,直接通过特征差异生成新颖风格数据,在医疗图像域间隙模糊的场景中依然有效。同时,基于模型“锐度”评估客户端模型的泛化贡献,通过客户端内和客户端间的分层聚合,缓解联邦场景下的模型贡献不平衡问题。尽管此类方法在形式上属于增强路径,但在建模逻辑上与训练优化机制深度耦合,部分研究也将其归类于学习策略范畴。

3.2.2 表示学习

在深层域泛化策略中,表示学习旨在构建具有稳定性、抽象性和语义一致性的特征空间,使模型能够在分布外数据上保持稳健的判别能力。与传统特征工程方法不同,深层表示学习不仅关注判别性能,更注重特征对跨域扰动的鲁棒性。近年来,相关研究主要围绕自监督预训练、解耦表示以及注意

力机制等方向展开。

自监督表示学习通过构建预定义任务(如拼图重建、对比学习等),在无需人工标签的前提下提取通用特征,降低对标注数据的依赖。针对电子健康记录的时间分布偏移问题,Guo等^[41]提出了电子健康记录基础模型,该模型基于Transformer和门控单元,运用自监督预训练策略,通过自回归任务,也就是预测患者下一天的医疗代码来学习数据中的全局模式。这种预训练方式使模型能够捕捉到电子健康记录序列数据中的复杂时间依赖关系和潜在规律,而无需依赖下游具体临床预测任务的标签。Wu等^[42]为解决医疗数据中的空间和时间域偏移,提出了用于域泛化的自学习框架,通过使用特征特定编码器将患者数据映射到多个潜在空间,无需域标签,通过层次聚类自动发现每个空间的潜在域,并为每个域训练分类器来提升模型的泛化能力。

解耦表示学习通过将特征空间划分为域不变的任务相关成分与域特定的无关成分,在保留关键语义信息的同时抑制干扰特征,广泛应用于图像分割与跨模态任务中。例如,Hong等^[43]提出“One Image as One Domain(OIOD)”假设,将每张图像视为独立域,通过内容编码器提取域不变的解剖结构特征,风格编码器提取域相关的风格特征,解码器则结合两者重建图像。在此过程中,通过膨胀目标区域掩码,引导解码器优先重建解剖边界及周围区域,增强分割精度,并随机生成风格编码,合成多样化风格图像,提升模型对风格变化的鲁棒性。Kim等^[44]则采用层次化Transformer作为编码器,分离域不变的内容特征和域相关的风格特征,结合StyleGAN2生成器合成跨域图像,并使用全局判别器和patch-wise判别器确保合成图像的真实性。

近年来,注意力机制的应用愈加广泛^[45]。这一机制在深层域泛化中通常被用于增强域不变特征的学习。例如,Gu等^[46]提出的域组合与注意力机制,通过分组卷积学习一组“基础域表示”,形成域表示库,并引入divergence约束,使基础表示多样化。域注意力模块的作用在于学习基础表示的线性组合系数,通过加权组合生成当前域的校准向量,用于特征图的动态校准,中和域偏差。Xu等^[47]提出了一种结合自适应特征混合和双交叉注意力正则化的域泛化方法,其中,自适应特征混合通过从均匀分布中随机采样特征统计量,然后将原始特征与增强特征沿通道维度混合,同时生成分布内和分布外样本;双交叉注意力正则化板块则利用交叉通道注意力机制,以原始深度特征为查询、生成特征为键值,重建鲁棒特征表示,同时反向地,以生成特征为查询、原始特征为键值,形成双向约束,引导模型学习域不变特征。

3.2.3 学习策略

在深层域泛化方法中,学习策略类技术通过优化训练机制,引导模型提升对分布外数据的适应能力,是增强模型鲁棒性的重要途径。图4展示了深层域泛化的分类,具体包括元学习、对抗训练、正则化策略、联邦学习策略与因果建模,表5详细列出了各深层域泛化方法的相同与不同之处以及其能解决的问题。

元学习旨在模拟“训练域-测试域”的泛化过程,通过构建元任务并在多个源任务上优化,使模型在遇到新域时能够

迅速适应。例如,Li等^[48]提出的一种基于episodic训练与任务增强的域泛化方法,基于模型无关元学习,通过模拟“元任务”实现知识迁移。每个元任务将源域划分为“元训练域”和“元测试域”,通过内外双层优化来提升模型对域偏移的适应能力。同时为了避免过拟合,提出“混合任务采样”,使元测试域不再仅来自于单一源域,而是通过插值多个源域的样本生成,增加元任务的多样性。

对抗训练通过引入域判别器,使特征提取器学习领域不可区分的表示,从而实现领域不变建模。例如,解耦自编码器^[49]通过对抗训练学习域不变特征表示,实现对多个域和任务的稳健泛化。该模型包含编码器、解码器、多个任务头和域头。编码器将输入数据映射到潜在空间,解码器重构输入,保留有用信息;任务头预测临床相关任务,域头识别域标签;同时引入梯度反转层,在训练时反转域头的梯度,使编码器学习域不变特征。

正则化策略则通过在训练目标中引入结构性或语义约束项,引导模型学习更稳定、泛化性更强的表示结构。例如,Matsun等^[50]在模型中引入互信息正则化项,通过最大化目标模型特征与大规模预训练模型特征之间的互信息,使目标模型的特征分布接近大规模预训练模型的泛化特征分布,同时保留分类任务的关键语义信息。

联邦学习策略在不共享原始数据的前提下,通过多中心协同训练实现跨域建模,是解决隐私与泛化双重挑战的关键技术路径。例如,Radwan等^[51]提出的MedUniverse框架,采用分阶段联邦学习,首先服务器通过掩码图像建模预训练基础模型,客户端下载该模型后用本地数据微调,再将微调权重上传到服务器聚合。Deng等^[52]提出联邦半监督域泛化,通过全局-局部协同优化实现跨域泛化。在全局层面,服务器采用泛化感知聚合策略,基于KL散度计算客户端模型与全局模型的“泛化差距”,对泛化能力弱的客户端赋予更高权重,平衡不同域的贡献;在局部层面,客户端通过双教师自适应伪标签优化融合全局模型和本地指数移动平均模型生成伪标签,并通过自适应不确定性阈值过滤噪声样本,同时引入扰动不变对齐机制,对编码器特征施加扰动,通过解码器输出一致性约束强制学习域不变特征。以上方法均基于联邦学习实现多中心协作,避免了原始数据共享,达到了保护隐私的效果。

域泛化的核心挑战是解决分布偏移,而因果机制在不同域中具有稳定性,因此成为提升泛化能力的关键工具^[53]。例如,Feder等^[54]提出的因果驱动的反事实数据增强方法,基于数据的因果结构,利用患者的用药、实验室结果等数据和大型语言模型生成反事实样本,打破标签与虚假属性的关联。Carvalho等^[55]提出的部分条件不变正则化,通过因果推断确保模型学习的表示,既不对不同环境的风格特征敏感,又保留了区分正常和异常的核心特征。两者均利用辅助信息来捕捉因果结构,基于因果推断框架,将数据生成过程建模为因果图,区分核心特征与虚假特征,并通过削弱虚假特征来提升性能。更加详细的因果策略介绍可参考;Sheth等^[56]将因果域泛化方法分为基于因果数据增强的不变性、基于因果表示学习的不变性和基于因果机制迁移的不变性3类,明确其在模型中的作用阶段,并分析了其在图像、文本和图数据中的应用。

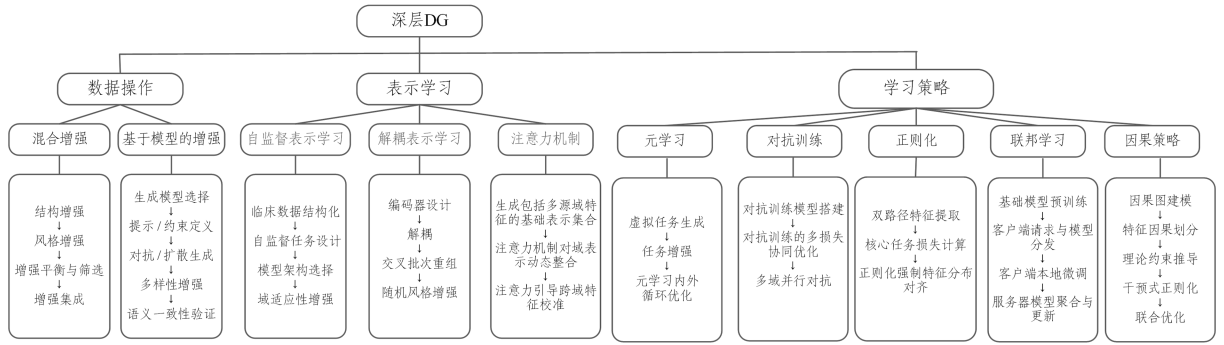


图4 深层 DG 的主要策略和流程

Fig. 4 Key strategies and processes of deep DG

表5 深层 DG 的核心策略比较

Table 5 Comparison of core strategies in deep DG

作者	核心策略	核心问题	优势	局限	适用场景
Liu 等	结构增强+风格增强	视网膜血管分割中的域偏移	结构增强针对性强;风格增强多样性高;泛化能力全面	计算成本高;结构增强依赖先验;对成像设备导致的深层风格差异覆盖不足	视网膜血管分割
Soleimani 等	测试时数据增强+多模态特征融合	ECG 跨域房颤检测的域偏移	轻量易部署;无需复杂的元优化过程,计算成本低	仅针对 ECG 信号设计,难以直接迁移	房颤检测
Ye 等	全局增强+局部区域增强+增强尺度平衡机制	医学图像分割标注数据少和域偏移	临床标注成本低;域扩散增强全面	复杂度高	多器官分割
Yu 等	组合提示引导的扩散模型+渐进式频谱扩散+持续学习策略	结肠息肉检测与诊断中的域偏移	语义引导的高保真性;多任务通用;针对性强	依赖多模态标注;计算成本高,生成效率低;语义偏差风险	小样本、形态不规则的息肉等
Che 等	新颖域对抗生成+锐度感知分层聚合	域偏移模拟不足,域标签模糊,泛化贡献不平衡	无需依赖模糊的域标签;缓解联邦学习中“贡献不平衡”问题	极端异质性适应不足;复杂度高;生成效率有限	隐私保护;小样本数 > 100
Guo 等	预训练学习全局模式,冻结权重迁移	电子健康记录中的时间分布偏移	时间偏移下鲁棒性强;随数据量扩展性能提升	校准性能异质性;依赖结构化电子健康记录代码	领域差异明确且偏移类型简单
Wu 等	解耦域发现+定制分类器	医疗数据中的空间和时间域偏移	无需域标签,自动发现细粒度潜在域;定制分类器适配不同患者特征	聚类质量影响性能;迭代训练增加计算成本	潜在领域多、偏移复杂且领域标签缺失
Hong 等	内容与风格解耦,通过多视图一致性和增强技术学习域不变特征	医学图像分割中的域偏移	无需域标签;EMA 增强边界分割;支持多模态数据	多损失函数训练复杂;极端风格偏移下可能失效	数据来源复杂、域标签缺失
Kim 等	内容与风格解耦,通过 Transformer 和生成模型强化特征分离	医学图像分割中的域偏移	特征解耦更彻底;生成器合成高质量跨域图像	计算成本高,小数据集上易过拟合;仅验证单模态	大样本单模态任务
Gu 等	并行域感知器+域注意力模块	医学图像分割中的域偏移	基础域表示库可复用,适配多场景;计算效率高	依赖基础表示的质量和数量;线性组合可能无法捕捉复杂域偏移	基础域已知且数量有限
Xu 等	自适应特征混合+双交叉注意力正则化	医学图像分割中的域偏移	分布外样本生成扩展了域覆盖范围;双向注意力约束;适用于多模态医学数据	计算成本高;依赖样本生成质量	域偏移较大
Li 等	Episodic 训练框架+任务增强+深层语义嵌入正则化	医学影像分类中的域偏移	缓解任务级过拟合;适用于多种医学影像任务	计算复杂度;参数敏感性;依赖源域多样性;未验证极端域偏移	多源域到单目标域;源域有限
Kreuter 等	解耦自编码器模型(编码器、解码器、多个任务头和域头)+梯度反转层	真实临床数据中的多域和多任务泛化	可通过增减任务头或者域头适应新城或者新任务,也可修改编码器或解码器适配表格数据或图像,适用性广泛	可能丢失部分细节;训练不稳定;依赖源域多样性	多域、多任务的复杂场景
Matsun 等	互信息正则化	糖尿病视网膜病变分类中的域偏移	性能优异;通用性强;实现简单	依赖 oracle 质量;不适用于小样本	数据充足、域偏移中等的糖尿病视网膜病变多类分类任务
Radwan 等	联邦学习框架+掩码图像建模+模型聚合与信任度验证	多中心医学影像分类中的域偏移	隐私保护彻底;多任务通用性	计算复杂度;依赖预训练数据;聚合偏差风险	多模态;保护数据隐私
Deng 等	联邦半监督学习+泛化感知聚合+双教师自适应伪标签优化+扰动不变对齐	联邦半监督医学影像分割中的域偏移	有效利用无标签数据;局部优化精细;模型稳健	超参数敏感;计算成本高;任务局限	针对分割任务;保护隐私
Feder 等	因果驱动的反事实数据增强	模型依赖虚假相关性导致的分布偏移	所需样本量小;适用于数据稀缺场景;可解释性强	辅助数据依赖;依赖大型语言模型生成质量	文本分类等可生成合理样本的场景
Carvalho 等	因果视图建模+部分条件不变正则化	模型依赖环境虚假相关性的导致的分布偏移	通用性强;无需生成样本;避免过度正则化	依赖环境标签;正则化权重敏感;不适用于文本等复杂模态	异常检测等样本生成困难的场景

4 方法对比与挑战

4.1 域泛化 vs 域适应

近年来,域适应方法凭借其相对成熟的技术体系和丰富的实证研究,在跨中心、跨设备的医学影像等临床建模任务中应用广泛。其优势在于能够利用目标域的数据,显式地学习源域到目标域的映射或特征对齐,从而有效缓解分布偏移问题,在临床实践中往往能获得较高的目标域精度。因此,域适应方法在此领域已有较多成功实践,它通过学习个体或者任务不变的特征表示来提升模型在新情境下的鲁棒性。脑电图信号是域适应方法在临床的重要应用领域之一。脑电图信号特征存在显著的个体差异和任务间分布偏移,其建模任务契合域适应设定,在癫痫检测^[57]、情绪识别^[58]与运动想象^[59]等领域已有较为广泛的实践。然而,样本获取受限等问题使得域适应方法在实际部署时面临显著瓶颈。

相比之下,域泛化方法在训练阶段完全规避了对目标域数据的依赖,其目标是从源域中学习领域不变的本质特征,以期在未见过的目标域上保持鲁棒性。这一特性使域泛化在目标域数据完全不可及或目标环境动态变化难以预知的临床场景中展现出独特的应用潜力。然而,域泛化的效能高度依赖源域数据的多样性和覆盖范围,它要求源域必须能充分反映潜在的目标域变异,这对临床数据集的构建提出了更苛刻的要求。目前,域泛化在临床应用的实证基础相对域适应较弱,其在复杂未知目标域上的实际精度往往面临更大的不确定性。针对域适应和域泛化在各方面的区别,表6进行了详细总结。

表6 DA和DG的对比

Table 6 Comparison between DA and DG

对比项	域适应	域泛化
目标域访问权限	可访问目标域数据 (无标签/少量标签)	无需访问目标域数据
源域多样性要求	较弱	较强
泛化性	较弱 (对特定目标域过拟合)	较强 (要求源域多样性强)
适用场景	环境变化可预知	环境变化不可预知
精确性	较强	较弱
隐私挑战	较大	较小
方法成熟度	较高	发展中

4.2 浅层方法 vs 深层方法

在模型架构层面,深层模型与浅层模型的对比同样需要考虑临床数据特性。深层模型的优势在于其强大的非线性特征表达能力,能够自动从原始高维、异构且常伴有噪声的临床数据中提取分层、抽象且更具判别性的鲁棒特征。这种能力使其在处理复杂临床模式识别任务时,通常显著优于依赖手工特征的浅层方法。深度学习的“端到端”特性也简化了特征工程流程,减少了对手工设计特征的依赖。但深层模型的资源消耗问题在临床场景中十分突出。例如,深层模型的参数规模巨大,基层医院的低算力服务器往往难以承受,且深层模型的训练需要大规模的标注数据,基层医院的单中心样本量根本无法满足。针对上述问题,研究者设计了一些轻量化的解决方案,例如通过知识蒸馏将深层模型的知识蒸馏到浅层模型^[60-61],或者进行模型剪枝^[62],也就是通过L1正则化减去

深层模型中的冗余,保留核心的域对齐能力等。

浅层模型的核心挑战在于其特征表示的瓶颈。它们高度依赖精心设计的手工特征工程,这需要丰富的领域知识,且构造的特征往往泛化能力有限,可扩展性不足,难以适应临床数据中复杂多变的环境和个体差异。但浅层模型在小样本场景下的数据效率、模型可解释性以及计算开销方面仍具优势,在资源受限或需要模型透明度的临床决策支持场景中使用更多。

4.3 临床方法与通用方法的区别

临床的DA/DG方法和通用DA/DG方法的差异主要表现在数据特点、约束条件和技术目标3个方面。首先,通用的DA/DG方法面对的大多是规模大、标注成本低、特征空间相对同构的数据集,分布偏移较为平滑;而临床方法处理的数据跨模态、标注成本极高、罕见病等小样本问题突出。其次,通用的DA/DG方法对隐私和可解释性要求较低,而临床方法必须满足“数据可用不可见”的隐私保护需求,对隐私和可解释性要求高,且临床方法的决策直接关系到患者生命健康,错误代价极高。最后,通用方法传统上以处理静态数据为主,而临床方法需要兼顾动态时序数据适配以及多模态融合和极端场景下的模型稳健性。

因此,研究者对通用的DA/DG方法进行了创新。例如,通过联邦学习^[51-52,63]、差分隐私等技术解决数据隐私与协同训练的矛盾,结合知识图谱、临床指南等注入医学先验知识,增强模型在小样本和跨模态情况下的泛化能力,开发动态的域适应机制实现对动态病程的持续适应,添加注意力可视化等可解释性模块^[46-47,64]来满足临床决策的可追溯性等。

4.4 临床应用中的挑战

域适应与域泛化方法在缓解协变量偏移、提升跨域泛化能力上颇具成效,但由于临床场景的特殊性和高约束性,其在实际的临床数据建模任务中仍然面临诸多挑战。

首先,在数据层面,除标准化难度和标注匮乏外,时序动态性和极端样本特性使得DA和DG方法的应用更加困难。在电子健康记录等时序数据中,不同中心的采样频率、时序特征定义差异会导致“时序分布偏移”随时间累积^[65-66],而传统域适应方法多针对静态数据,难以建模时序依赖关系的跨域对齐^[67]。同时,临床数据中罕见病、疑难病例等极端样本在单一源域中占比极低,使得域泛化模型难以学习到跨域共性特征。多模态数据的“模态内偏移”与“模态间关联偏移”也是挑战之一;单模态内部存在设备差异,模态间关联也可能因临床实践差异产生偏移,现有方法对这种关联一致性的建模能力明显不足。

其次是“概念偏移”。同一疾病的诊断标准可能因机构指南更新时差存在差异,产生“概念偏移”,而传统分布对齐方法易形成“伪对齐”^[68]。临床数据分布会随着治疗方案、设备等的更新发生变化,静态域泛化模型难以实时跟踪这种变化,而增量学习机制又可能引发“灾难性遗忘”。不同数据类型的偏移特性差异也需针对性应对:电子健康记录的特征空间异构性会使同构特征对齐方法失效;基因组数据的高维度(百万级特征)导致域泛化模型易受噪声干扰,如Ren等^[19]在单细胞数据研究中,即使采用多对抗机制,跨批次的基因表达谱对齐

仍会损失部分关键生物信号。相比之下,统一格式的医学影像数据的偏移更易被修正。

再次,模型性能与临床逻辑性的平衡也是挑战之一。过度对齐可能丢失临床关键信息,基于统计相关性的分布对齐可能学习跨越“伪相关”特征,使模型决策缺乏临床因果支撑^[69-71]。在小样本场景下,域适应过程易受噪声样本影响。浅层模型表达能力有限,深层模型的复杂结构又无法满足临床对可解释性的核心需求。

最后是“数据孤岛”问题。联邦域适应和域泛化可以作为解决方案,但仅交换模型参数会导致对齐不充分,采用同态加密传输特征又会使计算开销增加,无法满足急诊场景的实时性要求。并且,医院现有医院信息系统、实验室信息系统的接口差异要求模型额外开发适配层,而复杂的域适应模块存在推理延迟。

结束语 本文回顾了域适应和域泛化方法在临床数据建模中的研究进展,明确区分了浅层与深层策略。浅层域适应方法主要集中于实例加权 and 特征匹配,而深层域适应围绕对抗学习、分布差异和重构展开。浅层和深层域泛化方法则主要从数据增强、表示学习和学习策略三大技术路径进行分析。针对医学图像、电子健康记录、脑电图等任务场景,总结了当前方法在泛化性能、数据依赖性与可解释性等方面的表现差异。

浅层方法结构简洁、部署成本低,但存在表达力不足与泛化局限的问题。深层方法在特征建模、结构鲁棒性与机制可扩展性方面具有更高潜力,特别是在结合因果建模、联邦学习与自监督机制的研究中,深层方法正逐步拓展其在临床领域的实际应用能力。但深层方法也存在计算资源需求较高、可解释性较低的难题。

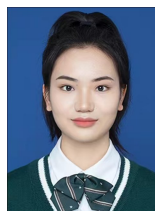
未来研究可从以下方向进一步推动进展:(1)发展可解释性强、训练稳定的跨域模型架构;(2)结合大模型与预训练策略提升泛化与迁移性能;(3)构建统一评估标准与数据共享平台,促进方法可复现性与对比研究。通过机制建模与临床需求的深度融合,有望构建更鲁棒、高效且可信的医疗 AI 系统。

参 考 文 献

- [1] HAH H, GOLDIN D S. How Clinicians Perceive Artificial Intelligence-Assisted Technologies in Diagnostic Decision Making: Mixed Methods Approach[J]. *Journal of Medical Internet Research*, 2021, 23(12): e33540.
- [2] LIANG J Y, QIAN Y H, LI D Y, et al. Granular Computing Theory and Methods for Big Data Mining[J]. *Scientia Sinica: Informationis*, 2015, 45(11): 1355-1369.
- [3] GAO S, HUANG X Y, WEI X Y, et al. Insulator Defect Detection Method Based on Domain Knowledge Transfer[J]. *Science Technology and Engineering*, 2023, 23(14): 6054-6062.
- [4] TAO J, XU H. Discovering Domain-Invariant Subspace for Depression Recognition by Jointly Exploiting Appearance and Dynamics Feature Representations [J]. *IEEE Access*, 2019, 7: 186417-186436.
- [5] GOETZ M, WEBER C, BINCZYK F, et al. DALSA: Domain Adaptation for Supervised Learning From Sparsely Annotated MR Images[J]. *IEEE Transactions on Medical Imaging*, 2016, 35(1): 184-196.
- [6] CHEN K, LIU J, WAN R, et al. Unsupervised Domain Adaptation for Low-Dose CT Reconstruction via Bayesian Uncertainty Alignment[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2024, 36(5): 8525-8539.
- [7] ZHAO X, SICILIA A, MINHAS D S, et al. Robust White Matter Hyperintensity Segmentation On Unseen Domain[C]// 2021 IEEE 18th International Symposium on Biomedical Imaging (IS-BI). IEEE, 2021: 1047-1051.
- [8] XIA J, ZENG L, LI T, et al. PKDG: Prior Knowledge based Domain Generalization Model for fundus image segmentation[J/OL]. *Research Square*, 2024. <https://doi.org/10.21203/rs.3.rs-4703557/v1>.
- [9] CARLONI G, TSAFTARIS S A, COLANTONIO S, CROCODILE: Causality Aids ROBustness via COntRastive DISentangled LEarning[C]// International Workshop on Uncertainty for Safe Utilization of Machine Learning in Medical Imaging. Cham: Springer, 2025: 105-116.
- [10] WACHINGER C, REUTER M. Domain Adaptation for Alzheimer's Disease Diagnostics[J]. *NeuroImage*, 2016, 139: 470-479.
- [11] CHEPLYINA V, PENA I P, PEDERSEN J H, et al. Transfer Learning for Multicenter Classification of Chronic Obstructive Pulmonary Disease[J]. *IEEE Journal of Biomedical and Health Informatics*, 2018, 22(5): 1486-1496.
- [12] XIA K, NI T, YIN H, et al. Cross-Domain Classification Model With Knowledge Utilization Maximization for Recognition of Epileptic EEG Signals[J]. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 2021, 18(1): 53-61.
- [13] LI W, ZHAO Y, CHEN X, et al. Detecting Alzheimer's Disease on Small Dataset: A Knowledge Transfer Perspective[J]. *IEEE Journal of Biomedical and Health Informatics*, 2019, 23(3): 1234-1242.
- [14] JIANG L, LIU S, MA Z, et al. Regularized RKHS-Based Subspace Learning for Motor Imagery Classification[J]. *Entropy*, 2022, 24(2): 195.
- [15] PENG Y, WANG W, KONG W, et al. Joint feature adaptation and graph adaptive label propagation for cross-subject emotion recognition from EEG signals[J]. *IEEE Transactions on Affective Computing*, 2022, 13(4): 1941-1958.
- [16] CUI J, JIN X, HU H, et al. Dynamic Distribution Alignment With Dual-Subspace Mapping for Cross-Subject Driver Mental State Detection[J]. *IEEE Transactions on Cognitive and Developmental Systems*, 2022, 14(4): 1705-1716.
- [17] GANIN Y, USTINOVA E, AJAKAN H, et al. Domain-Adversarial Training of Neural Networks[M]// Domain Adaptation in Computer Vision Applications. Cham: Springer, 2017: 189-209.
- [18] JIE L, LIANG P, ZHAO Z, et al. ADAN: An Adversarial Domain Adaptation Neural Network for Early Gastric Cancer Prediction[C]// 2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), 2022: 2169-2172.
- [19] REN P, SHI X, YU Z, et al. Single-cell assignment using multi-

- ple-adversarial domain adaptation network with large-scale references[J]. *Cell Reports Methods*, 2023, 3(9):100577.
- [20] SOMERS P, HOLDENRIED-KRAFFT S, ZAHN J, et al. Cystoscopic depth estimation using gated adversarial domain adaptation[J]. *Biomedical Engineering Letters*, 2023, 13(2):141-151.
- [21] QIN X, BUI F, HAN Z. Semantically preserving adversarial unsupervised domain adaptation network for improving disease recognition from chest x-rays[J]. *Computerized Medical Imaging and Graphics*, 2023, 107:102232.
- [22] BAO G, ZHUANG N, TONG L, et al. Two-Level Domain Adaptation Neural Network for EEG-Based Emotion Recognition[J]. *Frontiers in Human Neuroscience*, 2021, 14:605246.
- [23] LIU W, NI Z, CHEN Q, et al. Attention-Guided Partial Domain Adaptation for Automated Pneumonia Diagnosis From Chest X-Ray Images[J]. *IEEE Journal of Biomedical and Health Informatics*, 2023, 27(12):5848-5859.
- [24] HONG X, ZHENG Q, LIU L, et al. Dynamic Joint Domain Adaptation Network for Motor Imagery Classification[J]. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 2021, 29:556-565.
- [25] OFER A, OPHIR A, YOAV N, et al. Supervised autoencoder denoiser for non-stationarity in multi-session EEG-based BCI[J]. *Journal of Neural Engineering*, 2025, 22(2):026043.
- [26] CUI H, LI Y, WANG Y, et al. Toward Accurate Cardiac MRI Segmentation With Variational Autoencoder-Based Unsupervised Domain Adaptation[J]. *IEEE Transactions on Medical Imaging*, 2024, 43(8):2924-2936.
- [27] JIANG K, QUAN L, GONG T. Disentangled representation and cross-modality image translation based unsupervised domain adaptation method for abdominal organ segmentation[J]. *International Journal of Computer Assisted Radiology and Surgery*, 2022, 17(6):1101-1113.
- [28] WANG J, LAN C, LIU C, et al. Generalizing to Unseen Domains: A Survey on Domain Generalization[J]. *IEEE Transactions on Knowledge and Data Engineering*, 2023, 35(8):8052-8072.
- [29] GUO B, LU D, SZUMEL G, et al. The impact of scanner domain shift on deep learning performance in medical imaging: an experimental study[J]. *arXiv:2409.04368*, 2024.
- [30] ATHALYE C, ARNAOUT R. Domain-guided data augmentation for deep learning on medical imaging[J]. *PLOS One*, 2023, 18(3):e0282532.
- [31] KAISTI M, LAITALA J, WONG D, et al. Domain randomization using synthetic electrocardiograms for training neural networks[J]. *Artificial Intelligence in Medicine*, 2023, 143:102583.
- [32] AYODELE K P, IKEZOGWO W O, KOMOLAFE M A, et al. Supervised domain generalization for integration of disparate scalp EEG datasets for automatic epileptic seizure detection[J]. *Computers in Biology and Medicine*, 2020, 120:103757.
- [33] HE Y, SHEN X, XU R, et al. Covariate-shift generalization via random sample weighting[C]// *Proceedings of the AAAI Conference on Artificial Intelligence*. 2023:11828-11836.
- [34] LIU B, ZHANG Y, WANG S, et al. DGSSA: Domain generalization with structural and stylistic augmentation for retinal vessel segmentation[J]. *arXiv:2501.03466*, 2025.
- [35] SOLEIMANI M, TOOSI M H, MOHAMMADI S, et al. Using Test-Time Data Augmentation for Cross-Domain Atrial Fibrillation Detection from ECG Signals[J]. *arXiv:2503.13483*, 2025.
- [36] YE Z. SSL-DG: rethinking and fusing semi-supervised learning and domain generalization in medical image segmentation[J]. *arXiv:2311.02583*, 2023.
- [37] PATIL A, MEHTO A, NALBAND S. Enhancing skin lesion diagnosis with data augmentation techniques: a review of the state-of-the-art[J]. *Multimedia Tools and Applications*, 2025, 84:25325-25364.
- [38] APELLANIZ P A, PARRAS J, ZAZO S. Improving Synthetic Data Generation Through Federated Learning in Scarce and Heterogeneous Data Scenarios[J]. *Big Data and Cognitive Computing*, Multidisciplinary Digital Publishing Institute, 2025, 9(2):18.
- [39] YU J, ZHU Y, FU P, et al. Robust Polyp Detection and Diagnosis through Compositional Prompt-Guided Diffusion Models[J]. *arXiv:2502.17951*, 2025.
- [40] CHE H, WU Y, JIN H, et al. FedDAG: Federated Domain Adversarial Generation Towards Generalizable Medical Image Analysis[J]. *arXiv:2501.13967*, 2025.
- [41] GUO L L, STEINBERG E, FLEMING S L, et al. EHR foundation models improve robustness in the presence of temporal distribution shift[J]. *Scientific Reports*, 2023, 13(1):3767.
- [42] WU Z, YAO H, LIEBOVITZ D, et al. An Iterative Self-Learning Framework for Medical Domain Generalization[J]. *Advances in Neural Information Processing Systems*, 2023, 36:54833-54854.
- [43] HONG J, LIU B, LONG G, et al. Rethinking domain generalization in medical image segmentation: One image as one domain[J]. *arXiv:2501.04741*, 2025.
- [44] KIM H, SHIN Y, HWANG D. DiMix: Disentangle-and-Mix Based Domain Generalizable Medical Image Segmentation[C]// *Medical Image Computing and Computer Assisted Intervention—MICCAI 2023*. Cham: Springer, 2023:242-251.
- [45] QIN W Q, BAO X, CHEN X, et al. A method for predicting the synergy of anticancer drug combinations based on graph attention network [J]. *Journal of Nantong University(Natural Science Edition)*, 2025, 24(1):10-17.
- [46] GU R, ZHANG J, HUANG R, et al. Domain Composition and Attention for Unseen-Domain Generalizable Medical Image Segmentation[C]// *Medical Image Computing and Computer Assisted Intervention—MICCAI 2021*. Cham: Springer, 2021:241-250.
- [47] XU Y, ZHANG T. Boundless Across Domains: A New Paradigm of Adaptive Feature and Cross-Attention for Domain Generalization in Medical Image Segmentation[J]. *arXiv:2411.14883*, 2024.
- [48] LI C, LIN X, MAO Y, et al. Domain generalization on medical imaging classification using episodic training with task augmentation[J]. *Computers in Biology and Medicine*, 2022, 141:105144.
- [49] KREUTER D, TULL S, GILBEY J, et al. Dis-AE: Multi-domain & Multi-task Generalisation on Real-World Clinical Data

- [J]. arXiv:2306.09177,2023.
- [50] MATSUN A, MOHAMED D O, CHOKUWA S, et al. DGM-DR: Domain Generalization with Mutual Information Regularized Diabetic Retinopathy Classification[C]// Domain Adaptation and Representation Transfer. Cham: Springer, 2024: 115-125.
- [51] RADWAN A, OSMAN I, SHEHATA M S. Universal medical imaging model for domain generalization with data privacy[J]. arXiv:2407.14719,2024.
- [52] DENG Z, XU Z, ISSHIKI T, et al. FedSemiDG: Domain Generalized Federated Semi-supervised Medical Image Segmentation [J]. arXiv:2501.07378,2025.
- [53] CHEN K, WANG X H, SHI X L, et al. A graph neural network method for predicting rutting in asphalt pavements based on causal inference [J]. Journal of Nantong University (Natural Science Edition), 2025, 24(1): 18-27, 50.
- [54] FEDER A, WALD Y, SHI C, et al. Data augmentations for improved (large) language model generalization[J]. Advances in Neural Information Processing Systems, 2023, 36: 70638-70653.
- [55] CARVALHO J, ZHANG M, GEYER R, et al. Invariant anomaly detection under distribution shifts: a causal perspective[J]. Advances in Neural Information Processing Systems, 2023, 36: 56310-56337.
- [56] SHETH P, MORAFFAH R, CANDAN K S, et al. Domain Generalization—A Causal Perspective[J]. arXiv:2209.15177,2022.
- [57] CUI X, CAO J, LAI X, et al. Cluster Embedding Joint-Probability-Discrepancy Transfer for Cross-Subject Seizure Detection[J]. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 2023, 31: 593-605.
- [58] CHAI X, WANG Q, ZHAO Y, et al. Multi-subject subspace alignment for non-stationary EEG-based emotion recognition[J]. Technology and Health Care: Official Journal of the European Society for Engineering and Medicine, 2018, 26(S1): 327-335.
- [59] JIANG L, LIU S, MA Z, et al. Regularized RKHS-Based Subspace Learning for Motor Imagery Classification[J]. Entropy, 2022, 24(2): 195.
- [60] EKONG F, YU Y, PATAMIA R A, et al. RetVes segmentation: A pseudo-labeling and feature knowledge distillation optimization technique for retinal vessel channel enhancement[J]. Computers in Biology and Medicine, 2024, 182: 109150.
- [61] CUI H, YUWEN C, JIANG L, et al. Bidirectional cross-modality unsupervised domain adaptation using generative adversarial networks for cardiac image segmentation[J]. Computers in Biology and Medicine, 2021, 136: 104726.
- [62] KAUR S, GUMP A, XIAO Y, et al. CRoP: Context-wise Robust Static Human-Sensing Personalization[C]// Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies. 2025: 1-34.
- [63] BERNECKER T, PETERS A, SCHLETT C L, et al. Fednorm: Modality-based normalization in federated learning for multimodal liver segmentation[J]. arXiv:2205.11096,2022.
- [64] WANG S, YU L, LI K, et al. DoFE: Domain-Oriented Feature Embedding for Generalizable Fundus Image Segmentation on Unseen Datasets[J]. IEEE Transactions on Medical Imaging, 2020, 39(12): 4237-4248.
- [65] WONG A, OTLES E, DONNELLY J P, et al. External validation of a widely implemented proprietary sepsis prediction model in hospitalized patients [J]. JAMA Internal Medicine, 2021, 181(8): 1065-1070.
- [66] GUO L L, PFOHL S R, FRIES J, et al. Evaluation of domain generalization and adaptation on improving model robustness to temporal dataset shift in clinical medicine[J]. Scientific Reports, 2022, 12(1): 2726.
- [67] COHEN S N, FOSTER J, FOSTER P, et al. Subtle variation in sepsis-III definitions markedly influences predictive performance within and across methods[J]. Scientific Reports, 2024, 14(1): 1920.
- [68] ZHENG J, CAI X, QIU S, et al. Spurious forgetting in continual learning of language models[J]. arXiv:2501.13453,2025.
- [69] ZECH J R, BADGELEY M A, LIU M, et al. Variable generalization performance of a deep learning model to detect pneumonia in chest radiographs: a cross-sectional study[J]. PLoS Medicine, 2018, 15(11): e1002683.
- [70] CARUANA R, LOU Y, GEHRKE J, et al. Intelligible Models for HealthCare: Predicting Pneumonia Risk and Hospital 30-day Readmission[C]// Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 2015: 1721-1730.
- [71] DEGRAVE A J, JANIZEK J D, LEE S I. AI for radiographic COVID-19 detection selects shortcuts over signal [J]. Nature Machine Intelligence, 2021, 3(7): 610-619.



CHEN Xiu, born in 2001, master. Her main research interest is intelligent medical engineering.



ZHANG Yuanpeng, born in 1984, professor, Ph.D supervisor. His main research interest is medical artificial intelligence.