Wave2Word：用于神经重症监护中脑电图-文本对齐与多任务表征学习的多模态Transformer框架

Argha Kamal Samanta[a,1], Deepak Mewada[b,1], Monalisa Sarma[c], Debasis Samanta[b]

[a]印度理工学院克勒格布尔分校电子与电气通信工程系，印度西孟加拉邦克勒格布尔，*721302*[b]印度理工学院克勒格布尔分校计算机科学与工程系，印度西孟加拉邦克勒格布尔，*721302*[c]印度理工学院克勒格布尔分校苏比尔·乔杜里质量与可靠性学院，印度西孟加拉邦克勒格布尔，*721302*

arXiv:2601.00670v1 [cs.HC] 2 Jan 2026

摘要

连续脑电图（脑电图）常用于神经重症监护，以监测癫痫发作及其他有害脑活动，包括常见且具有临床意义的节律性和周期性模式。尽管深度学习方法已在癫痫发作检测方面取得较高准确率，但现有方法大多仍以癫痫发作为中心，依赖离散标签监督，且主要采用基于准确率的指标进行评估。当前脑电图建模实践的一个核心局限在于，学习得到的表征与临床工作流程中对脑电检查结果的解读和总结之间对应关系较弱。有害脑电活动表现出重叠模式、分级专家共识和时间持续性，仅靠分类目标难以充分捕捉这些特征。本研究提出一种多模态脑电图表征学习框架，将信号域建模与结构化临床语言监督相结合。首先，将原始脑电图转换为纵向双极导联和时频表示。其次，采用双路基于Transformer的编码器分别建模互补的时间中心型和以频率为中心的依赖关系，并通过自适应门控机制进行融合。第三，通过对比目标，将脑电图嵌入与结构化专家共识描述对齐。最后，在标准分类损失之外，引入脑电图条件文本重建损失作为表征级约束。在受控的训练-验证-测试划分下进行实验评估，六分类测试准确率达到0.9797。消融分析显示，移除对比对齐后，跨模态检索性能从Recall@10为0.3390降至0.0045，而分类准确率几乎没有变化。这些发现表明，对于具有临床意义的脑电图建模，判别准确率并不能可靠反映表征质量。与结构化临床语言进行显式对齐，能够揭示基于准确率的评估无法观察到的语义组织和优化行为差异，为超越分类性能评估脑电图表征提供了有据可依的基础。

关键词： 脑电图，神经重症监护，有害脑活动，多模态学习，脑电图—语言对齐

1. 引言

在现代重症监护病房中，持续脑电图（EEG）监测常规用于监测急性脑损伤、昏迷或不明原因意识状态改变的患者；在这些情况下，仅依靠临床检查并不充分。大型多中心前瞻性研究报告称，接受持续脑电图监测的危重成人中，约有 10–40% 出现脑电图发作，其中超过 80% 为非惊厥性发作，因此临床上无明显表现 [1, 2, 3, 4]。除癫痫发作外，节律性和周期性脑电模式更为常见，包括局灶性周期性放电（LPDs）、全身性周期性放电（GPDs）、侧化节律性δ活动（LRDA）和全身性节律性δ活动（GRDA）。多个重症监护病房队列报告称，超过半数接受监测的患者出现这些模式 [5, 6, 7, 8]。

这些脑电图模式具有重要的临床意义。既往研究表明，节律性或周期性脑电活动与死亡率升高、重症监护病房住院时间延长、神经功能恢复延迟及功能结局不良独立相关 [6, 3, 9, 10, 11]。不良结局与模式负荷、持续时间和时间演变的相关性，比与单一类别异常是否存在的相关性更强 [4, 12, 13]。与此同时，患者的持续脑电图监测通常持续 24–72 小时甚至更久，因此产生了大量判读工作，需要专家持续审阅 [14, 8]。这些因素促使人们开发自动化工具，以支持而非取代临床判读。

这一场景中的一项关键挑战是，有害脑电活动并非由一组彼此离散且互斥的类别构成。Clinical Neurophysiology 文献将癫痫发作、节律性活动和周期性放电描述为处于发作期–发作间期连续谱上，其特征是逐渐过渡、形态重叠以及随时间演变 [5, 15, 9]。脑电图模式可能在同一段记录中共存，也可能随时间在节律性、周期性和发作状态之间转换 [4, 11]。因此，专家解读往往依赖于相对主导性，

电子邮箱： arghakamal25@gmail.com（Argha Kamal Samanta），deepakmewada96@gmail.com（Deepak Mewada），monalisa@iitkgp.ac.in（Monalisa Sarma），dsamanta@iitkgp.ac.in（Debasis Samanta）[1]这些作者对本文的贡献相同。

持续性和时间背景，而非简单判断其是否存在。

这种复杂性也体现在专家一致性上。即使是经验丰富的神经生理学家，对节律性和周期性脑电模式的评定者间一致性也仅为中等水平，报告的 Cohen κ系数通常介于 0.4 至 0.6 之间[16, 6, 17]。为应对这种差异，美国临床神经生理学会（ACNS）引入了标准化术语，用于界定有害脑电模式，同时明确编码主导性、不确定性和专家意见分歧 [15]。在常规临床实践中，脑电检查结果通常以专家间的一致率汇总，而非给出确定性标签，这反映了解读的概率性质 [18, 17]。忽略这种结构的计算表述与临床报告规范不符。

近年来，用于脑电图自动分析的机器学习方法发展迅速。卷积神经网络、时序卷积模型和基于 Transformer 的架构在癫痫发作检测、睡眠分期和通用脑电图分类任务中表现出色 [19, 20, 21, 22, 23]。然而，系统综述显示，大多数研究仍以癫痫发作为中心，或采用二元异常—正常表述[21, 24]。LPDs、GPDs、LRDA 和 GRDA 等非癫痫发作性有害模式常被合并到粗略的异常类别中，或被排除在模型输出之外 [25, 26]。

这种建模选择与临床脑电图实践存在不匹配。离散标签优化假设类别边界清晰，而临床医生会对模式的混合、分级置信度和时间演变进行推理 [5, 15]。因此，仅针对分类准确率进行优化的模型可能学到具有判别性、但不足以支持报告级解读或不确定性沟通的表征。

更广泛地说，许多现有的脑电图学习框架依赖于与临床工作流程相冲突的简化假设。尽管已有证据表明脑电图异常具有连续性，但这些异常仍被视为离散实体。尽管脑电图解读通过结构化语言加以外化，脑电图仍被建模为单模态的信号处理问题。尽管专家意见分歧是临床术语中固有且明确编码的特征，监督仍被假定为标签完备。这些假设压制了不确定性，限制了对表征失败模式的洞察，同时更重视基准测试性能，而非临床对齐。

这种脱节在脑电图解读的沟通方式上表现得最为明显。在临床实践中，脑电检查结果通过书面报告传达，使用标准化语言描述主要和次要模式、时间背景以及专家意见一致性，而非数值评分 [15, 27]。语言使临床医生能够表达不确定性、传达概率性判断，并将脑电检查结果与更广泛的临床背景相结合 [17, 18]。不支持或无法保留这种描述性结构的自动化系统，无法对脑电图解读的最终输出进行建模。

多模态表征学习的最新进展指明了一种可能的方向。研究表明，将高维信号与语言对齐，可以在视觉和医学影像领域生成具有语义结构的嵌入 [28, 29, 30, 31, 32]。将语言与生物医学数据相结合的基础模型进一步表明，语言监督能够超越仅依赖标签的训练，对表征进行正则化[33, 34, 35]。尽管取得了这些进展，脑电图在这一范式中仍未受到足够关注，尽管脑电图解读本质上是由语言介导的。

受这些观察结果的启发，本研究探讨将脑电图表征与结构化专家临床语言对齐，是否能为有害脑电活动建模提供有益约束。其核心前提是，同时支持区分能力和临床连贯描述的表征，或许能更好地反映实践中脑电图的判读与记录方式。从这一视角来看，语言对齐被视为一种表征探针，而非最终应用。

所提方法概述.. 所提出的框架模拟了临床脑电图判读的关键步骤。首先，将原始头皮脑电图转换为纵向双极导联，以减少参考依赖伪影并凸显局部电压梯度。随后，对各通道进行短时傅里叶变换，以捕捉与美国临床神经生理学会定义的模式相关的节律和频谱结构。双路基于Transformer的编码器分别建模以时间为中心和以频率为中心的互补依赖关系，从而实现长程上下文整合。利用对称对比目标，将脑电图嵌入与结构化专家临床描述对齐，促使其依据临床语义而非仅依据标签身份进行组织。此外，辅助的脑电图条件文本解码器会重建专家风格的摘要，并作为表征层面一致性约束。

报告层面的示例输出。以下示例展示了这种表述方式的含义，并对比了专家共识描述与模型生成的摘要：

样例01
真实标注： 专家意见完全一致，均指出存在局灶性周期性放电（100%）。
生成结果： 专家意见完全一致，均指出存在局灶性周期性放电（100%）。

样例02
真实标注： 专家意见不一，8%指出存在偏侧性节律性δ活动，54%指出存在广泛性节律性δ活动，38%指出存在其他模式。

生成结果： 专家意见显示一致性不高，其中 8 % 认为是局灶性节律性δ活动，54 % 认为是全身性节律性δ活动，38 % 认为是其他模式。

在此，文本生成并不被视为最终应用，而是一种严格的探针，用于揭示脑电图嵌入是否在类别身份之外保留了具有临床意义的语义。

贡献..... 本研究的贡献总结如下：

1. 提出一种具有临床依据的框架，用于建模美国临床神经生理学会定义的六类有害脑电活动，以弥补以往研究过度关注癫痫发作的不足。2. 提出一种多模态脑电图—文本表征学习方法，将脑电图嵌入与结构化专家临床描述对齐。3. 采用联合优化策略，结合分类、对比对齐和脑电条件文本重构，对表征学习加以约束。4. 在受试者独占协议下，通过准确率、跨模态对齐、架构消融和优化稳定性评估表征质量。

论文结构。 第2 节回顾相关文献。第3 节介绍所提出的框架。第 4 节呈现实验结果。第5节讨论其影响与局限性。

2. 相关工作

脑电图（EEG）自动分析已得到广泛研究，其动因在于需要处理长时间记录，而在相关应用场景中，专家解读既耗时又受资源限制。本节回顾与神经重症监护脑电相关的既有研究，重点探讨现有方法如何表征脑电图、定义监督方式以及评估性能。讨论侧重于具体的建模实践，而非重复临床动机。

神经重症监护中的脑电图自动分析。 早期脑电图自动分析系统依赖于从时域统计量、谱功率、相干性指标、熵和启发式规则中提取的手工特征 [36, 37, 38]。这些方法通常针对癫痫发作检测、爆发抑制识别或背景异常筛查等有限目标而设计。其性能高度依赖专家定义阈值，并且容易受到噪声、电极配置和受试者间变异的影响，因此在异质性较强的重症监护病房环境中鲁棒性有限 []。

后续的大型观察性研究和前瞻性研究证实，脑电图发作、周期性放电和节律性脑电图模式在危重患者中十分常见；若识别或治疗延迟，则与不良神经系统结局相关 [1, 5, 6, 3, 7]。这些发现推动了重症监护病房自动脑电图监测系统的开发。然而，在这一场景中研究或部署的大多数自动化方法都侧重于癫痫发作检测或异常—正常二元判别。临床定义的非癫痫发作模式，例如局灶性或全身性周期性放电以及节律性δ活动，往往被合并为粗略类别，或被排除在模型输出之外 [25, 26]。

卷积神经网络与局部上下文建模。深度学习的出现使脑电图分析从手工设计特征转向数据驱动的表示学习。卷积神经网络（CNN）表明，可以直接从原始或经过少量预处理的信号中学习判别性脑电图表征。DeepConvNet 和 ShallowConvNet 等架构为通用脑电解码任务建立了强大的基线 [19], 而 EEGNet 则强调参数效率以及对低资源环境的适用性[20]。

基于卷积神经网络的模型已广泛应用于癫痫发作检测、睡眠分期和脑机接口任务[39, 40, 41, 24]。其显著特点是采用固定的局部感受野。虽然这种设计对短时事件十分有效，但对于由持续性、规律性和时间演变等特征定义、且跨越较长时间窗口的脑电图现象，其建模能力有限，例如周期性放电或持续性节律活动 [5, 9, 11]。因此，大多数基于卷积神经网络的研究都在短时片段上进行，并使用离散标签，未能明确建模临床相关时长范围内的长程时间结构。

基于*Transformer*的模型与长上下文脑电编码。Transformer架构的引入旨在解决卷积模型时间背景有限的问题。多项研究报告称，使用Transformer编码器或混合CNN–Transformer架构，在脑电情绪识别、睡眠分期和一般分类任务中取得了更好的性能 [23, 42, 43]。自注意力能够灵活整合不同时间点的信息，这对非平稳脑电信号尤为有利。

尽管具有这种表征能力，基于Transformer的脑电模型仍大多是单模态的，通常采用离散分类目标进行训练。此类模型在神经重症监护脑电中的应用有限，且很少对周期性放电或节律性δ活动等临床定义的模式类别进行显式建模。评估主要依赖准确率或曲线下面积指标，对表征结构、鲁棒性或与临床判读的一致性的分析则较为有限 [21, 22, 24]。

时频表示与多视图融合。时频分析，包括短时傅里叶变换和小波分解，常用于表征脑电中的振荡结构和瞬态事件 [44, 45]。基于频谱图运行，或结合时域与频域输入的深度学习模型，已在多项脑电任务中展现出更好的性能 [46, 40, 47]。

多视图和多分支架构通过融合互补的空间、时间和频谱表征来拓展这一思路 [48, 49, 50]。这些方法通常依赖拼接或固定加权等静态融合策略。尽管临床证据表明，癫痫发作、周期性放电和节律性活动在时间与频谱组织结构上存在根本差异[5, 15]，但这类设计并不会根据模式特定的信号特征调整表征的主导性。

多模态学习与脑电–语言整合。多模态表征学习已成为医学人工智能的核心，尤其体现在放射学和病理学中的图像–文本模型。语言对齐表征有助于报告生成、跨模态检索和结构化

解读 [30, 31, 32]。CLIP 等对比学习框架表明，与语言对齐能够生成语义组织良好的嵌入，超越仅依赖标签的监督 [28, 29]。

语言对齐基础模型进一步表明，语言监督可以在生物医学领域中规范表征 [33, 34, 35]。相比之下，在这一范式中，脑电图受到的关注有限。现有的脑电图—文本研究规模较小，依赖弱监督或事后监督，而且很少使用标准化临床术语 [51, 52]。在自动脑电图研究中，脑电图表征与专家临床语言之间的显式对齐仍不常见。

ACNS 定义的脑电图模式类别覆盖范围。在脑电图机器学习文献中，各类别的覆盖范围明显失衡。已发表的研究主要集中于癫痫发作检测，而其他由美国临床神经生理学会术语定义的重要临床模式，包括局灶性和全身性周期性放电以及节律性δ活动，则代表性不足 [15, 25]。即使纳入研究，这些模式也常被合并为宽泛的异常类别，或作为次要标签处理 [26]。这种覆盖范围的局限，使我们难以了解现代架构如何表征不同的有害脑电模式，以及表征方式的选择如何影响临床相关场景中的可分性。

评估实践与表征层面分析。脑电图深度学习研究中的评估主要依赖准确率和曲线下面积等点估计指标。多项综述指出了一些长期存在的问题，包括受试者泄漏、对训练动态的报告有限，以及鲁棒性分析不足 [21, 22, 24]。在更广泛的医学人工智能领域，研究者也提出了类似担忧，认为以准确率为中心的评估对于临床部署存在局限性[53, 54]。

很少有脑电图研究考察学习得到的表征是否与临床语义对齐、是否支持不确定性沟通，或能否在架构扰动下保持稳定。尽管表征层面分析和跨模态评估对于向临床工作流程转化具有重要意义，但目前仍不常见。

总结。表 1 总结了具有代表性的既有工作，并突出了反复出现的方法学特征，包括以癫痫发作为中心的任务定义、单模态信号建模、离散标签监督、静态特征融合以及以准确率为中心的评估。尽管自动脑电图分类已取得重大进展，但对于临床定义的有害脑电活动，其语义对齐、类别覆盖范围和表征质量仍存在亟待解决的问题。

3. 所提方法

所提方法旨在学习有害脑活动的表征，使其具备良好的分类判别能力，同时与使用标准化临床术语总结脑电检查结果的方式保持一致。在本文中，对齐是针对基于专家共识投票形成的结构化专家摘要描述来定义的，而非自由形式叙述推理。该框架将脑电图的信号域建模与结构化临床语言监督相结合。每个架构组件都对应于脑电图解读中常用的一个独立分析步骤。

图1 展示了所提流程的完整概览。方法描述遵循图中从左到右的顺序，依次介绍脑电图信号构建、时频变换、双视图表征学习、自适应融合、多模态对齐和联合优化。

3.1. 问题表述与学习目标

设$\mathbf{X}_{\text{raw}} \in \mathbb{R}^{C_{\text{raw}} \times T_{\text{raw}}}$ 表示以频率f_s采样的多通道脑电图记录，其中C_{raw} 为头皮电极的数量，T_{raw} 为时间采样点的数量。每个脑电图片段对应一个经临床标注的事件，并与一个离散标签相关联

$$y \in \mathcal{Y} = \{0, 1, 2, 3, 4, 5\}, \tag{1}$$

该标签表示癫痫发作、偏侧性周期性放电、全身性周期性放电、偏侧性节律性δ活动、全身性节律性δ活动以及其他模式。

每个脑电图片段还配有一条结构化专家意见描述，该描述由专家共识投票得出。该描述表示为一个词元序列

$$\mathbf{t} = (w_1, w_2, \ldots, w_L), \tag{2}$$

其中，w_i 表示固定词汇表中的一个词元，L 表示最大序列长度。

学习目标联合优化三个相关任务：(i) 根据脑电图对有害脑活动进行分类，(ii) 对脑电图表征与结构化专家摘要进行对齐，以及 (iii) 根据脑电图嵌入重建专家摘要。这些目标反映了临床实践中脑电检查结果的分类与沟通方式，并为图 1 所示的多模态结构提供了依据。

3.2. 脑电图输入构建与时频表征

框架的第一阶段将原始脑电图记录转换为结构化表征，同时保留空间关系、时间背景和频谱内容。此阶段引入的误差或偏差会传递至下游组件，因此需要采用明确且以生理学为依据的预处理方法。

3.2.1. 纵向双极导联构建

原始脑电通道被转换为纵向双极导联配置（LBM），这是临床脑电图判读中使用的标准表示方式 [55]。双极导联突出局部电压梯度，并降低对参考依赖伪影的敏感性。

表1：脑电图自动分析及相关多模态学习领域的代表性既往研究，按照临床范围、脑电图表示、监督范式，以及相对于神经重症脑电图解读尚未解决的方法学问题进行整理。

研究（年份、发表场所）	临床范围	脑电图表示	学习范式	已观察到的方法学局限
[36] (1982，临床脑电图 *Neurophysiol*)	癫痫发作检测	手工设计的时序特征	基于规则的检测	狭窄的任务范围；需要大量专家调优；在不同患者中的鲁棒性有限
[37] (2004，*Clin Neurophysiol*)	重症监护病房脑电图筛查	频谱特征与启发式方法	基于规则的系统	对导联配置和噪声敏感；缺乏结构化的脑电图模式分类体系
[38] (2006，*IEEE TBME*)	EEG 异常检测	时频特征	统计学习	依赖特征工程；跨领域能力有限
[1] (2004，《神经病学》)	重症监护病房癫痫发作负荷	临床脑电图判读	观察性研究	数据集泛化能力 确立患病率和结局；未进行自动化建模
[5]（2012，《临床神经生理学杂志》）	发作期–发作间期连续谱	临床脑电图形态学	概念性 / 观察性	化建模 展示脑电图模式的连续性；但未在机器学习表述中得到解决
[19]（2017，*Human Brain Mapping*）	通用脑电图解码-ing	原始脑电图、卷积神经网络（深度-卷积网络、浅层卷积网-vNet）	监督分类	局部感受野；短时序上下文；非重症监护病房任务重点
[20] (2018年，*J Neural Eng*)	脑机接口、通用脑电图	紧凑型卷积神经网络（脑电图-Net）	监督分类	以准确率为中心；单模态；无不确定性或语义建模
[39] (2017，*Clin Neurophysiol*)	新生儿癫痫发作检测测	基于卷积神经网络的分类器	监督分类	仅关注癫痫发作；特定数据集验证
[40](2018年，*Comput Biol Med*)	癫痫发作检测	短时傅里叶变换 + 卷积神经网络	监督分类	狭窄的任务界定；排除非癫痫发作脑电图模式
[25] (2020年，*Clin Neurophysiol*)	重症监护病房脑电图异常-ity	深度卷积神经网络	异常/正常二元分类	将美国临床神经生理学会定义的脑电图模式归并为粗粒度标签
[26] (2021年，*IEEE TBME*)	重症监护病房脑电图监测器-ing	基于卷积神经网络的深度学习-ing	二元分类	未区分节律性或周期性脑电图模式
[22] (2019年，*J Neural Eng*)	癫痫发作检测	CNN/RNN混合模型	监督分类	对长程时间演化的建模有限-变
[21]（2019年，*IEEE TNNLS*）	脑电图深度学习	评述	系统分析	指出了受试者泄漏、鲁棒性有限
[23] (2022年，*NeuroImage*)	综述情绪识别 tion	Transformer编码器	监督分类	分析不足、临床依据薄弱 非临床任务；单模态；仅准确率评估
[42] (2023年，*IEEE TBME*)	脑电图模式识别-识	基于Transformer的架构	监督分类	非重症监护病房数据集；无美国临床神经生理学会模式定位
[48] (2021年，*NeuroImage*)	脑电图分类	多视图卷积神经网络	静态特征融合	非模式自适应融合
[49] (2022年，*IEEE TBME*)	脑状态分类	多分支深度学ing	监督分类	启发式融合；无语义或语言监督——监督
[50] (2023，*IEEE TNNLS*)	通用脑电图分类——分类	时频卷积神经网络 (TSception)	监督分类	特征多样性有所提升；仍是单模态，且以准确率为中心
[51] (2022年，*MICCAI*)	脑电图–文本表征表征学习	多模态深度学习ing	弱对齐	规模有限；未采用 ACNS 术语；仅进行定性评估性评估
[30] (2019年，*AAAI*)	放射学报告	图像-文本对齐	多模态学习	展现报告级建模的价值（非脑电图领域）
[28] (2021年，*ICML*)	视觉-语言（通用）	对比对齐（CLIP）	多模态对比学习	未应用于脑电图；启发了信号—语言对齐范式
[33] (2019年，*ACL*)	临床自然语言处理	Bio-ClinicalBERT	语言预训练	未整合生理信号
[34] (2023, *Nature*)	医学基础模型	多模态表征式	语言对齐学习ing	脑电图研究代表性不足；重点主要放在影像和电子健康记录
[35] (2023, *Nature*)	多模态临床AI	基础模型范式	以语言为中心的监督式sion	在脑电图或神经重症领域的应用有限care

令$\mathbf{X}_{C_1}$和$\mathbf{X}_{C_2}$表示构成纵向双极导联对的成对电极信号。双极脑电信号的计算方式为

$$\mathbf{X}_{\text{bip}} = \mathbf{X}_{C_1} - \mathbf{X}_{C_2}, \tag{3}$$

得到$\mathbf{X}_{\text{bip}} \in \mathbb{R}^{C\times T}$。每个通道表示解剖位置相邻的电极之间的电位差，从而提高空间可解释性以及对全局噪声的鲁棒性。

3.2.2. 围绕标注的 *Offset* 进行窗口提取

每个脑电图片段均围绕临床标注的偏移时间 τ 提取。给定固定的窗口时长 T_w 秒，提取的片段定义为

$$\mathbf{X} = \mathbf{X}_{\text{bip}}[:,\, t_0 : t_0 + T_w f_s], \tag{4}$$

其中，t_0 的选取使窗口以 τ 为中心。这确保了各样本具有一致的时间背景，并使输入符合专家标注实践。

3.2.3. 时频变换

许多病理性脑电图模式由节律性、频率成分和时间规律性定义，而不是由绝对波形形态定义。为对这些特性进行显式建模，本文对每个通道应用短时傅里叶变换。

对于通道 c，其短时傅里叶变换定义为

$$\mathcal{S}_c(f,t) = \sum_{n=0}^{N-1} x_c[n]\, w[n-t]\, e^{-j2\pi f n}, \tag{5}$$

其中，$w(\cdot)$ 是一个汉宁窗。保留幅度后得到一个

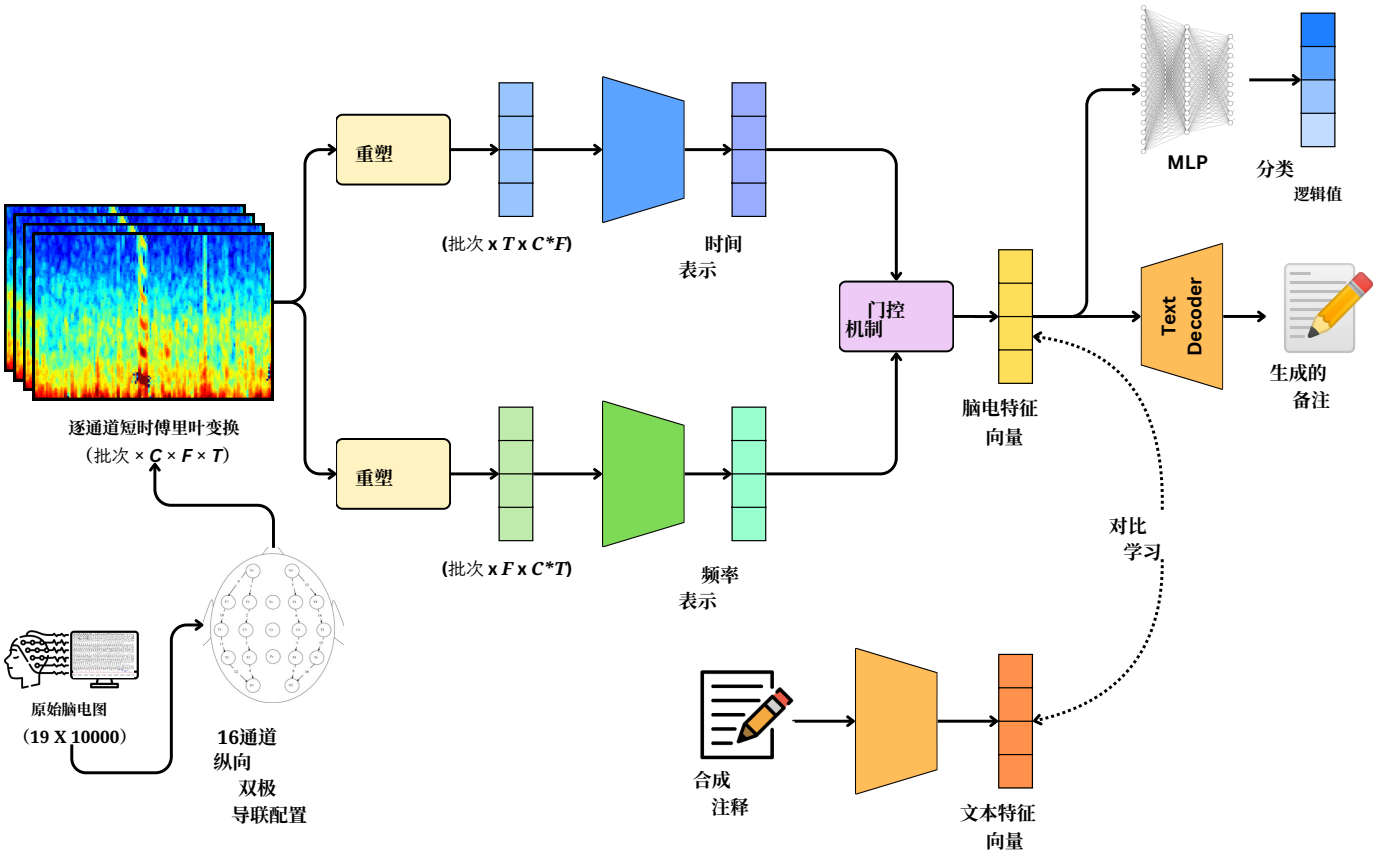


图1：所提出的多模态脑电图—文本学习框架概览。原始脑电图信号被转换为纵向双极导联，并围绕临床标注的偏移进行分段。逐通道短时傅里叶变换生成时频表示，随后由两个互补的基于Transformer的编码器分支进行处理，分别侧重于以时间为中心和以频率为中心的依赖关系。自适应门控机制将这些表征融合为统一的脑电图嵌入。与此同时，基于专家共识投票生成的结构化临床笔记由特定领域的语言模型进行编码。脑电图嵌入与文本嵌入通过对比学习进行对齐，同时辅助解码器从脑电图嵌入重建临床文本。最终的脑电图嵌入用于有害脑活动六分类。

时频张量

$$\mathbf{Z} \in \mathbb{R}^{C\times F\times T_f}. \tag{6}$$

时频表示是脑电图分析中的标准方法，因为周期性放电和节律性δ活动的特征是持续的频谱结构 [44, 45]。张量 **Z**是双视图编码器的共享输入。

3.2.4. 频谱图域数据增强

为提高鲁棒性，训练期间对**Z** 应用随机数据增强，包括时间掩蔽、频率掩蔽、添加高斯噪声、幅度缩放和时间平移。数据增强在频谱图域进行，以保留具有生理意义的结构，同时增强模型对轻微扰动的不变性。

3.3. 双视图脑电图编码

张量 **Z** 包含交织的空间、时间和频谱依赖关系。使用单个编码器对所有依赖关系进行联合建模在计算上效率低下，并会削弱归纳偏置。因此，该框架在编码之前将 **Z** 分解为两个互补视图。

3.3.1. 时间中心视图与频率中心视图

时间中心视图构造如下

$$\mathbf{Z}_t \in \mathbb{R}^{T_f\times(CF)}, \tag{7}$$

强调各通道随频率变化的时间演变。频率中心视图构建如下：

$$\mathbf{Z}_f \in \mathbb{R}^{F\times(CT_f)}, \tag{8}$$

强调跨通道和时间聚合的特定频率结构。

这种分解方式与临床脑电图判读相呼应：在整合之前，时间演变和节律结构会分别加以考量。

3.3.2. 线性投影与*Transformer*编码

每个视图都被投影到维度为 d 的潜在空间中：

$$\mathbf{H}_t = \mathbf{Z}_t\mathbf{W}_t, \quad \mathbf{H}_f = \mathbf{Z}_f\mathbf{W}_f. \tag{9}$$

每个投影序列都由一组Transformer编码器层 [56]进行处理。位置嵌入保留

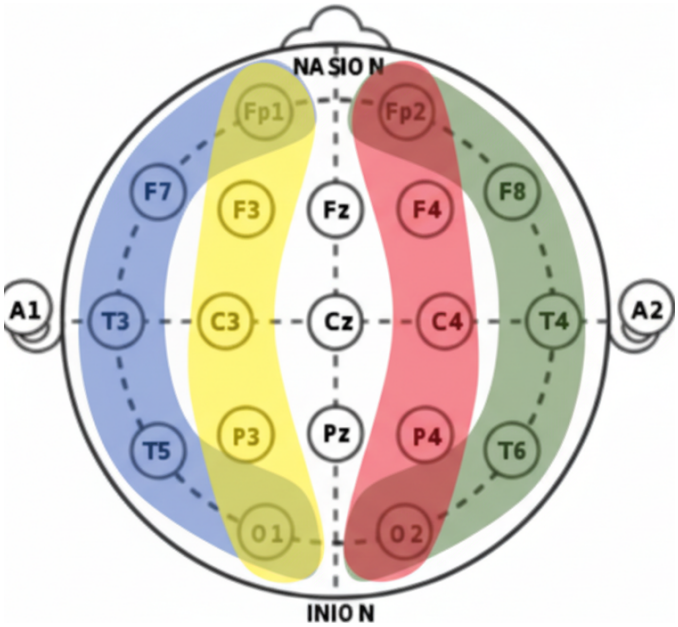


图2：纵向双极导联（LBM）的示意图，该导联配置用于生成脑电图通道。每个双极通道表示解剖位置相邻的电极之间的电压差，突出局部空间梯度并减少参考依赖伪影。

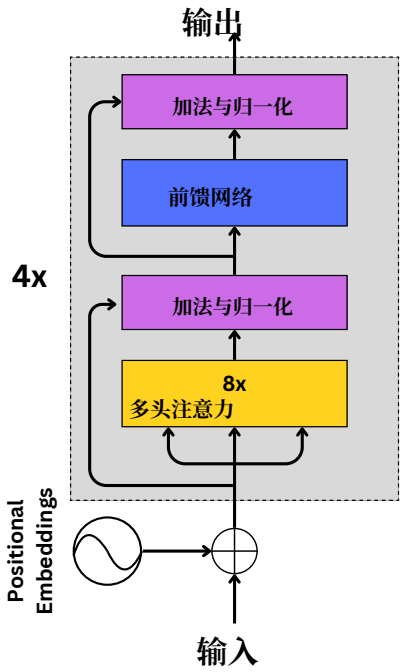


图 3：用于以时间为中心和以频率为中心的脑电图分支的 Transformer 编码器结构。每个编码器由堆叠的多头自注意力层组成，并包含位置编码、前馈网络、残差连接和层归一化。

顺序信息，而自注意力则支持长距离交互：

$$\mathrm{Attn}(\mathbf{Q},\mathbf{K},\mathbf{V}) = \mathrm{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^\top}{\sqrt{d}}\right)\mathbf{V}. \tag{10}$$

3.3.3. 全局聚合

逐词元输出通过平均池化进行聚合：

$$\mathbf{h}_t = \frac{1}{C}\sum_{i=1}^{C}\mathbf{H}_t^{(i)}, \quad \mathbf{h}_f = \frac{1}{F}\sum_{i=1}^{F}\mathbf{H}_f^{(i)}. \tag{11}$$

3.4. 自适应门控与特征融合

时间嵌入和频率嵌入通过可学习的、针对样本的门控机制进行融合。拼接后的嵌入

$$\mathbf{h}_{\mathrm{cat}} = [\mathbf{h}_t;\mathbf{h}_f] \tag{12}$$

通过以下方式映射为融合系数：

$$(\mu_t,\mu_f) = \mathrm{softmax}(\mathbf{W}_g\mathbf{h}_{\mathrm{cat}}). \tag{13}$$

最终的脑电图嵌入为

$$\mathbf{h}_{\mathrm{eeg}} = [\mu_t\mathbf{h}_t;\mu_f\mathbf{h}_f]. \tag{14}$$

门控机制支持跨异质脑电图模式的表征灵活性，并非旨在保证分类准确率单调提升。

3.5. 临床文本建模与多模态对齐

结构化临床描述是根据专家投票分布，使用标准化模板确定性生成的。这些描述通过 Bio-ClinicalBERT 进行编码，从而得到

$$\mathbf{h}_{\mathrm{text}} = \mathbf{W}_p\,\mathrm{Bio_ClinicalBERT}_{\mathrm{CLS}}(\mathbf{t}), \tag{15}$$

其中仅对最终的编码器层进行微调 [33]。

临床描述为合成文本，并遵循固定格式。因此，文本模态对脑电图表征学习起到受控语义约束的作用，而非语言变异性的来源。脑电图嵌入与文本嵌入通过对称对比损失 [28]进行对齐。

3.6. 脑电图条件文本重建

自回归 Transformer 解码器根据$\mathbf{h}_{\mathrm{eeg}}$重建临床文本。重建损失为

$$\mathcal{L}_{\mathrm{rec}} = -\sum_{i=1}^{L}\log p(w_i \mid w_{<i},\mathbf{h}_{\mathrm{eeg}}). \tag{16}$$

文本重建用作诊断探针，用于评估脑电图嵌入是否保留了类别身份之外与临床相关的描述性信息；它并非旨在作为独立的报告生成系统。

3.7. 分类与联合优化

使用多层感知机对脑电图嵌入进行分类：

$$\hat{y} = \mathrm{softmax}(\mathbf{W}_c\mathbf{h}_{\mathrm{eeg}}). \tag{17}$$

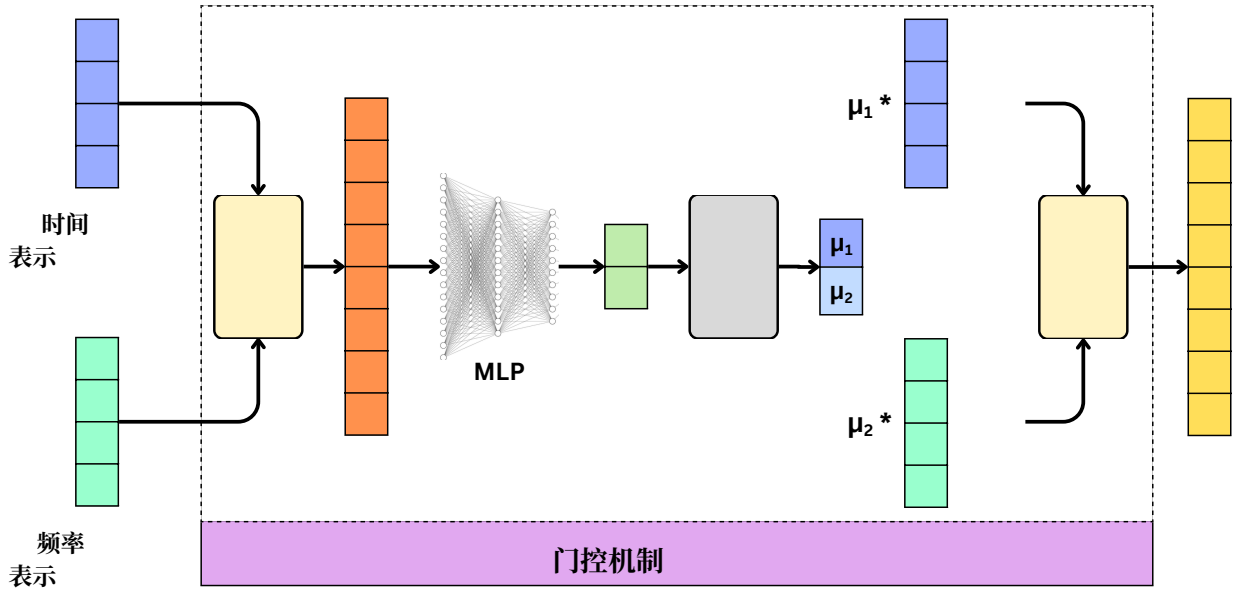


图 4：用于融合以时间为中心和以频率为中心的脑电图表征的自适应门控机制。可学习系数按样本确定各分支对最终脑电图嵌入的相对贡献。

分类损失为

$$\mathcal{L}_{\text{cls}} = -\log p(y \mid \mathbf{h}_{\text{eeg}}). \tag{18}$$

所有目标均进行联合优化：

$$\mathcal{L} = \lambda_1 \mathcal{L}_{\text{cls}} + \lambda_2 \mathcal{L}_{\text{con}} + \lambda_3 \mathcal{L}_{\text{rec}}, \tag{19}$$

其中 $\lambda_i = \exp(\alpha_i)$ 是可学习正标量 [57]。可学习权重用于平衡各目标之间的梯度幅值，并稳定多任务优化。

4. 实验与实验结果

本节在严格的受试者独占协议下，对所提出的框架进行定量评估。分析考察六分类准确率、多模态对齐、架构敏感性、参数效率和优化行为。我们结合表格和图示解读结果，以避免仅依据单一指标得出结论。由于总体准确率的变化范围较窄，因此采用相对差异报告效应量，以突出有意义的趋势。

4.1. 实验目标

实验评估围绕以下五个目标展开：

1. 量化六分类有害脑活动分类性能。2. 使用跨模态检索指标评估脑电–文本对齐。3. 通过消融分析各架构组件的贡献。4. 分析性能与模型复杂度之间的权衡。5. 考察训练稳定性和收敛行为。

4.2. 数据集

所有实验均使用*HMS – 有害脑活动分类（HBAC）*数据集 [58]，，该数据集由哈佛医学院、临床数据动画中心（CDAC）、Persyst 和 Jazz Pharmaceuticals 联合开发。该数据集包含从约 1,950 名接受重症监护和连续神经生理监测的患者中采集的 17,089 份独立脑电图记录。数据集提供了 106,800 个带标签的 50 秒脑电图片段，统一采样频率为 200 Hz；每个片段均配有一张同步频谱图，覆盖以带标签区间为中心的 10 分钟时间窗口。信号采集自 19 个标准 10–20系统电极（Fp1、F3、C3、P3、F7、T3、T5、O1、Fz、Cz、Pz、Fp2、F4、C4、P4、F8、T4、T6、O2）以及一个心电图（EKG）参考通道，因此每个样本包含 20 通道双极记录。癫痫病学家专家依据美国临床神经生理学会（ACNS）术语，对每个 50 秒窗口的中央 10 秒进行标注，并通过投票划分为六类具有临床意义的类别：癫痫发作（SZ）、侧化周期性放电（LPD）、广泛性周期性放电（GPD）、侧化节律性δ活动（LRDA）、全身性节律性δ活动（GRDA）和其他。多数票决定最终共识标签（专家共识），同时保留各类别的个人投票数（癫痫发作投票、LPD 投票、GPD 投票、LRDA 投票、GRDA 投票、其他投票），用于量化评定者间一致性。最终类别分布较为均衡，包括 20,933 个癫痫发作样本、18,861 个 GRDA 样本、18,808 个其他样本、16,702 个 GPD 样本、16,640 个 LRDA 样本和 14,856 个 LPD 样本。元数据文件 `train.csv`中的每一行都指定了脑电图片段（脑电图 ID、脑电图子 ID）、其在完整记录中的时间偏移（脑电图标签偏移秒数）以及对应的患者标识符（患者 ID）。原始脑电图信号以单独的`.parquet` 文件形式存储在`train_eegs/`目录下，

列名表示电极通道，另有一列为心电图（EKG）。

4.3. 实验设置

所有实验均使用基于 PyTorch 实现的统一训练与评估流程。该配置旨在确保不同消融实验间的架构可比性、结果的可复现性，以及高容量模型的稳定优化。

硬件环境。 训练与评估均在 NVIDIA H100 GPU 上进行。该加速器支持处理长脑电序列和高分辨率时频表示的 Transformer 编码器。所有报告的运行时间和收敛行为均对应此硬件配置。

软件栈。 所有模型均使用 PyTorch 和启用 CUDA 的后端实现。启用自动混合精度训练，以提高计算效率并降低内存消耗，同时保持数值稳定性 [59]。

训练方案。 所有变体均训练固定数量的训练轮次。不采用提前停止。此设计确保观察到的差异源于架构或目标层面的变化，而非训练时长。优化采用 AdamW，并使用解耦权重衰减 [60]。除非消融实验明确调整，否则所有实验均采用相同的学习率调度方案和正则化参数。

批处理与数据加载.. 所有运行均使用固定批次大小加载脑电图片段。数据加载采用并行处理，以避免输入瓶颈。在训练集划分中按受试者进行打乱，以确保训练、验证和测试受试者之间相互独立。

优化目标.. 完整模型优化由分类、对比对齐和文本重建损失组成的联合目标。在消融实验中，移除单个组成部分，同时保持其余目标和权重不变。这样可以单独评估各目标的贡献。

初始化与可复现性.. 所有实验均使用相同的随机种子配置。权重初始化方案和优化器超参数保持不变。按照既定的可复现性准则 [61]，评估使用在最后一个训练轮次结束时获得的检查点。

评估方案.. 在留出的测试集上评估性能。所有变体均使用相同的评估脚本计算测试准确率和检索指标。不应用测试时增强、集成或校准。

4.4. 定量结果总结

表2 总结了所有模型变体的结果。评估指标包括测试准确率、跨模态检索性能、可训练参数数量、验证集损失和训练时长。该表为后续分析提供了数值参考。

4.5. 消融效果全局视图

图5 汇总展示了消融实验结果。各子图将在后续小节中进行分析。

4.6. 分类准确率与架构敏感性

如图所示。5(a)，各变体的测试准确率介于 0.9720 至 0.9811 之间，相对差异约为 0.93%。精简型 *Transformer* 和无数据增强 变体的性能下降最为明显，这表明表征深度以及对多样化信号实现的接触程度，是分类鲁棒性的主要决定因素。

移除门控机制后，准确率略有提升。鉴于模型的工作区间较窄，这一提升只对应少量额外的正确预测，并表明对于该数据集，固定融合足以优化决策边界。

仅 EEG 模型仅使用 1.97M 个参数，即达到 0.9766 的准确率，相当于基线性能的 99.27%，而参数量仅约为基线的 5.7%。

4.7. 跨模态检索与表征结构

图 5（b）报告了脑电图–文本检索性能。完整模型的 Recall@10 达到 0.3390。移除对比目标后，Recall@10 降至 0.0045，相对下降 98.7%，尽管分类准确率几乎没有变化。这表明判别性准确率与语义对齐之间存在明显的解耦。

移除文本重建会使 Recall@K 评估指标整体下降，表明重建能够细化对齐，而不是决定对齐。移除门控会使检索性能略有下降，揭示了峰值分类准确率与语义灵活性之间的权衡。

4.8. 模型复杂度与效能权衡 *ffs*

图 5（c）展示了参数扩展的边际收益递减。从 1.97M 增加参数至 34.38M，准确率相对提升约 0.32%。

无文本重建变体将参数数量减少了 50% 以上，同时保留了大部分分类和检索性能，表明其在效率与性能之间实现了良好平衡。相比之下，在参数数量相近的情况下减少 Transformer 深度会导致性能下降，凸显了架构结构相较于原始容量的重要性。

4.9. 训练动态与优化稳定性

图 5（d）和（e）展示了完整模型平滑且高度耦合的训练曲线与验证曲线。损失迅速收敛，验证准确率与训练准确率 closely 跟随，表明优化过程稳定。

图6 比较了各变体的收敛情况。未采用多模态监督的模型初期收敛更快，但验证方差更高，表明优化稳定性较差。

表 2：受试者互斥评估下的消融研究结果。评估指标包括测试准确率、脑电图–文本检索性能（召回率@K）、可训练参数数量、验证损失和训练轮数。

模型	Acc.	R@1	R@5	R@10	参数	验证损失
完整模型（基线）	0.9797	0.0694	0.2263	0.3390	34.38M	2.3×10^{-6}
无对比损失	0.9784	0.0000	0.0018	0.0045	34.38M	7.6×10^{-7}
无文本重建	0.9743	0.0654	0.2191	0.3210	16.34M	2.2×10^{-6}
仅脑电图（无文本）	0.9766	0.0000	0.0000	0.0000	1.97M	6.2×10^{-7}
冻结的 Bio-ClinicalBERT	0.9793	0.0252	0.1073	0.1799	20.20M	2.1×10^{-6}
无门控机制	0.9811	0.0712	0.2259	0.3287	34.38M	2.3×10^{-6}
精简型 Transformer	0.9720	0.0667	0.2250	0.3291	33.85M	2.2×10^{-6}
无数据增强	0.9725	0.0694	0.2209	0.3269	34.38M	8.3×10^{-6}

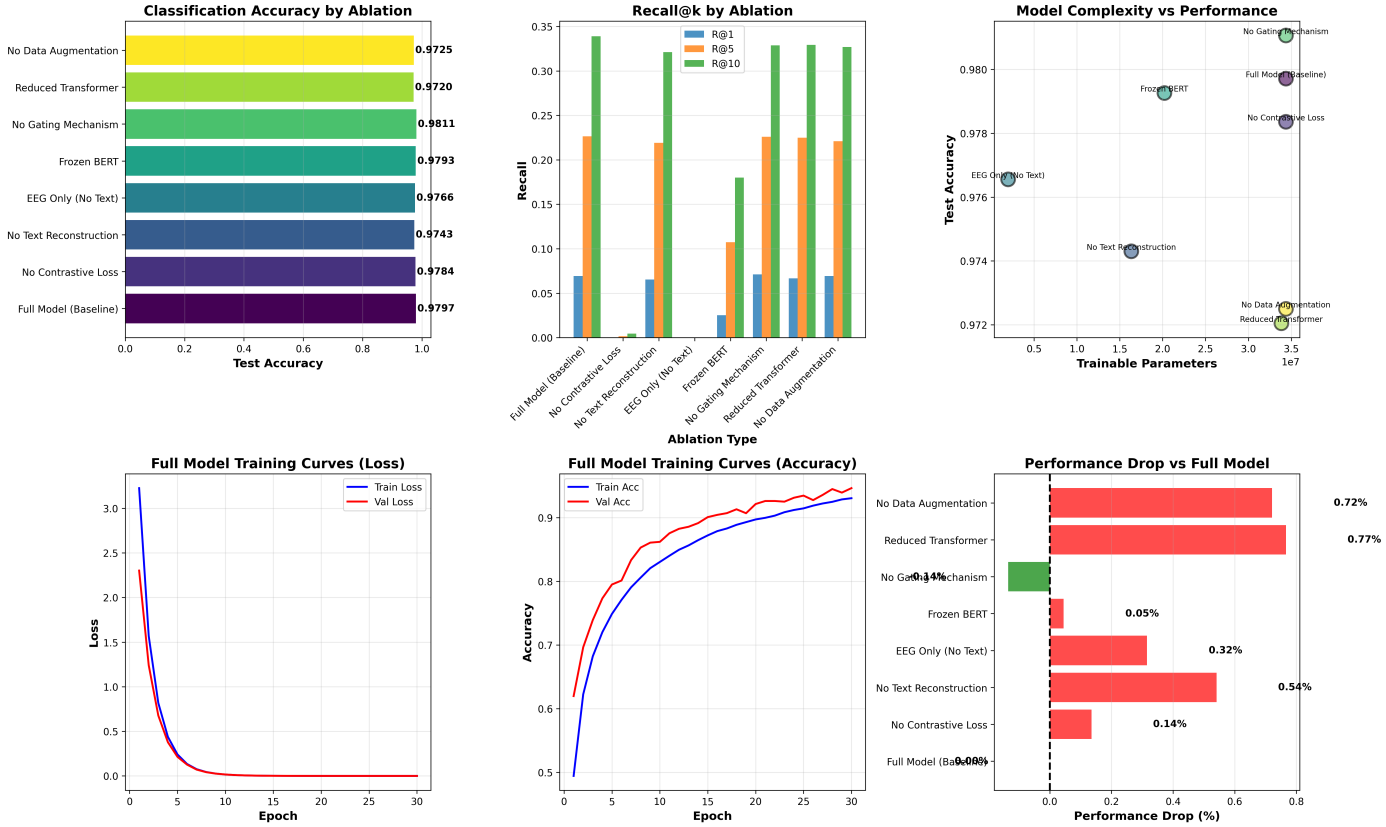


图 5：综合消融分析，展示了 (a) 测试准确率、(b) 脑电图–文本检索性能、(c) 准确率与参数数量的关系、(d) 完整模型的训练损失和验证损失、(e) 完整模型的训练准确率和验证准确率，以及 (f) 相对于基线的性能变化。

4.10. 基线收敛特性

图 7详细展示了基线模型的收敛情况。验证准确率单调上升，到第 30 个训练轮次时达到约 0.946；与此同时，验证损失降至 $2_{\cdot}3\times10^{-6}$，表明优化条件良好。

4.11. 综合解读

在所有分析中，分类准确率、语义对齐和优化稳定性受部分相互独立的因素影响。编码器深度和数据增强对准确率影响最大，对比学习决定语义对齐，而辅助目标则能稳定优化。模型的准确率相近，但表征质量可能存在显著差异，这凸显了面向临床的脑电图建模需要进行多维评估。

5. 讨论

本节结合既往脑电建模研究解读实验结果，并阐述这些结果对神经重症监护中表征学习的意义。重点在于探讨结果揭示的架构设计、监督策略和评估实践方面的启示，而非重复前文讨论过的数值结果或临床动机。

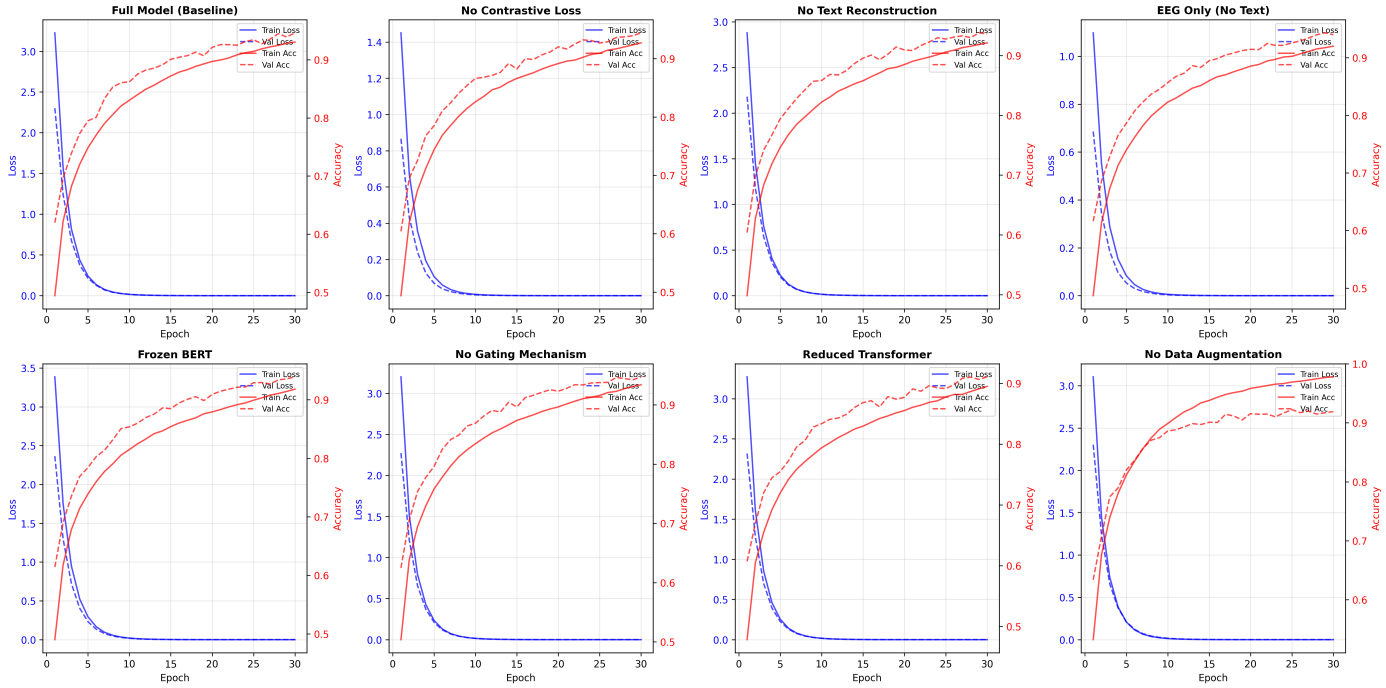


图6：所有消融变体的训练曲线和验证曲线。缺少多模态监督的模型早期收敛更快，但验证方差更大。

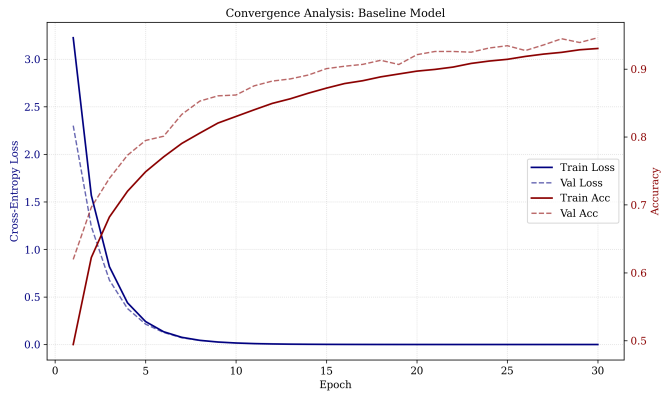


图 7：完整模型的详细收敛行为，显示出稳定的训练和验证动态。

超越以癫痫发作为中心的脑电建模。现有的大多数脑电图深度学习研究主要针对癫痫发作检测，非癫痫发作异常则往往被排除在外，或被归并为粗略的异常类别 [19, 20, 25, 26]。尽管这一研究重点推动了癫痫发作基准的发展，但也限制了我们对模型如何表征其他临床上重要的脑电图模式的认识，而这些模式在神经重症监护中十分常见，包括周期性放电和节律性δ活动 [5, 6, 7]。

通过在统一框架中明确建模美国临床神经生理学会定义的六种有害脑电图模式，本研究得以直接比较现代架构如何编码不同的病理动态。这一更广泛的研究范围回应了临床脑电文献和近期机器学习综述中指出的不足：对非癫痫发作模式的覆盖有限，以及先前研究缺乏表征层面分析 [15, 21, 24]。

长上下文建模的作用。与缩减上下文的变体相比，基于Transformer的编码器表现出更强的鲁棒性，这与先前的观察一致：自注意力机制更善于捕捉脑电信号中的长程时间依赖[23, 42]。这一表现符合临床认知，即节律性和周期性脑电模式的定义特征是持续性和时间规律性，而非孤立事件 [5, 9]。

然而，结果也表明，仅凭架构容量并不能生成具有临床结构的表征。仅接受离散分类监督训练的 Transformer编码器虽然准确率很高，却缺乏有意义的语义组织。这一发现与先前关于脑电图 Transformer 的研究相呼应：这些研究报告了准确率提升，却未考察表征结构，凸显出监督信号应体现临床语义，而非仅反映标签可分性的必要性。

显式脑电–文本对齐的影响。引入显式的脑电–文本对比目标，会显著改变表征几何和跨模态检索性能，但对分类准确率的影响很小。移除这一目标后，即使判别性能得以保留，对齐行为也几乎完全崩溃。

这种解耦与视觉-语言及医学图像-文本模型中的证据相呼应：对比对齐对于在学习到的嵌入中构建语义结构是必要的[28, 29, 30, 32]。在脑电图领域，既有的信号–语言研究在规模和评估严谨性方面仍然有限 [51, 52]。本研究结果提供了定量证据，表明多模态监督会以以准确率为中心的评估无法捕捉的方式改变脑电图表征。

文本重建作为一种表征约束。脑电图条件下的文本重建对分类性能的提升有限，但能持续改善表征紧凑性和优化稳定性。这种表现表明，重建主要是对表征完整性的约束，而非独立的任务目标。

在多模态医疗人工智能中也观察到了类似模式，其中生成目标用于探究潜在表征是否能保留类别身份之外的临床相关信息 [62, 34]。在这种情况下，成功重建表明，脑电图嵌入保留了足够的描述性信息，可用于生成结构化临床摘要；这补充了对比对齐，而不会重复其效果。

自适应融合与表征灵活性。与静态融合相比，自适应融合使对齐行为更加稳定，也降低了对架构扰动的敏感性，尽管分类准确率的峰值差异不大。这种权衡与自适应与基于混合的架构相关的既有研究结果一致：这类架构通常更注重在异质输入下保持稳健，而非追求边际准确率提升 [63, 48]。

鉴于有害脑电模式呈现出不同的时间和频谱特征，模式依赖融合提供了一种合理的机制来体现这种异质性。相比之下，大多数现有的多视图脑电图模型依赖静态融合策略 [49, 50], ，这限制了它们根据不同模式类型调整表征侧重的能力。

对脑电图表征学习的启示。 综合来看，结果表明，分类准确率、语义对齐和优化稳定性受部分相互独立的因素影响。编码器深度和数据增强主导判别性能，对比对齐决定语义组织，而辅助目标会影响训练动态。因此，准确率相近的模型在表征质量和可解释性方面可能存在显著差异。

这些发现支持在面向临床的脑电图建模中采用多维评估方案，并提示不应仅依据准确率进行评估。将脑电图表征与结构化临床描述进行显式对齐，可提供一种实用机制，以探究并塑造表征质量，使其更能反映临床判读。

局限性与未来方向。 仍存在若干局限。所有实验均在单一数据集上开展，且采集特征固定；研究未评估跨机构泛化 [54]。采用结构化的基于模板的文本描述限制了语言变异性，无法涵盖自由形式脑电图报告的全部范围。此外，本研究未明确探讨不确定性校准和概率置信度建模。

未来的工作应探讨跨数据集迁移、不确定性感知目标以及纵向脑电图建模。多模态基础模型的最新进展为扩展脑电图—语言对齐提供了潜在途径，同时能够保持临床依据 [64, 34]。

6. 结论

本研究探讨了如何对具有临床危害的脑电活动进行建模，使其既能体现判别性能，也能反映神经重症监护实践中采用的描述性结构。本研究提出的框架并未将脑电图解读视为纯粹的信号级分类问题，而是将信号域建模与结构化临床语言监督相结合，以探究并约束表征学习。

实证结果表明，仅有较高的分类准确率并不能保证表征具有临床意义。准确率几乎相同的架构，在语义对齐、优化稳定性和表征结构方面仍存在显著差异。显式的脑电图–文本对比监督在塑造嵌入几何方面发挥着核心作用，而辅助重建目标和自适应融合机制则会影响稳定性与表征灵活性，但不会主导准确率表现。

通过将评估范围扩展到以准确率为中心的指标之外，本研究强调了面向临床的脑电图建模中表征层面分析的重要性。研究结果表明，即使语言监督具有结构化且受约束的特点，多模态监督仍能作为一种实用机制，用于评估脑电图表征并引导其形成具有临床可解释性的结构。

尽管当前评估仅限于单一数据集和基于模板的临床描述，但该框架为研究不同监督范式下的脑电表征质量奠定了基础；这些范式更能反映脑电检查结果在实践中被解读和传达的方式。本研究为持续推动脑电图机器学习的发展作出了贡献，助力构建不仅准确，而且符合临床推理和报告规范的模型。

参考文献

[1] J. Claassen, S. A. Mayer, R. G. Kowalski, R. G. Emerson, L. J. Hirsch，危重患者持续脑电图监测中脑电图发作的检测，《神经病学》62 (10) (2004) 1743–1748. [2] J. Claassen, F. S. Taccone, P. Horn, M. Holtkamp, N. Stocchetti, M. Oddo，危重患者脑电图监测应用建议，《重症医学》39 (2013) 1337–1351. [3] M. B. Westover, M. M. Shafi, M. T. Bianchi, L. M. Moura, D. O'Rourke, E. S. Rosenthal，危重成人脑电图监测期间发生癫痫发作的概率，Clinical Neurophysiology 126 (3) (2015) 463–471. [4] A. Rodrıguez Ruiz, J. Vlachy, J. Y. Lee, E. J. Gilmore, H. A. Haider, N. Gaspard，周期性和节律性脑电图模式与危重患者癫痫发作的关联，《JAMA神经病学》74 (2) (2017) 181–188. [5] B. Foreman, J. Claassen，定量脑电图在脑缺血检测中的应用，Journal of Clinical Neurophysiology 29 (2) (2012) 98–103. [6] N. Gaspard, L. J. Hirsch, S. M. LaRoche, C. D. Hahn, M. B. Westover，重症脑电术语的评估者间一致性，Epilepsia 55 (9) (2014) 1366–1373. [7] R. Saab, J. J. Millichap，重症监护病房中的脑电图监测：适应证与结局，Journal of Clinical Neurophysiology 37 (1) (2020) 20–28. [8] N. S. Abend, K. E. Chapman, T. G. Zelleke，重症监护病房中的脑电图，《柳叶刀·神经病学》12 (12) (2013) 1214–1224. [9] A. F. Struck, G. Osman, N. Rampal, S. Biswal，持续脑电图监测下危重患者癫痫发作的时间依赖性风险，《神经病学年鉴》82 (2) (2017) 177–185. [10] G. B. Young，重症监护病房中的持续脑电图监测，Journal of Clinical Neurophysiology 35 (3) (2018) 195–201.

[11] E. T. Payne, A. T. Berg，危重疾病中的脑电图监测与癫痫发作负担，Epilepsy Currents 19 (4) (2019) 233–238. [12] E. T. Payne, N. Zhao, H. Frndova, K. McBain, R. Sharma，癫痫发作负担与危重患儿短期结局独立相关，Brain 137 (5) (2014) 1429–1438. [13] C. R. Newey, A. Sarwal, S. Hantus，重症监护病房中的脑电图，Seminars in Neurology 40 (1) (2020) 89–100. [14] J. D. Pandian, G. D. Cascino，重症监护病房中的数字视频脑电图监测，Journal of Clinical Neurophysiology 21 (5) (2004) 341–352. [15] L. J. Hirsch, S. M. LaRoche, N. Gaspard，美国临床神经生理学会标准化重症脑电图术语，Journal of Clinical Neurophysiology 30 (1) (2013) 1–27. [16] P. A. Gerber, K. E. Chapman, S. S. Chung，危重患者脑电图模式解读的观察者间一致性，Epilepsia 49 (6) (2008) 965–971. [17] J. J. Halford，脑电图解读的评定者间一致性，Journal of Clinical Neurophysiology 32 (6) (2015) 512–517. [18] S. Beniczky, H. Aurlien, J. C. Brøgger，脑电图的标准化计算机辅助结构化报告，Clinical Neurophysiology 131 (3) (2020) 698–705.[19] R. T. Schirrmeister, J. T. Springenberg, L. Fiederer，使用卷积神经网络进行深度学习以实现脑电图解码，Human Brain Mapping 38 (11) (2017) 5391–5420. [20] V. J. Lawhern, A. J. Solon, N. R. Waytowich，EEGNet：一种用于基于脑电图的脑机接口的紧凑型卷积神经网络，Journal of Neural Engineering 15 (5) (2018) 056013. [21] Y. Roy, H. Banville, I. Albuquerque，基于深度学习的脑电图分析：系统综述，J Neural Eng 30 (9) (2019) 2531–2553.[22] A. Craik, Y. He, J. L. Contreras-Vidal，脑电图分类任务中的深度学习，Journal of Neural Engineering 16 (3) (2019) 031001. [23] Y. Song, Q. Zheng, F. Liu，基于Transformer的脑电图情绪识别，NeuroImage 245 (2022) 118699. [24] A. Shoeibi, N. Ghassemi, R. Alizadehsani, M. Rouhani, H. Hosseini-Nejad, A. Khosravi, S. Nahavandi，深度学习技术在利用脑电信号自动诊断多发性硬化中的应用，Neural Computing and Applications 33 (2021) 15259–15284. [25] M. C. Tjepkema-Cloostermans, J. Hofmeijer，缺氧后昏迷中的脑电图异常，Clinical Neurophysiology 131 (2020) 202–209. [26] S. Mousavi, F. Afghah，用于脑电图异常检测的深度学习，IEEE Transactions on Biomedical Engineering 68 (1) (2021) 20–30. [27] W. O. Tatum，脑电图报告：哪些内容至关重要，Journal of Clinical Neurophysiology 33 (4) (2016) 294–302. [28] A. Radford, J. W. Kim, C. Hallacy, et al.，通过自然语言监督学习具有迁移能力的视觉模型，载于：国际机器学习会议论文集，2021年. [29] C. Jia, Y. Yang, Y. Xia, et al.，扩展视觉及视觉语言表征学习，载于：IEEE/CVF计算机视觉与模式识别会议论文集，2021年. [30] J. Irvin, P. Rajpurkar, M. Ko, et al.，CheXpert：大型胸部X线影像数据集，载于：AAAI人工智能会议，2019年. [31] B. Boecking, N. Usuyama, et al.，提升医学影像模型对混杂因素的鲁棒性，Medical Image Analysis (2022). [32] S.-C. Huang, L. Shen, et al.，多模态医学基础模型，Nature (2023). [33] E. Alsentzer, J. R. Murphy, et al.，公开可用的临床BERT嵌入，载于：ACL论文集，2019年. [34] M. Moor, O. Banerjee, et al.，通用型医学人工智能基础模型，Nature (2023). [35] W. Hong, L. Wu, et al.，面向医疗保健的多模态基础模型，Nature (2023). [36] J. Gotman，脑电图中癫痫发作的自动识别，Electroencephalography and Clinical Neurophysiology 54 (5) (1982) 530–540. [37] S. B. Wilson, M. L. Scheuer, C. Plummer, B. Young, S. Pacia，癫痫发作检测：REVEAL算法评估，Clinical Neurophysiology 115 (10) (2004) 2280–2291. [38] M. A. Navakatikyan, P. B. Colditz, C. J. Burke, T. E. Inder，基于波序列分析的新生儿癫痫发作检测算法，IEEE Transactions on Biomedical Engineering 53 (3) (2006) 511–520. [39] A. Antoniades, L. Spyrou, C. C. Took, S. Sanei，癫痫发作的深度学习脑电图中的癫痫发作检测，Clinical Neurophysiology 128 (12) (2017) 2287– 2295.[40] U. R. Acharya, S. L. Oh, Y. Hagiwara, J. H. Tan, H. Adeli，利用脑电信号自动检测和诊断癫痫发作的深度卷积神经网络，生物学与医学中的计算机 100 (2018年) 270–278.[41] N. R. Waytowich, V. J. Lawhern, J. O. Garcia，异步稳态视觉诱发电位分类的紧凑型卷积神经网络，Journal of Neural Engineering 15 (6) (2018年) 066031. [42] Y. Yuan, Z. Xiong, D. Wu，基于Transformer的脑电图分类注意力网络，IEEE Transactions on Biomedical Engineering 70 (5) (2023) 1456–1466.[43] Z. Chen, X. Li, Z. Zhang，使用Transformer网络对脑电图进行长程时间建模，IEEE生物医学与健康信息学杂志 27 (3) (2023) 1204–1215. [44] A. V. Oppenheim, R. W. Schafer，《离散时间信号处理》，第2版，Prentice Hall，1999.[45] C. Brunner, A. Delorme, S. Makeig，EEGLAB——用于电生理学研究的开源MATLAB工具箱，神经科学方法杂志 234 (2014) 1–17. [46] P. Bashivan, I. Rish, M. Yeasin, N. Codella，利用深度循环卷积神经网络从脑电图中学习表征，载于：国际学习表征会议，2016年。[47] B. Hu, Z. Zhang, L. Zhang，用于脑电图表征学习的多尺度深度神经网络，NeuroImage 225 (2021) 117438. [48] W. Zhou, Z. Zhang, J. Luo, W. Chen，用于脑电图表征学习的时空频谱注意力，NeuroImage 237 (2021) 118190. [49] X. Wang, X. Li, Z. Zhang，用于脑电图脑状态分类的多分支卷积神经网络，IEEE Transactions on Biomedical Engineering 69 (3) (2022) 1077–1088. [50] Y. Ding, N. Robinson, S. Li, C. Guan，TSception：捕捉时间与空间动态以实现脑电图解码，IEEE神经网络与学习系统汇刊 34 (6) (2023) 2875–2888. [51] H. Zhang, S. Chen, K. Han，在语言监督下学习脑电图表征，载于：MICCAI论文集，2022年。[52] Y. Liu, H. Wang, Y. Zhang，用于脑状态解码的脑电图-文本多模态学习，传感器 23 (4) (2023) 1987. [53] C. J. Kelly, A. Karthikesalingam, M. Suleyman, G. Corrado, D. King，实现人工智能临床影响的关键挑战，BMC医学 17 (2019年) 195. [54] M. Roberts, D. Driggs 等，医学影像机器学习中的常见陷阱，自然·机器智能 (2021年)。[55] W. O. Tatum，《脑电图解读手册》，第2版，Demos Medical Publishing，纽约，NY，2014年。[56] A. Vaswani, N. Shazeer, N. Parmar 等，注意力机制就是你所需要的一切，载于：神经信息处理系统进展，2017年。[57] A. Kendall, Y. Gal, R. Cipolla，利用不确定性为场景几何和语义的损失加权的多任务学习，2018年。[58] J. Jing, Z. Lin, C. Yang, A. Chow, S. Dane, J. Sun, M. B. Westover, HMS——有害脑活动分类，Kaggle，访问日期：2026-01-02 (2024). [59] P. Micikevicius, S. Narang 等，混合精度训练，载于：国际学习表征会议，2018年。[60] I. Loshchilov, F. Hutter，解耦权重衰减正则化，载于：国际学习表征会议，2019年。[61] J. Pineau, P. Vincent-Lamarre 等，提升机器学习研究的可复现性，机器学习研究杂志 22 (2021年) 1–20. [62] S.-C. Huang, A. Pareek 等，利用深度学习融合医学影像与电子健康记录，自然·医学 29 (2023) 14–28. [63] N. Shazeer, A. Mirhoseini 等，规模惊人的神经网络：稀疏门控专家混合层，arXiv预印本 arXiv:1701.06538 (2017). [64] R. Bommasani, D. A. Hudson 等，基础模型的机遇与风险，arXiv预印本 arXiv:2108.07258 (2021).