# 1.多模态融合方法

人工智能发展的灵感是模仿人类的多个感官从外界获取信息，例如视觉、听觉、嗅觉、触觉。由于人类生活在复杂信息相互交融的环境中，面临着大量、多样化的数据，获取信息更多以多维度的信息融合的方式，例如在自动驾驶领域，视觉传感器可以获取路况画面，激光雷达可以检测车辆前方物体的距离，声音传感器可以根据汽笛声做到听声定位，来弥补视觉盲区的道路信息。这些多模态数据蕴含着丰富的信息，对人们理解和处理现实世界的问题具有重要意义。随着文本图像的生成[1.1]、图文检索[1.2]、视觉问答[1.3]、机器人[1.4]和智能医疗[1.5]等领域发展，为了使机器能从多模态的信息中综合地提取信息，需要赋予机器理解、推理和学习的能力，因此需要将不同模态的信息进行融合，以获得更全面准确的分析和决策能力。

多模态数据是相同语义信息在不同的空间维度下的记录。而不同维度的数据具有不同的性质、结构和表征，这种异构性可能会影响模型的学习和应用。多模态信息融合的目标是减少这些异质性的差距，使用不同模态信息共同完成相同的任务。例如，对于同一信息的表达，既可以使用图像，也可以使用声音或文字。当某一模态数据丢失时，其他模态可以弥补这一缺失。这些数据之间的相互作用能够产生更完整的信息。它们以不同的方式相互作用产生更完整的信息。深度学习的发展是以数据为基础，多模态所蕴含的信息多于单一模态，优于单模态的学习为该领域的发展带来了全新的挑战[1.6]，例如在自动驾驶[1.7]、情感识别[1.8]和医疗诊断等领域。任何一个任务场景所传递的信息都不是单一的，而是以多种方式进行传递，不同的方式就会蕴含不完全一样的信息，利用好多种模态信息，挖掘更全面的信息，是多模态融合技术的研究目标，多模态融合适用于人工智能赋能的任何领域，是非常具有前景的领域。多模态融合方法有联合融合方法、协同融合方法、编码器融合方法以及分裂融合方法四种，多模态信息对齐方法有显式对齐和隐式对齐两种

对于本模型使用的是协同融合方法和显示对齐方法，相较于联合融合方法在输入阶段就进行融合，直接将多个模态的数据拼接为一个输入。由于两种模态之间的信息密度、数据格式和特征空间差异，简单拼接可能导致信息不充分利用或失真，未能充分考虑模态间的复杂交互和相互依赖；而协同融合方法

# 2．抑郁症

# 抑郁症,也称为抑郁障碍,是一种常见情绪障碍,其主要表现为思维迟滞、意志减退及社会功能受损,严重者可致自伤行为和自杀意念[2.1]。 世界卫生组织2021年报道,抑郁症是导致全球精神和身体残疾的主要原因,据估计,全球共有超过3.5亿人患病,到 2030 年为止,抑郁症可能成为第一致残疾病[2.2]。 有研究发现, 抑郁症终身患病率可达 18. 1%[2.3],国内抑郁症患病率为 6. 9%[2.4]。 然而,在精神卫生领域,对抑郁症的诊断主要依靠问诊和量表筛查,极大依赖于医师的临床经验和患者的配合程度,具有很高的误诊率和漏诊率[2.5]。 相关数据显示,全科医师对抑郁症的诊断识别率达 47. 3%[2.6]。因此,寻找客观有效的诊断方法对早发现、早治疗具有十分紧迫的现实意义。

抑郁症是全世界主要致残原因，也是造成全球疾病总负担的主要因素之一[3.1]，其主要特点有持续的悲伤、失去兴趣或快乐等。抑郁症会给患者身心带来极大的影响且容易复发，当抑郁症发展到中度或重度时，将成为一个严重的健康疾患甚至导致患者自杀[3.2]，从而对家庭和社会造成严重的损失。

目前，抑郁症的诊断主要基于国际疾病分类标准（international classification of diseases，ICD）和精神疾病诊断统计手册（diagnostic and statistical manual of mental disorders，DSM），这种诊断方式易受患者的主观信念和医师的诊断经验影响[3.3]。此外，现代社会对于抑郁症等精神类疾病的接受程度普遍较低，抑郁症患者自身也常由于存在病耻感和对精神疾病缺乏了解等原因而没有选择求医，结果造成众多患者因未得到及时、准确的诊断而错失最佳治疗机会[1]。

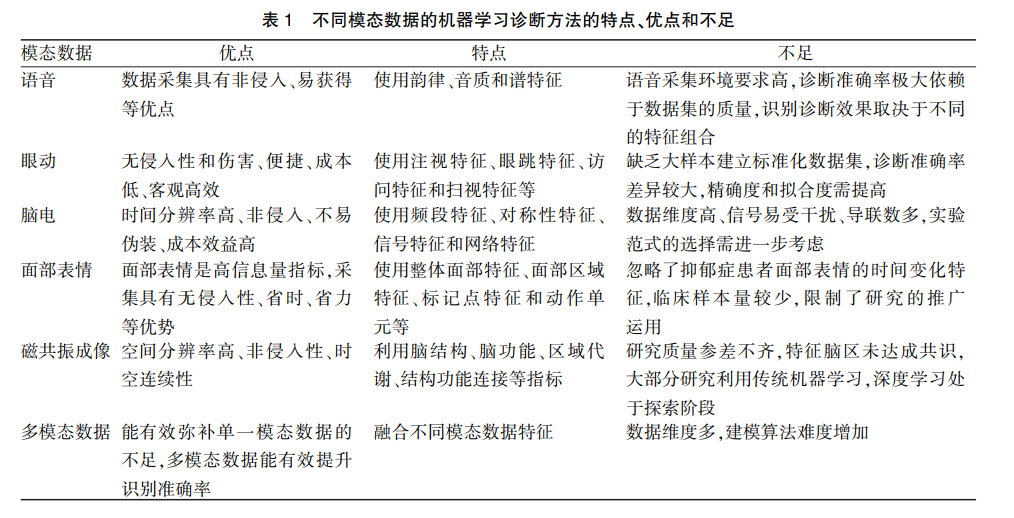
随着互联网的兴起，人们越来越倾向于在社交媒体上分享自己的生活状态，且患有精神疾病的人也以倾诉自己的精神状态作为一种解脱[3.4]。

# 3.抑郁症的机器学习研究

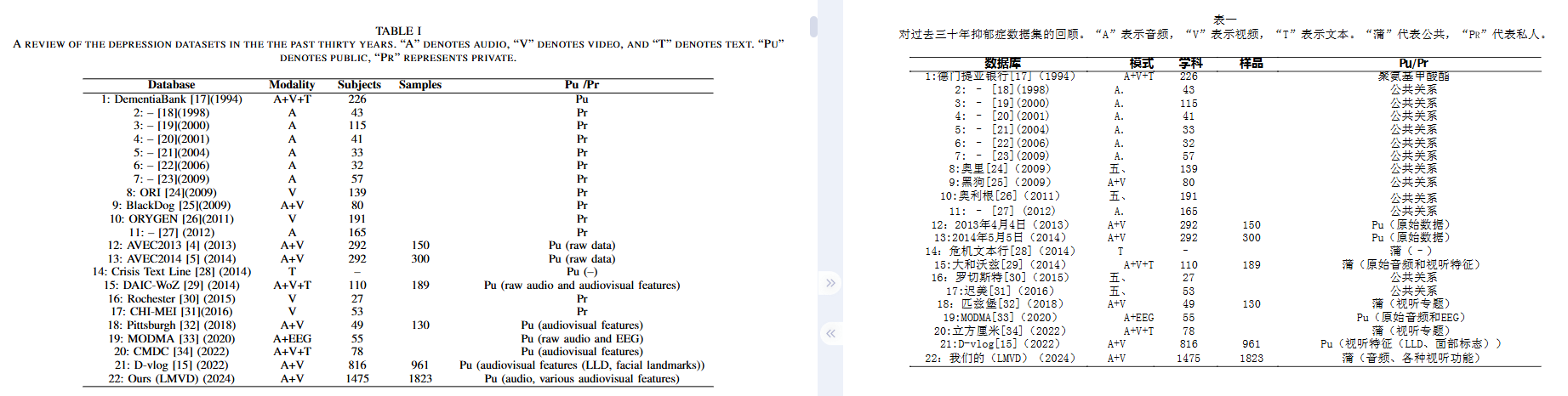
近年来,随着人工智能的日趋成熟,其应用已逐渐渗透至精神医学领域,对抑郁群体的精神疾病的早期预测、识别,以及调整最优治疗方案等方面具有出色的表现,在抑郁症的辅助诊断中发挥着不可或缺的作用。 其中,机器学习(machine learning,ML)作为人工智能的核心技术,其优势在于可以从海量的数据中,学习事物背后的规律,并对未来数据进行有效预测。 机器学习是一门研究计算机如何从数据中学习的科学学科,指计算机对已有的数据进行自主学习,以便于在下一次执行相同的任务中表现出更高效率的技术。 按照模型结构的深度可以将机器学习分为传统机器学习和深度学习[2.7]。 机器学习因具有很强的数据处理能力,可用于建立临床预测模型,辅助临床医师诊断和判断疾病预后情况,有助于精准医疗的实施。具有很强的数据处理能力,可用于建立临床预测模型,辅助临床医师诊断和判断疾病预后情况,有助于精准医疗的实施。

将不同模态数据的特点、优点和不足进行总结：基于语音数据、基于眼动数据、基于基于面部表情。

基于多模态的：由于抑郁症的临床表现和影响因素相当复杂,使用单一模态的数据极有可能忽略个体差异对结果的影响。 近年来,有研究者使用机器学习融合多模态数据进行建模,取得良好的预测识别效果。 有学者提取语音、眼球活动及头部姿势三个模态数据的特征,采用支持向量机进行分类,结果发现多模态特征融合后的识别准确率显著高于单个模态[2.32]。 侯峰等[2.33]采用情绪脸-词 Stroop 任务,收集 31 例重度抑郁症患者和 31 名健康对患者的行为数据及 ERP 数据,提出行为脑电并行时序卷积神经网络的学习方法,结果发现,与单模态数据结果相比,多模态在准确性和 F1 分值等指标上具有明显的提升。 此外, 还有研究基于量表[2.34]、 文本[2.35]、炎症标志物[2.36]、脉搏和呼吸信号[2.37]、用户画像[2.38]、蛋白组学[2.39]和步态信息[2.40]等模态数据建立机器学习模型用于抑郁症分类研究,未来可融合不同模态信息,提升模型诊断准确率。



# 4.抑郁症机器学习数据集



# 5.抑郁症+多模态+transformer

# 赵小明等针对传统的多模态融合方法在抑郁症检测中忽略了模态之间的交互性、未能充分提取出更全面的特征表示的问题,本研究提出一种基于多模态特征增强网络的抑郁症检测方法,该方法有效地集成了视频、音频和远程光电容积脉搏(photoplethysmographic,rPPG)信号3种模态,通过模态间Transformer、模态内Transformer和多头自注意力机制,共同学习输入模态序列每个时间步的模态内和模态间的动态关系,达到了特征增强的目的。最终, 拼接3个模态增强后的特征获得全面特征表示。在AVEC2013公共数据集上的实验结果显示,该方法的平均绝对误差为7.07,优于单模态抑郁症检测,表明该方法有效促进了模态之间的交互,并实现了特征增强,在自动抑郁症检测任务中展现出显著的有效性。[4]

[4]的引言作为绪论，[4]的相关研究作为文献综述，参考[4]的内容写内容