УДК 004.891.3

将视频多模态情感分析运用在临床抑郁检测中

黎博毅 (Li BoYi), 何润海 (He RunHai), 章恒睿 (Zhang HengRui), Natalia·Khajynova
白俄罗斯国立信息与无线电大学 (BSUIR)
e-mail: liboyer854@gmail.com, khajynova@bsuir.by

Summary. *The current clinical diagnosis of depression in the medical community relies on self-rating scales and physician interviews, but this approach is limited by the expertise of clinicians and the uneven distribution of medical resources. This paper proposes the use of video multimodal techniques in clinical diagnosis, aiming to improve the efficiency and accuracy of depression detection in clinical settings.*

根据世界卫生组织的数据，全球有 2.8 亿人患有抑郁症。仅近三年，全球新增抑郁症患者超过 7000 万人。抑郁症，被称为“21 世纪最大的杀手”。抑郁症是一种严重的心理疾病，不仅会对个人的心理以及身体产生极大的危害，而且也会给家庭、社会带来不利的影响。〞早发现、早治疗 〞被认为是这种疾病的最佳治疗方案，这表明需要对抑郁症进行早期筛查。传统的抑郁症诊断依赖于自我评估量表和医生访谈，但这种方法受限于临床医生的专业知识和医疗资源的不均衡分布。人工智能的快速发展为抑郁症的识别提供了一个新的解决方案，有望弥补上述不足。

抑郁症患者语言上常表现为消极、厌世, 表情常表现为皱眉和更少的微笑, 声音常表现为语速较慢、停顿较多，利用人工智能可以很好的捕捉到这些特征。所以，通过情感分析辅助识别抑郁症是一种趋势，已经有一些研究通过分析社交文本、语音信号或面部图像来检测抑郁症，然而，由于抑郁症的表现形式多样，基于单一特征的抑郁症识别并不能获得足够的信息，导致识别不准确，故论文提出在临床诊断中使用视频多模态情感分析技术来提升抑郁症识别的准确率以及效率。

视频多模态情感分析（Video multimodal sentiment analysis）是指首先将视频中包含的视觉、听觉、文本等多模态信息提取出来，在采取特征融合后进行情感分析，并且综合分析结果，从而得出更有效的结论。在临床抑郁症的诊断中，可以在经过被诊断者允许的情况下，收集其日常生活的视频，因为在视频中可以很好的呈现出被诊断者在日常交流时体现出的表情变化，语气转变以及用词的倾向。将视频分为三种类型的数据集，分别为：视频中提取出的文本信息、视频中的语音信息、包含表情变化的关键帧，这些数据集可在各自对应的模型中进行处理，最后通过 Attention 机制进行模态融合。

Attention 机制，简单来说，人类在观察外界事物时，通常不会把它作为一个整体来看待，而是倾向于根据自己的需要有选择地获取被观察事物的一些重要部分。同样，在深度学习中，Attention 机制可以帮助模型对输入的每一部分给予不同的权重，提取更多的关键和重要的信息，从而使模型能够做出更准确的判断，而不会给模型的计算和存储带来更多的开销。Attention 机制有效实现了多模态信息的互补性和多模态贡献度计算, 保证了多模态信息融合的合理性和准确度。

人类情感的表达体现在声音、表情、肢体动作等多种模态中，而且是一个发展变化的过程。多模态情感识别涉及对多个模态信号进行处理、各个模态情感特征的学习、多模态特征之间的融合、多模态之间的交互建模等，因此，多模态情感分析技术与传统单模态情感分析相比有着在模态融合上的优势。并且将人工智能应用到临床医学诊断中可以实现不受空间、时间、资源的约束，例如将视频多模态情感分析技术运用到抑郁症诊断中，医生只需要被诊断者提供日常生活的视频，视频在经过相应的处理，提取出对应信息，即可由人工智能模型快速得出诊断结果。

随着人工智能的不断发展，未来人工智能在抑郁症诊断领域将的运用会越来越频繁，抑郁症的诊断也会越来越高效，使得抑郁症在早期被发现的概率也越高。