心理科学进展 2026, Vol. 34, No. 3, 424–440 https://doi.org/10.3724/SP.J.1042.2026.0424
Advances in Psychological Science https://cstr.cn/32111.14.2026.0424

·研究方法(Research Method)·

大模型在抑郁症筛查与诊断中的应用*

谢 宇[1] 郑弘欣[1] 刘怡资[1] 禹红刚[2] 杨成赫[2]

([1]安徽师范大学教育科学学院, 芜湖 241000) ([2]中国电信股份有限公司安徽分公司, 合肥 230001)

摘 要 抑郁症是一种常见的精神障碍, 严重影响患者的社会功能和生活质量。近年来, 大模型凭借其强大的语义理解和多模态数据处理能力, 在抑郁症早期筛查与辅助诊断中展现出显著优势。构建抑郁症筛查和诊断大模型通常包括: 数据准备、模型选择、模型训练和模型评估四个步骤。大模型在抑郁症筛查与诊断中, 主要通过情境化语义表征、注意力机制、多模态行为捕捉及生成式预测等关键技术实现。但当前研究仍存在算法偏见、诊断特异性、幻觉现象、隐私安全及伦理问题等挑战。未来应加强大模型心理干预的整合应用, 聚焦临床转化路径, 构建更为精细、动态且具备文化适应性的抑郁症数字表型, 实现心理健康服务的数智化转型。

关键词 大模型, 抑郁症, 早期筛查, 辅助诊断

分类号 R395

1 引言

抑郁症是一种常见的精神障碍, 主要表现为持续的情绪低落、兴趣或愉悦感减退, 常伴认知、睡眠、食欲等功能性失调, 严重影响患者的社会功能和生活质量, 且并发自杀风险的增加(American Psychiatric Association, 2013)。据世界卫生组织(World Health Organization, WHO)估计, 全球有超过 3 亿的人群罹患抑郁症, 占世界总人口的 4.4% (World Health Organization, 2017)。传统抑郁症筛查和诊断手段以标准化量表和结构化临床访谈为主, 存在主观性强、效率低、早期识别困难等不足(Insel & Cuthbert, 2015)。导致大量的抑郁症患者无法得到及时的诊断和治疗, 加重了患者的疾病负担。因此, 建立客观、高效、准确的抑郁症筛查和诊断手段, 已经成为心理学和精神医学领域亟待解决的重要科学问题。

近年来, 人工智能(Artificial Intelligence, AI)尤其是机器学习方法在心理学领域的应用取得了很大的进展。早期研究利用机器学习方法对文本、语音、行为等数据进行分析, 在抑郁症的识别、诊断和风险评估等方面展现出良好的应用潜力(董健宇 等, 2020)。随着数据规模的指数级增长和模型复杂度的提升, 基于 Transformer 架构的大模型正在推动研究范式革新(Bommasani et al., 2021)。大模型往往具有数十亿甚至数百亿的参数规模, 在海量、多模态数据上进行预训练, 具备强大的上下文语义理解、复杂逻辑推理和高质量内容生成能力, 突破了传统机器学习在处理复杂、高维、多模态数据方面的局限。

大模型凭借其在处理复杂问题的能力, 实现了在通用场景领域的突破, 其应用迅速扩展至医疗场景(陈晓红 等, 2024)。与在其他医疗领域的广泛应用不同, 大模型在心理健康领域面临着独特的挑战, 特别是应对心理的复杂性与满足个性化干预的需求(Omar et al., 2024)。如表 1 所示, 大模型较传统量表和机器学习方法, 在评估效度、效率、可及性、个性化方面展现出巨大潜力。虽然探索适用于各类常见心理问题的通用型大模型是目前临床工作者和人工智能开发者的共同愿景,

收稿日期: 2025-06-28

* 安徽省高等学校思想政治教育研究会 2024 年度高校思想政治教育研究专项课题(2024SZX012); 中国电信股份有限公司大中小一体化智能心育研发项目(24AHEKYF5020)。

通信作者: 谢宇, E-mail: xiey@ahnu.edu.cn

表 1　大模型、传统量表和机器学习的比较

维度	大模型	传统量表	机器学习
数据来源	海量的多模态数据	标准化问题的主观报告	结构化或非结构化数据
评估效度	高	高	中
评估效率	极高	低	高
可及性	极高	中	高
评估客观性	高	低	中
可解释性	低	高	中
部署成本	高	低	中
个性化	高	弱	中

但目前研究主要聚焦于抑郁症领域。一方面, 就心理问题的代表性而言, 抑郁症是患病率高、研究最深入的心理障碍之一。另一方面, 从技术的可行性来看, 抑郁症相关的公开数据集相对充足, 能支撑大模型的训练、微调和验证。在抑郁症上的研究突破, 其方法和技术可扩展到其他心理障碍, 最终为心理问题的通用型大模型奠定基础。

目前, 大模型在抑郁症应用领域的探索已初见成效。大模型能够准确地识别患者的情绪特征和情感状态, 不仅可以通过分析社交媒体中的文本进行抑郁症的自动化筛查(Yang, Cao, et al., 2024), 还可以分析在多模态语义信息中蕴含的抑郁症生物标志(Jiang et al., 2024)。但是, 目前抑郁症筛查和诊断大模型的应用研究尚处在早期探索阶段, 其主要步骤、应用场景及风险挑战还需要进一步梳理与评估。因此, 本研究系统梳理大模型在抑郁症筛查与诊断中的研究进展、主要步骤、应用场景、风险挑战及未来研究方向, 以期深化对大模型应用于抑郁症领域的理论认知和实践启示, 为抑郁症筛查和诊断的大模型研究提供参考。

2　方法

本研究采用系统综述的方法, 对 2017 年以来大模型在抑郁症诊断、筛查和预测领域的应用研究进行文献检索。使用的数据库包括中文数据库(知网、万方、维普)和英文数据库(Web of Science、IEEE、EMBASE、PubMed)。中文检索词组合为: “ChatGPT”或“生成式人工智能”或“大语言模型”或“自然语言处理”或“BERT”或“大模型”和“抑郁”或“抑郁症”。英文检索词组合为: “ChatGPT” OR “Generative Artificial Intelligence” OR “Natural Language Processing” OR “Large Language Models” OR “LLMs” OR “BERT” OR “Foundation Model” OR “GAI” AND “Depression” OR “Major Depressive Disorder” OR “MDD”。本研究将文献检索的起始年份设定为 2017 年, 主要考量是 Transformer 架构于该年由 Vaswani 等人正式提出(Vaswani et al., 2017)。该架构的问世具有里程碑意义, 它不仅为大模型领域的研究奠定了基础, 也迅速吸引了学界与业界的广泛关注, 并推动了相关投入的显著增长。因此, 文献检索的时间范围设定为 2017 年 1 月 1 日至 2025 年 4 月 10 日。文献的纳入标准包括: (1)文献类型为实证类期刊论文或会议论文; (2)研究主题聚焦于大模型在抑郁症领域的筛查和诊断; (3)文章包括明确的研究问题、方法、结果和结论。排除标准为: 综述文章、病例报告、无全文的会议摘要、社论和预印本。

文献筛选流程如下: 首先, 将所有检索结果导入文献管理软件后去除重复文献, 由两位研究人员独立对文献的题名和摘要进行初步筛选。之后, 两人对文献进行交叉核对, 若出现分歧, 通过讨论解决, 难以达成一致意见时由第三位研究人员裁决。最后, 两名研究人员阅读初筛文献全文, 并根据纳入与排除标准进行精筛, 最终纳入 55 篇文献。文献筛选过程详见图 1。两名研究人员独立提取文献的关键信息, 包括第一作者、发表年份、国家、样本量、研究对象、数据类型、所应用的基础大模型, 以及性能评估方法等, 主要内容详见网络版附表 1。

为系统评估纳入文献的方法学质量, 本研究采用 QUADAS-2 (Quality Assessment of Diagnostic Accuracy Studies 2)工具(Whiting et al., 2011)。依照该工具, 由两名研究员独立对每篇文

献的四个领域(流程与时间、金标准、被测技术、患者选择)进行偏倚风险评估。评估者根据QUADAS-2的问题，将各领域的偏倚风险归类为“低风险”“高风险”或“不清楚”。文献质量评估结果见图2。

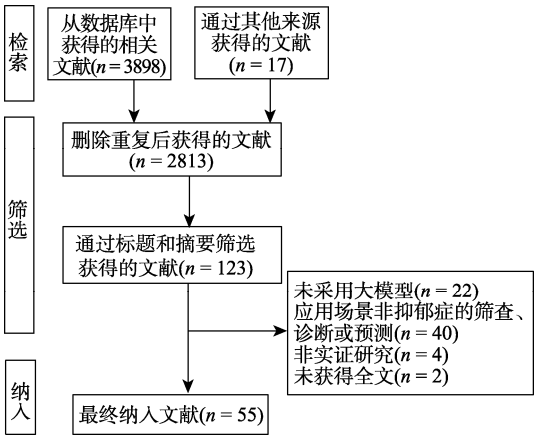

图1 文献筛选流程图

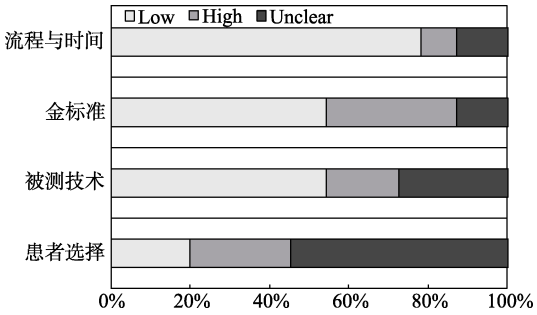

图2 文献质量评估图

3 应用步骤

将大模型应用于抑郁症筛查与诊断，目标在于构建出既具备高精度识别能力，又能在抑郁症诊疗中落地使用的模型。构建抑郁症筛查和诊断大模型通常需要经过4个阶段，依次为数据准备、模型选择、模型训练和模型评估。图3展示了建立抑郁症筛查和诊断大模型技术流程的基本框架和主要环节，表2介绍了大模型的关键技术概念。

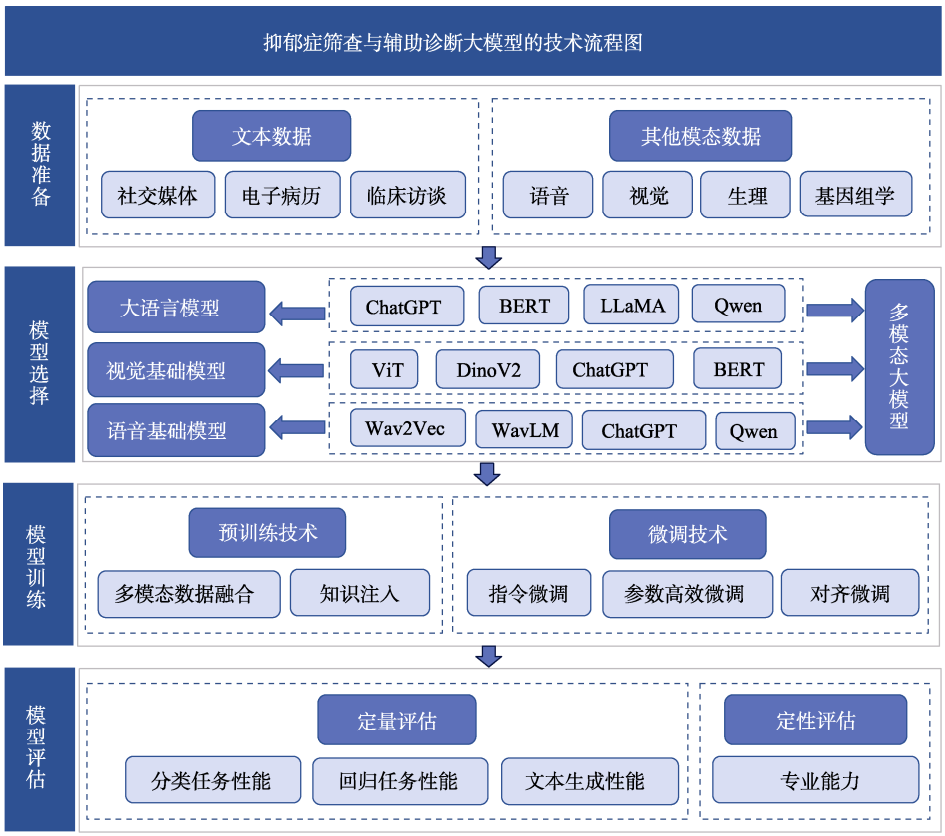

图3 建立抑郁症筛查和诊断大模型技术流程图

表 2　大模型在抑郁症筛查与诊断应用中的关键技术概念解释

技术术语	定义	在抑郁症研究中的应用价值
预训练	在海量语料上, 采用自监督学习范式对模型进行初始训练, 学习通用的语言表征。	为模型赋予了理解人类语言的底层能力, 是识别抑郁症语言模式的基础。
微调	利用特定任务的有标注数据集进行监督学习, 使模型适应特定领域的应用。	将通用大模型特化为抑郁症筛查和诊断领域的专用模型。
指令微调	通过“指令–输出”对数据集进行训练, 使模型学会泛化地遵循人类指令的能力。	提升模型执行复杂临床任务指令的能力。
参数高效微调	仅训练模型的一小部分参数, 以极低的计算和存储成本实现对下游任务的适应。	大幅降低了针对抑郁症定制模型的资源门槛, 使得快速迭代和部署成为可能。
对齐微调	使模型的输出与人类的偏好、价值观和社会规范对齐, 常采用基于人类反馈的强化学习等技术。	保障模型在临床应用中的安全性与伦理性, 避免生成有害、偏颇或不负责任的诊断建议。
适配器	在预训练模型各层中插入小型、可训练的模块, 微调时仅更新这些模块的参数。	提供了一种模块化的模型定制方案, 可为不同评估任务训练专用适配器。
低秩适应	通过在模型权重矩阵旁注入可训练的低秩矩阵来模拟参数更新, 从而以少量参数实现高效微调。	兼顾了微调性能与计算效率, 是目前为特定临床语境快速定制模型的主流高效方法之一。
思维链	通过在提示中加入逐步推理的示例, 引导模型在回答复杂问题时生成推理过程。	提升模型在复杂诊断任务中的推理能力和结果的可解释性。

3.1　数据准备

数据收集是构建大模型的基础环节, 决定了模型的性能和适用范围。社交媒体的推文为抑郁症的筛查和临床诊断提供了大量数据, 其中的文本信息能较好地表征用户的情感状况。Beniwal 和 Saraswat (2024)开发了一种应用于 Twitter、Facebook、Instagram 等多种社交平台的大模型(BERT-CNN), 文本通过分词、去除停用词、特殊字符、数字、标点符号和重复字符等步骤进行清洗, 同时将表情包转换成文本, 提取相关的情感信息。Kerasiotis 等(2024)使用 Reddit 的“抑郁症严重程度数据集”进行抑郁症早期识别, 该数据集包含 3,553 条推文, 分为最小、轻度、中度和重度四个抑郁严重等级。为增强数据集的鲁棒性, 研究者利用数据增强方法对少量类别进行了样本扩充处理, 采用平衡采样解决类别不均衡问题。社交媒体数据的预处理不仅仅要对文本信息进行精细化处理, 同时也要综合多模态信息和语言特性, 以提升抑郁症筛查模型的准确性和泛化能力。

电子病历数据因其包含丰富的纵向临床信息、诊断记录、用药历史及实验室检查结果, 已成为抑郁症筛查与诊断大模型的重要数据源。Meng 等(2021)从所在医院的电子病历系统中筛选了 43, 967 名因特定基础疾病(如心肌梗死、乳腺癌或肝硬化)就诊的患者记录, 获取了病人的结构化诊断编码国际疾病分类第 9 版(International Classification of Diseases - 9, ICD - 9)、操作编码、药物清单、人口学信息及通过对临床笔记提取的非结构化文本特征, 为模型辅助诊断抑郁症提供了数据基础。McCoy 等(2025)利用大语言模型直接从电子病历数据库的临床笔记中提取抑郁症严重程度信息, 收集了 15, 000 条门诊临床笔记及患者的健康问卷-9 (Patient Health Questionnaire-9, PHQ-9)评分作为参考。

临床访谈数据源于临床医生或研究人员与患者的直接对话, 是一种主动获取的信息。它作为高价值数据源, 能为大模型提供丰富的上下文和与诊断高度相关的信息。其核心价值在于, 对话过程中蕴含了患者的情感状态、认知模式与症状表现等多维度的词语及语义信息。Lorenzoni 等(2024)通过设计引导问题的方式获得的患者自述, 并利用GPT-4从文本中评估抑郁严重程度, 验证了主动收集的对话对大模型驱动心理评估的价值。Sadeghi 等(2023)通过标准化的虚拟访谈员主动收集得到了 DAIC/E-DAIC 数据集(Extended Distress Analysis Interview Corpus, E-DAIC), 用于从访谈转录文本中预估 PHQ-8 评分。访谈后转录的文本由临床医生进行诊断标注或症状评分, 为大模型训练提供了可靠的“金标准” (Jarvers et al., 2024)。

虽然基于文本的数据目前已被广泛地用于抑

郁症的筛查和诊断，但抑郁症作为一种在生理、心理及行为层面均具有复杂表现的精神障碍，单一文本模态数据难以全面捕捉其病理特征。因此，收集并融合多种不同来源的多模态数据，如语音、视觉(面部表情、行为活动)、生理信号乃至基因组学信息，已成为提升抑郁症筛查与诊断性能的重要方向。Wang 和 Zhang (2024)将对话信息基于 BERT 的文本分析和基于图像识别模型(Vision Transformer, ViT)的音频频谱图信息结合在一起处理，融合了文本的语义内容与音频的声学特征，显著提升了抑郁症诊断的准确性。Englhardt 等(2024)利用移动和可穿戴传感器(如智能手机和智能手表)收集了包括步数、屏幕使用时间、电话通话时长、位置信息及睡眠时间等在内的丰富行为健康数据，提取行为特征筛查抑郁症。不同数据来源在大模型研究中的应用特性如表 3 所示。

3.2 模型选择

基座大模型通过在海量数据上预训练获得丰富知识和通用能力，是构建抑郁症筛查与诊断大模型的基础。目前，可作为抑郁症研究的基座大模型包括大语言模型、视觉基座模型、语音基座模型和多模态大模型。大语言模型具有强大的自然语言处理能力和活跃的开源生态系统，在抑郁症筛查和诊断领域表现出比较突出的应用潜力。大语言模型能够准确捕获语言中的情感、认知与行为线索，能够理解上下文，因此可以实现情感分析、症状筛查、生成心理评估报告等功能。Shin 等(2024)基于 ChatGPT 训练了从用户日记文本中筛查抑郁症的方法。Lorge 等(2025)运用 BERT 模型对难治性抑郁症临床相关因素进行跨度提取。Gu 等(2024)以 ChatGLM3 为基座模型，开发了一个具有心理状态追踪模块的抑郁症诊断对话系统。此外, LLaMA (Carstensen et al., 2024)、Qwen (Xu et al., 2025)等基座模型也展现出抑郁症筛查与诊断的初步应用潜力。在选择大语言模型方面，需兼顾专业性与交互性需求，需要借助抑郁症数据集进行微调，从而让大模型能够在抑郁症领域的任务中表现出优秀的性能。

视觉基座模型是经过大规模图像或视频数据集预训练后、能够提取通用视觉特征表示的模型。该模型可被应用于对抑郁症相关的非语言行为指标进行视觉通道上的分析。抑郁症常伴随有消极的面部表情、视线回避、肢体活动减少等症状，因此对面部区域进行视觉特征分析能够筛查和诊断抑郁症(He et al., 2021; Zhou et al., 2018)。Zhang 等人(2023)采用基于 Transformer 架构的 ViT 模型，通过知识蒸馏提高大模型对于抑郁症筛查任务的可部署性和准确性，在图像分类和情感分析任务中都有较为良好的表现。为了从视觉模态中识别抑郁症的客观行为指标，首先需要采集高质量的视觉数据。

语音基座模型需要对大规模语音数据集进行自监督训练，提取语音信号的通用声学特征表示。语音信号中蕴含丰富的情感: 韵律、音色、音调、停顿、节奏等声学特征是体现情绪状态和心理健康状况的重要指标(Schuller, 2018)。Gerczuk 等(2023)以 Wav2Vec 2.0 语音基础模型为基础，利用个性化元数据调整实现了对抑郁症的筛查。尽管语音基座模型对于识别抑郁症的语音特征已经越来越敏感，但其对于大模型的作用深度与作用广度都还有待进一步拓展。

3.3 模型训练

为增强模型在抑郁症相关任务的性能，通常采取将基座大模型先进行领域适应性预训练，再进行大模型微调，以提高抑郁症大模型输出结果的准确性。领域适应性预训练通过利用抑郁症领域的专业知识，对基座大模型再进行二次预训练。抑郁症数据具有高度的多样性与复杂性，因此大模型必须经过大量、多源的数据进行训练，例如临床病历、临床访谈、心理量表结果、患者

表 3 不同数据类型的比较

	社交媒体数据	电子病历数据	临床访谈数据	多模态数据
数据规模	极高	中等	低	低
数据质量	低	较高	高	较高
临床诊断相关性	中等	高	极高	极高
采集与处理成本	低	中等	高	极高
模型实现复杂度	中等	较高	较高	极高

自述、社交媒体信息等。通过领域适应性预训练, 大模型能够更好地对抑郁症患者存在的核心症状表现进行学习, 从而使大模型在抑郁症筛查和诊断的任务中有更强的适应性及泛化性。Danner 等人 (2023) 利用抑郁分析访谈语料库 (Distress Analysis Interview Corpus-Wizard of Oz, DAIC-WOZ) 和扩展抑郁分析访谈语料库(Extended Distress Analysis Interview Corpus, Extended-DAIC)数据集中的临床访谈文本数据对大模型进行预训练, 显著提升了大模型的抑郁症筛查性能。Englhardt 等(2024)利用 GLOBEM 数据集中多源被动感知数据(智能手机和可穿戴设备收集的感知数据, 如步数、GPS 位置、电话使用情况、社交活动等), 采用多模态数据融合的方法进行大模型的预训练, 探索了非文本数据在抑郁症大模型训练中的应用。

当完成领域适应性预训练后, 还要使用领域内小规模、高相关的专业数据集对大模型进行微调, 进一步强化大模型在抑郁症筛查和诊断任务上的性能。大模型的主要微调技术如图 4 所示, 包括指令微调、参数高效微调和对齐微调。指令微调指向大模型中输入大量“指令−输出”对作为训练样例, 指导大模型遵守特定指令(例如“请根据以下文本判断抑郁风险等级并陈述理由”)做出任务结果的输出, 可以让模型在保持权重相对不变的情况下提升其对特定任务理解的能力。Shin 等人(2024)利用思维链提示技术, 通过提供结构化的分析流程(例如包括情境与态度提取、情绪词汇识别及基于此信息的分类), 针对大模型的内部推理过程施加了更精细的任务指令, 结果表明通过指令微调能够显著提升大模型在抑郁症筛查任务的性能。参数高效微调方法(如低秩适配、适配器)仅训练模型中的少量参数, 从而极大地节省了算力和时间成本, 可以实现在有限的资源情况下对大模型进行专门定制, 从而有力地提高其性能。Gu 等人(2024)采用低秩适应对模型进行微调, 将心理状态跟踪模型与回应生成模型联合优化, 预测患者当前心理状态和下一步行动, 根据对话历史和患者的当前心理状态生成回复。对齐微调的目标是大模型的生成输出内容与人类偏好、社会价值观、伦理规范、特定领域的专业要求等一致。例如, 可以通过人类反馈强化学习等方法以确保大模型抑郁症筛查和诊断任务的生成答案不仅准确, 并且安全、负责和有益, 以避免产生偏见、不准确或有害的信息。Yang, Zhang 等人(2024)提出利用人类反馈强化学习结合少量示例对模型进行微调, 使之生成准确且符合临床规范的抑郁症症状分析报告。对齐微调是保障大模型在抑郁症领域应用的重要环节, 其核心价值在于实现技术应用的伦理可接受性。

3.4　模型评估

在抑郁症的筛查与诊断任务中, 最为重要的评价指标是大模型的分类能力, 即区分抑郁症患者和非抑郁症患者的能力。分类性能的评价常用混淆矩阵中的四个基本概念: 真阳性(True Positives, TP), 即实际为抑郁症并被正确识别为抑郁症; 真阴性(True Negatives, TN), 即实际为非抑郁症并被正确识别为非抑郁症; 假阳性(False Positives, FP), 即实际为非抑郁症而被错误地识别为抑郁症; 假阴性(False Negatives, FN), 即实际为抑郁症而错误地识别为非抑郁症。 基于混淆矩阵, 研究者常用的评价指标包括: 准确率(Accuracy, ACC)、精确率(Precision, PPV)、召回率(Recall, Sensitivity, SEN)、F1 分数(F1-score)以及 ROC 曲线下面积(Area Under the ROC Curve, AUC)。上述指标可对大模型进行效能的评价, 衡量预测阳性准确率(PPV)、避免漏诊的能力(SEN), 以及不同分类阈值下整体性能。

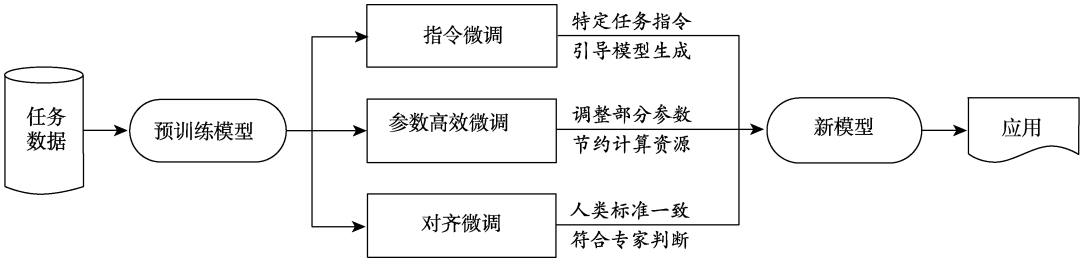

图 4　大模型训练流程

虽然对抑郁症分类判断是大模型评价的基础性能, 但对于临床实践来说严重程度的判断有利于指导诊疗决策。通过回归任务, 大模型可以预测出患者抑郁症的程度, 评价模型的性能主要通过平均绝对误差(Mean Absolute Error, MAE)、均方根误差(Root Mean Squared Error, RMSE)等指标衡量。MAE 和 RMSE 衡量大模型预测的程度与临床实践程度之间的偏离程度, 反映了大模型准确描述抑郁症状严重程度的能力, 从而为抑郁症的干预和治疗提供证据。

大模型在自然语言生成方面具有内在优势, 能够生成抑郁症相关文本内容。生成内容的性能主要根据自然语言处理领域成熟的指标进行评价, 包括双语评估替换指标(Bilingual Evaluation Understudy, BLEU)、面向召回率的摘要评估辅助指标(Recall-Oriented Understudy for Gisting Evaluation, ROUGE)、显式词序匹配翻译评估度量(Metric for Evaluation of Translation with Explicit Ordering, METEOR)和不同 n 元语法比例(Distinct n-grams, DIST-n)。这些指标从词汇相似度、信息覆盖率、语义准确性、流畅度、内容丰富性及新颖性等多个维度, 评估大模型生成文本的临床适用性。

尽管客观量化指标能提供重要的模型性能评估数据, 但由于抑郁症筛查和诊断属于高度敏感的临床问题, 仅以客观量化指标评定可能存在一定的局限性, 因此, 依据临床专业能力的定性评估是量化评价必要的补充, 有助于确保大模型输出结果的临床适用性和合规性。定性的专业能力评估工作通常由具有丰富临床经验的心理专业人员、精神科医生等专业人士实施, 主要包括两个方面: 一是大模型输出的临床质量与准确性, 包括筛查或诊断建议的清晰度、易理解性, 以及对症状模式和危险因素识别的一致性; 二是大模型在临床应用中的实用性与用户接受度, 包括模型能否顺利集成至现有临床工作流程、能否提升工作效率, 以及临床医生使用大模型工具的接受意愿等。表 4 总结了纳入分析文献的大模型性能评估指标。

表 4 大模型在抑郁症筛查与诊断任务中的性能评估指标

指标类型	具体指标	指标说明	测试方法	评分范围
分类任务性能	准确率(ACC)	正确识别抑郁与非抑郁个体的比例	(TP + TN) / (TP + TN + FP + FN)	0%-100%
	精确率(PPV)	预测为抑郁的个体中, 实际为抑郁的比例	TP / (TP + FP)	0%-100%
	召回率(SEN)	实际为抑郁的个体中, 被成功识别为抑郁的比例	TP / (TP + FN)	0%-100%
	F1 分数(F1)	精确率和召回率的调和平均值	2 × (PPV × SEN) / (PPV + SEN)	0-1
	AUC	ROC 曲线下面积, 评估分类器整体区分能力	通过 ROC 曲线计算	0.5-1
回归任务性能	平均绝对误差(MAE)	预测的抑郁得分与真实抑郁得分之差绝对值的平均数	$(1/n)\,\Sigma \mid y_i - \hat{y}_i \mid$	≥0
	均方根误差(RMSE)	预测的抑郁得分与真实抑郁得分之差平方的均值的平方根	$\sqrt{[(1/n)\,\Sigma (y_i - \hat{y}_i)^2]}$	≥0
	相关系数(r)	预测的抑郁得分与真实抑郁得分之间的线性相关强度	皮尔逊或斯皮尔曼相关	−1 到+1
文本生成性能	双语评估替换指标(BLEU)	衡量生成文本与参考文本的表面相似度	与参考文本比较	0-1
	面向召回率的摘要评估辅助指标(ROUGE)	衡量生成文本对参考文本信息的覆盖程度	与参考文本比较	0-1
	显式词序匹配翻译评估度量(METEOR)	综合精确率、召回率、同义词与语序	词对齐与映射	0-1
	不同 n 元语法比例(DIST-n)	衡量生成文本中独特 n-gram 的比例	独特 n-gram 占比	0-1
专业能力	模型输出的临床质量与准确性	输出的清晰度、相关性、一致性	访谈、问卷	定性评价
	临床应用的实用性与用户接受度	在真实场景中的可行性与用户采纳意愿	访谈、问卷	定性评价

4　基于大模型的抑郁症筛查和诊断的研究应用

4.1　早期筛查

抑郁症的早期筛查是疾病预防和及时干预的关键步骤, 对患者的治疗效果有着重要价值。目前临床较为常用的方式是基于心理量表的标准化筛查, 但量表评估结果容易受到患者主观报告偏差及回忆偏倚等因素的影响, 从而导致筛查结果可靠度及一致性的下降(Smith et al., 2013)。基于文本语义、语音声学、面部表情、行为动作等多模态数据的大模型智能筛查, 研究者能逐步建立更加准确的抑郁症筛查模型, 为抑郁症早期筛查提供新方法。

文本是情感与认知的信息载体, 大模型可通过强大的自然语言处理能力, 对非结构化文本语料中的抑郁症相关数据进行提取与分析。社交媒体平台存储着海量的用户生成文本, 能够反映其真实的情感状态和认知模式。Shen 和 Paik (2023)基于 Twitter 数据库(共包含 160 万条推文)训练了 BERT 模型, 在抑郁倾向筛查任务中得到了 83.28%的预测准确率。Bendebane 等(2025)开发的 DAC-BERT 通过对推文文本的分析, 在识别高风险人群的准确率达到了 96.50%。El-Ramly 等人(2021)在 CairoDepv1.0 社交媒体数据集中训练多方言阿拉伯语 BERT 模型(Multi-dialect Arabic BERT, MARBERT)和阿拉伯语 BERT 模型(Arabic BERT, ARABERT)对阿拉伯语推文进行抑郁筛查, 准确率分别达到 96.93%和 96.07%。

语音是情绪表达的重要渠道, 大模型在处理复杂声学表征方面展现出较高性能。抑郁症患者的声学表现有韵律减少、停顿增多及言语不流畅等特征(Stasak et al., 2019; Menne et al., 2024)。Gupta 等人(2024)开发的 RADIANCE 系统采用滤波器组视觉变换器(FilterBank Vision Transformer, FBViT)处理语音信息, 在 CMDC 数据集上的抑郁症筛查准确率达到 94.44%。Wu 等人(2023)提出了用自监督学习预训练基础模型的方法, 在 DAIC-WOZ 数据集上取得了 0.89 的 F1 分数, 验证了大模型在抑郁症语音筛查作用。Jiang 等人(2024)采用多模态大模型, 从在线访谈的视频中提取了面部表情、语音、对话转录文本以及生理信号等多模态特征, 在重度抑郁症筛查任务中达到了 0.77 的 AUC。

尽管前述研究已在一定程度上体现了大模型在抑郁症筛查和诊断应用的前景, 特别是针对文本、语音数据方面效果显著, 但研究还存在以下不足: 一是现有研究的高准确率均基于特定的、清洗过的大数据集所得, 在更泛化、异质性更高的人群中是否仍具有稳健性仍需要进行检验; 二是跨文化环境下抑郁症表现具有一般性差异, 不同文化语境下抑郁症大模型的适应性还需要进一步探索; 三是不同研究分别采用了不同评估指标和基准方法, 大模型的横向比较仍需要慎重审慎。

4.2　辅助诊断

抑郁症的临床诊断中, 医生需要综合考虑患者自述、家属描述、量表测评、生理生化结果等大量的非结构化数据。大模型可以快速读取并理解相关信息, 提取出相对简明扼要的病情摘要, 依据抑郁症临床诊断标准辅助临床决策。Guo 和 Guo (2024)基于 GraphRAG 框架, 采用精神障碍诊断与统计手册 5(The Diagnostic and Statistical Manual of Mental Disorders-5, DSM-5)和 ICD-11 中的抑郁症诊断标准, 建立抑郁症的知识图谱作为大模型推理的基础知识库, 利用 ChatGPT o1 大模型生成思考链模板, 结合诊断标准及分析步骤, 最终获得 0.84 的抑郁症诊断准确率。对于抑郁症诊断任务, 大模型在准确性、专业性、可靠性和实用性等方面, 也获得了临床医生的高度评价。D'Souza 等人(2023)对大模型的诊断专业性做了基准性评估。在临床病例测试中, 共收集了 100 例包含抑郁症在内的多种心理障碍诊断病历。ChatGPT 3.5 取得了较好表现: 61%的病例获得"优秀", 31%的病例获得"良好", 仅有 8%的病例被评为"需要改善"。

虽然大模型已经展现出处理复杂临床信息的能力, 但将其应用于临床进行抑郁症诊断还存在着如下三方面的障碍。首先是大模型诊断建议的可靠性取决于大模型训练数据和知识库的质量, 需要确保知识库的持续更新, 并适应临床指南的变化。其次是抑郁症的临床诊断并非简单的症状标准条目匹配, 对于患者的个体差异、社会文化背景、非言语线索等因素的综合考量是大模型在当前阶段力所不能及的。最后是目前模型解释的"黑箱"问题, 大模型的可解释性不足, 给临床医

生的提示参考带来了障碍。

4.3 风险预测

抑郁症是一种较为复杂的精神障碍，与遗传、环境、神经生物学等因素密切相关。研究指出，抑郁症具有很强的遗传倾向，其遗传率在30%~50% (Kendall et al., 2021; Flint, 2023)。虽然一部分具有家族倾向的个体并没有表现出严重抑郁的症状，但其在情绪调节、认知能力和社会行为等方面可能和正常个体存在明显差异(King et al., 2006; MacKenzie et al., 2019)。Gao 等人(2024)选取了 84 个包含血液和脑区数据的基因表达综合数据库，用分箱法和 gene2vec 将基因表达数据进行预训练，基于 Performer 编码器构建了 DP-BERT 模型。该模型在预训练时通过屏蔽策略学习基因之间交互信息，在微调阶段进一步学习与抑郁症相关的基因表达特征，结果发现，DP-BERT 在抑郁症数据集上的准确率达到 0.939，AUC 为 0.979。大模型可以从海量的基因数据信息中提取相关特征，能够为预测抑郁症提供生物学参考。

大模型解析基因数据进行抑郁症风险预测为从生物学层面理解乃至干预抑郁症提供了新的思路，但如何从生物学层面提升大模型预测抑郁症的临床应用价值，仍面临巨大挑战。首先，由于抑郁症具有高度异质性，其发生由遗传与环境因素共同驱动。当前仅依赖基因表达数据的预测模型，因忽略了环境因素的关键驱动作用，存在方法论上的局限，导致其预测效能受限。同时，基因表达数据本身存在着非常高的动态性和组织特异性，虽然 Gao 等人(2024)的研究涵盖了血液及相关脑区，但是抑郁症相关脑区的样本获取在实际应用当中也较为困难。因此，大模型预测抑郁症的性能，仍有待于在大规模前瞻性队列研究中得到验证，以确认其可重复性与外部有效性。

5 机制解释与理论支撑

大模型在抑郁症筛查与辅助诊断领域展现出良好的性能，深入分析其机制表明，大模型的计算原理与心理学对抑郁症的理论解释之间有一致性。这种跨领域的理论契合是理解大模型有效性的基础，也为计算心理学提供了新的范式。

5.1 情境化语义表征

大模型的动态的、上下文感知的语义表征能力使其能够在理解抑郁症患者语言的过程中，超越对个别负性词汇的测量，在完整语境中捕捉患者特有的思维认知。根据抑郁症的认知模型，抑郁症的一个主要特征就是系统性的负性认知偏差。这一认知特点并非是源于少数的负面词汇，而是每一位患者特有的功能失调性图式，使个体自动地、有偏向地对事件信息进行筛选、加工和解释，进而产生抑郁情绪(Beck, 2008)。因此，关注在相关语境下的负性语义解读，是理解抑郁症的关键。

传统的自然语言处理方法给每个单词或词组分配一个固定、上下文无关的向量表示。比如在静态嵌入模型中，“我感觉自己像个负担”中的“负担”与“这个责任我能负担得起”中的“负担”，它们的向量表示是完全相同的，这显然不能表现出人类语言的丰富性。大模型的一项根本性改进是实现了词义的动态上下文表征，取代了传统的静态词向量。其技术原理就是模型处理一个词时并不孤立地对待，而是把它放在整个句子或段落的上下文中，并考虑该词对所有上下文词语的依赖程度。再通过一个深度、多层次的神经网络模型对输入的文本反复进行加工，在每层加工过程中，一个词的向量表示被其周边词语的信息所精炼。最终，大模型能够为同一词汇在不同语境中生成动态向量，这从数学上实现了对词语多义性的精准区分与表征。这个计算过程就是在数学上高度模拟了人类的语言理解中依赖语境的认知过程，也是解读抑郁症语义的重要基础。

5.2 注意力机制

在大模型中，注意力机制对文本序列各信息的加权聚焦计算过程可看作是对抑郁症患者的注意偏向认知过程的计算模拟，使得模型会优先处理与抑郁症高度相关的诊断线索。抑郁症的注意偏向包括了外部注意偏向和内部注意偏向(Mennen et al., 2019)。外部注意偏向指的是个体在感知外部环境时，有倾向地关注负性刺激(如悲伤面孔、负面词汇)。内部注意偏向则是内在的表征，包括思维、记忆和情感，而非直接对外部刺激导致的注意偏向。抑郁症患者的内部注意偏向体现为一种认知控制缺陷：一旦负性想法或记忆被激活，个体无法将注意力从中抽离出来，出现对负性内容的认知固着及抑制困难。

大模型的注意力机制模拟了抑郁患者的外部

注意偏向和内部注意偏向。一方面, 注意力机制在单层网络中分配注意力权重的过程, 模拟了外部注意偏向。比如在识别句子“尽管这个项目取得了成功, 但是我总是觉得自己是个失败者”时, 大模型能通过学习把注意力分配到“失败者”等表达自己负性评价的词汇上, 实现了对环境中负性信息的选择性捕捉。另一方面, 注意力机制在多层网络间的传递模拟了内部注意偏向的认知固着特点。当一个底层被赋予高权重的负性概念, 其形成的语义表达会持续向上一层级进行传递, 并始终占据优势地位, 持续影响大模型对整个文本的理解。这种跨层级的持续影响力, 是对抑郁患者认知控制缺陷的一种模拟计算。

5.3　多模态行为捕捉

大模型能够整合视觉、语音、文本等多种模态信息, 通过量化抑郁症患者的行为表现, 实现抑郁症的客观化识别。运动迟滞是抑郁症主要的行为特征, 表现为言语、动作及认知加工的全面减慢, 其生物学机制与神经环路异常关系紧密。行为层面, 患者表现为动作减少、面部表情僵硬、言语迟滞、认知迟滞等行为特征(Buyukdura et al., 2011)。

多模态大模型可以融合视觉、语音、文本等各类信息, 对抑郁症进行全面、精准的评估。对于视觉信号, 大模型主要基于卷积神经网络提取面部关键点位置特征并定位眼眉、嘴角等面部单位, 量化表情僵化程度, 捕捉抑郁症患者常见的面部特征。大模型也能分析语音信息, 提取停顿时长、语速、语调等声学参数, 发现言语迟缓特征。对于文本信息, 大模型可以通过语义编码, 检测消极词与认知迟缓相关的语言表达。由于抑郁症的复杂性, 单模态信息难以呈现抑郁症的全面特征, 多模态大模型通过提取与处理单模态信息, 并通过跨模态联合学习, 完成对自然语言理解、情绪评估、非言语行为分析等复杂任务的处理(陈露等, 2023)。这种机制不仅实现了对运动迟滞临床表现特征的量化捕获, 还通过多层级特征融合, 有效提高了识别的特异性与敏感性, 为抑郁症的筛查提供了计算基础。

5.4　生成模型的预测范式

预测加工框架(Predictive Processing, PP)为抑郁症的认知缺陷提供了整合的理论框架, 其原理与大语言模型的机制具有高度的同构性。它认为大脑并不是一个被动的信息接收器, 而是一个主动的预测生成系统(Friston, 2005; Clark, 2013)。根据内部生成模型对感觉输入持续进行预测, 感觉输入与预测间的失配会产生“预测误差”, 大脑通过调整先验或行动将预测误差最小化。据此, 抑郁症的病理机制可以理解为一种预测性学习的失调: 个体对消极先验赋予了过高的精度(Kube et al., 2020)。对消极先验的过度确信, 会导致负面信息被赋予过高权重, 而对与之冲突的积极信息所产生的预测误差则被压制性忽略。这种认知偏差抑制了信念的有效更新, 从而形成了一种自我延续的、陷于负面信念的适应不良回路(Badcock et al., 2017)。该模型也可用于躯体症状的解释, 即对疲乏、疼痛等机体内部感受的负面预测主导了患者的生理体验。

作为基于海量数据训练的模型, 大模型由于其庞大的参数实质上构成了关于语言和世界知识的复杂内部生成模型, 其优化过程同样是对“预测误差” (即模型预测词元与真实词元间的差异)的最小化。大模型识别抑郁症的内在机制, 并非简单依赖于对临床术语的理解, 而是通过概率统计的方式, 量化用户语言语料与模型内部训练语言间的差异。换而言之, 抑郁症患者语言中高频出现的消极的、绝对化的词汇可以被视为其“高精度消极先验”的语料标签; 其在积极语境下对正面事件的弱化, 则对应于其对“积极预测误差”的认知衰减机制; 对其躯体不适的高频表达则表征了患者“内感受预测”的失调。这些独特的语言表征在统计分布上显著偏离常态, 因而被大模型识别为抑郁症的重要信号。

6　挑战

6.1　算法偏见

大模型训练多聚焦于普通成人样本, 这导致模型在向青少年、老年人及不同文化背景的少数群体等特殊人群迁移应用时, 可能表现不佳, 甚至产生算法偏见。大模型的数据源主要来自互联网公开文本, 这些数据不可避免地会包含社会中普遍存在的刻板印象、歧视性语言和文化偏见(Zhao et al., 2023; Wei et al., 2023; Sadeghi et al., 2023)。大模型学习上述有偏的数据之后, 将会在决策过程中再现或扩大原有数据集中的偏见。已有研究发现, 针对特定群体的大模型会延续源数

据的偏见，从而低估其内部抑郁表现的复杂性(Long et al., 2024)。尽管算法偏见难以被彻底根除(Townson, 2023; Mittermaier et al., 2023)，但为了改善大模型的公平性和有效性，研究人员必须采取系统性的对策。一是在数据层面，构建和标注不同年龄、文化以及少数群体的用户数据集，提高数据的质量。二是在算法层面，可以考虑引入领域自适应、对抗性训练等方式来提升模型针对不同人群数据分布的泛化性。三是在评估层面，应避免过分地追求大模型的准确率，提倡采用分层评估方式针对亚群体的特征进行分类评价。

6.2 诊断特异性

诊断特异性也是大模型在抑郁症的筛查和诊断中面临的另一项挑战，即如何精确地将抑郁症与其它具有高度共病性的心理障碍(尤其是焦虑症)区分开来。很多抑郁症大模型面对含有焦虑症状的文本时会出现召回率高、精确率低的情况，将大量焦虑症状的文本错误识别为抑郁症信号。主要原因是由于抑郁症和焦虑症在临床症状和自然语言表达层面上深度重叠，二者基本都包含情绪低落、精力减退、睡眠障碍、注意缺陷和负性认知等核心症状特征，因而二者非常难以区分，这也是多数研究选择分别处理单一精神障碍的原因(Bendebane et al., 2025)。为有效应对抑郁症共病问题，提高大模型对抑郁症的诊断特异性，未来研究需要从二元分类转向多类别或多标签分类框架。研究人员不能简单地将任务设定为判断是否抑郁，而应通过标注高质量的数据，构建能同时识别“抑郁症”“焦虑症”“正常”以及其他相关心理障碍的多类别分类模型。此外，还可以进一步研究多标签分类模型，允许模型为一个患者同时标注“抑郁症”和“焦虑症”两个标签，这也更加接近临床共病的真实情况。

6.3 幻觉现象

幻觉是指大模型生成的文本结果所包含的事实性错误或不合适的上下文内容。在抑郁症筛查与诊断应用中，大模型可能会生成与患者病情不符但看似合理的诊断结论，或提供非事实性的、过于乐观或过分悲观的医疗建议(Berrezueta-Guzman et al., 2024)。大模型生成的回复表达较为流畅，但其“权威性”与“专业性”可能让用户忽略潜在的准确性问题(Liu et al., 2024)。幻觉现象的根源在于大模型的“黑箱”特性导致其决策过程不透明，致使输出结果具有不确定性。为了应对幻觉风险，在技术层面应发展可解释人工智能技术，打开模型决策黑箱。通过呈现模型做出判断所基于的关键用户输入，可解释人工智能技术可以使其推理过程更透明，便于专业人士及时发现可能存在的幻觉内容。应用流程层面，关键是建立人机协同的工作流，即确保所有模型输出都经过专业人士审核。另外，使用结构化输出等交互设计手段，也可以约束模型生成过程，降低幻觉风险。

6.4 隐私安全

大模型在抑郁症领域的应用，面临严峻的数据隐私与安全挑战。大模型训练的数据来自海量的使用者信息采集，存在数据所有权、使用许可权、数据代表性等伦理问题(Cohen, 2023)。抑郁症患者的个人信息往往比较敏感，如果患者的个人信息被非法获取或未经授权的外泄，不仅侵犯了患者的基本权力，甚至会给患者造成二次伤害，加重抑郁症症状(Liu et al., 2024)。如果大模型没有稳健的安全措施，也可能受到恶意攻击，产生数据泄露的危机(Abdulai & Hung, 2023)。为应对隐私安全风险，未来的研究不仅应遵循数据最小化原则，即仅收集服务于特定功能所必需的最少数据，更需积极采用前沿的隐私保护计算技术。例如，联邦学习作为一种先进的分布式训练框架，允许模型在用户的本地设备上进行训练，仅将加密后的模型参数上传至中央服务器进行聚合，而原始敏感数据始终不离开用户端，从根本上降低了数据集中存储带来的泄露风险。在此基础上，还可结合差分隐私技术，通过在上传的参数中注入经过精确计算的噪音，为数据提供数学上可证明的隐私保障，使得攻击者即便获取了最终模型也难以反推出任何个体的具体信息。

6.5 伦理问题

大模型在临床实践中的角色定位与人机关系也引发了深刻的伦理反思。抑郁症筛查和诊断大模型具有明显的成本优势和便捷性，患者可能会将大模型视为其心理问题的主要诊断者和决策者(Ma et al., 2023)。医生和心理治疗师也可能过度参考大模型生成的建议，弱化自身在临床评估中的主动性(Elyoseph et al., 2024a; Perlis et al., 2024; Blease et al., 2024)。这种趋势可能会导致抑郁症诊疗实践的去人性化，将抑郁症患者简化为数据

化标签, 忽视其整体性(Palmer & Schwan, 2022; Haque & Waytz, 2012)。因而, 大模型未来的发展需要计算机科学家、心理学家和伦理学家的跨学科交叉合作, 共同建立负责任的大模型应用框架。这既需要在大模型的设计过程中嵌入一套算法伦理制约机制, 使算法程序运行中的人类价值得以合理规范, 更重要的是构建起人机分工的界限, 始终将大模型作为辅助人类专家的临床决策、增强临床诊疗效率的工具而非代替专业判断的决策者。

7　展望

7.1　心理干预的整合应用

除了在抑郁症的筛查与诊断领域展现出巨大潜力外, 大模型在心理干预中的整合应用, 正成为一个备受瞩目的前沿方向。传统的心理咨询对话机器人(ChatBot)受限于有限的理解能力, 往往难以提供真正个性化和共情性的支持。而大模型凭借其强大的自然语言生成、理解和推理能力, 正在从根本上改变这一现状, 其应用不再局限于简单的信息提供或症状记录, 而是深度整合到干预过程的核心环节。Sharma 等(2023)提出的“共情增强器”模型可实时分析咨询师的文本并生成更高共情水平的改写建议, 通过人工智能辅助而非替代人类的方式, 强化了治疗联盟的核心要素。Sabour 等(2023)研发的 Emohaa 系统则整合了结构化认知干预与生成式情感支持, 随机对照试验结果已证实其在缓解抑郁症状方面效果显著。构建从筛查、诊断到干预的全流程一体化大模型成为未来研究的重要方向, 最终实现对个体心理健康的个性化、全周期闭环管理。

7.2　临床转化路径

从实验环境走向临床部署的转化路径, 是大模型在抑郁症领域落地的关键。现有大模型的验证主要基于回顾性数据, 结论能否有效泛化至真实临床环境尚需继续验证。已有研究发现, 人工智能工具在临床实践中部署后, 其性能与实验室结果存在显著差异(Akhlaghi et al., 2024)。因此, 后续研究应当设计前瞻性临床试验, 在临床实际的诊疗流程中对模型提升工作效率、辅助诊断决策、改善患者结局等方面进行评估。此外, 提高模型可解释性从而建立临床信任十分重要, 可采用可视化的方式直观地向医生呈现模型决策的过程(Joyce et al., 2023)。例如, 可通过注意力机制将最能影响模型决策的文本高亮显示出来, 从而直观地向临床医生展示模型的思考过程, 这是未来医工合作的攻关重点。最后, 部署的可行性与数据的安全性也是大模型通往临床应用的现实障碍。鉴于心理健康领域文本数据的高度敏感性, 发展可在医疗机构内部安全运行的私有化、轻量化模型已成为必然的技术方向(Blease & Torous, 2023)。Taylor 等人(2024)发现在单个图形处理器(Graphics Processing Unit, GPU)这样的有限计算资源上高效运行先进模型的可行性, 从而为在资源紧张的临床环境中实现安全、本地化的大模型应用开辟了现实路径。

7.3　未来技术创新的研究方向

未来, 大模型需要从单一数据源的模式识别迈向对个体复杂性的深度理解, 构建更为精细、动态且具备文化适应性的抑郁症数字表型。大模型需解决多模态数据融合、时间序列建模的局限, 通过对齐标准化框架设计实现对文本、视觉、语音和生理等多源异构数据的深度融合, 结合患者的情绪、行为等长期时间序列数据的动态建模, 大模型将能够实现更为全面和立体的抑郁风险识别(Jiang et al., 2024; Tlachac et al., 2023)。此外, 抑郁症的语言表达行为具有一定文化特异性, 在单一文化背景下训练的模型难以在不同人群中有效泛化(Teferra et al., 2024)。为此, 未来要构建跨文化抑郁语义知识图谱, 对不同文化背景下的抑郁症语言标记、情绪隐喻、行为模式与社会心理因素进行系统性刻画, 并以此作为先验知识支撑大模型训练, 使模型能够从浅层语言模式识别提升到深层文化语境理解, 有效加强抑郁症筛查和诊断的适用性。

综上所述, 在抑郁症的早期筛查、辅助诊断乃至风险预测等领域, 大模型均展现出了良好的应用前景。大模型正在以前所未有的方式挖掘和处理海量的文本、语音、图像、生理和行为数据, 为心理健康的数智化服务提供了新的技术可能。但是, 目前研究仍存在算法偏见、诊断特异性、幻觉现象、隐私安全及伦理问题等挑战。因此, 未来应加强大模型心理干预的整合应用, 聚焦临床转化路径, 构建更为精细、动态且具备文化适应性的抑郁症数字表型, 实现心理健康服务的数智化转型。

参考文献

陈露, 张思拓, 俞凯. (2023). 跨模态语言大模型: 进展及展望. *中国科学基金, 37*(5), 776−785.

陈晓红, 刘浏, 袁依格, 王俊普, 李大元, 邱建华. (2024). 医疗大模型技术及应用发展研究. *中国工程科学, 26*(6), 77−88.

董健宇, 韦文棋, 吴珂, 妮娜, 王粲霏, 付莹, 彭歆. (2020). 机器学习在抑郁症领域的应用. *心理科学进展, 28*(2), 266−274.

张冬瑜, 庄沐霖, 靳森源, 刘馨月. (2025). 基于隐喻信息和指令调优的心理疾病检测. *数据分析与知识发现*, https://doi.org/10.11925/infotech.2096-3467.2024.0450

Abdulai, A. F., & Hung, L. (2023). Will ChatGPT undermine ethical values in nursing education, research, and practice. *Nursing Inquiry, 30*(3), e12556.

Akhlaghi, H., Freeman, S., Vari, C., McKenna, B., Braitberg, G., Karro, J., & Tahayori, B. (2024). Machine learning in clinical practice: Evaluation of an artificial intelligence tool after implementation. *Emergency Medicine Australasia, 36*(1), 118−124.

Al Masud, G. H., Shanto, R. I., Sakin, I., & Kabir, M. R. (2025). Effective depression detection and interpretation: Integrating machine learning, deep learning, language models, and explainable AI. *Array, 25*, 100375.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). https://doi.org/10.1176/appi.books.9780890425596

Badcock, P. B., Davey, C. G., Whittle, S., Allen, N. B., & Friston, K. J. (2017). The depressed brain: An evolutionary systems theory. *Trends in Cognitive Sciences, 21*(3), 182−194.

Beck, A. T. (2008). The evolution of the cognitive model of depression and its neurobiological correlates. *American Journal of Psychiatry, 165*(8), 969−977.

Bendebane, L., Laboudi, Z., Saighi, A., & Bouziane, S. E. (2025). Fine-tuning the BERT model to predict depression and anxiety using multi-labeled Twitter data. *2025 4th International Conference on Sentiment Analysis and Deep Learning (ICSADL)* (pp.586−591). https://doi.org/10.1109/ICSADL65848.2025.10932995

Beniwal, R., & Saraswat, P. (2024). A hybrid BERT-CNN approach for depression detection on social media using multimodal data. *The Computer Journal, 67*(7), 2453−2472.

Berrezueta-Guzman, S., Kandil, M., Martín-Ruiz, M. L., Pau de la Cruz, I., & Krusche, S. (2024). Future of ADHD care: Evaluating the efficacy of ChatGPT in therapy enhancement. *Healthcare, 12*(6), 683.

Blease, C., & Torous, J. (2023). ChatGPT and mental healthcare: Balancing benefits with risks of harms. *BMJ Mental Health, 26*(1), e300884.

Blease, C., Worthen, A., & Torous, J. (2024). Psychiatrists' experiences and opinions of generative artificial intelligence in mental healthcare: An online mixed methods survey. *Psychiatry Research, 333*, 115724.

Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., ... Liang, P. (2021). On the opportunities and risks of foundation models. *ArXiv*. https://doi.org/10.48550/arXiv.2108.07258

Buyukdura, J. S., McClintock, S. M., & Croarkin, P. E. (2011). Psychomotor retardation in depression: Biological underpinnings, measurement, and treatment. *Progress in Neuro-Psychopharmacology and Biological Psychiatry, 35*(2), 395−409.

Carstensen, C., Small, N., Bhaskar, J., Lopez, B., Shrestha, A., & Rundensteiner, E. A. (2024). MInDS: Using large language models to screen for depression. *2024 IEEE MIT Undergraduate Research Technology Conference (URTC)* (pp.1−5). https://doi.org/10.1109/URTC65039.2024.10937571

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences, 36*(3), 181−204.

Cohen, I. G. (2023). What should ChatGPT mean for bioethics? *The American Journal of Bioethics, 23*(10), 8−16.

Danner, M., Hadzic, B., Gerhardt, S., Ludwig, S., Uslu, I., Shao, P., ... Rätsch, M. (2023). Advancing mental health diagnostics: GPT-based method for depression detection. *2023 62nd Annual Conference of the Society of Instrument and Control Engineers (SICE)* (pp.1290−1296). https://doi.org/10.23919/SICE59929.2023.10354236

D'Souza, R. F., Amanullah, S., Mathew, M., & Surapaneni, K. M. (2023). Appraising the performance of ChatGPT in psychiatry using 100 clinical case vignettes. *Asian Journal of Psychiatry*, *89*, 103770.

El-Ramly, M., Abu-Elyazid, H., Mo'men, Y., Alshaer, G., Adib, N., Eldeen, K. A., & El-Shazly, M. (2021). CairoDep: Detecting depression in Arabic posts using BERT transformers. *2021 Tenth International Conference on Intelligent Computing and Information Systems (ICICIS)* (pp.207−212). https://doi.org/10.1109/ICICIS52592.2021.9694178

Elyoseph, Z., Gur, T., Haber, Y., Simon, T., Angert, T., Navon, Y., ... Asman, O. (2024a). An ethical perspective on the democratization of mental health with generative AI. *JMIR Mental Health, 11*, e58011.

Elyoseph, Z., Levkovich, I., & Shinan-Altman, S. (2024b). Assessing prognosis in depression: comparing perspectives of AI models, mental health professionals and the general public. *Family Medicine and Community Health, 12*(Suppl. 1), e002583.

Englhardt, Z., Ma, C., Morris, M. E., Chang, C. C., Xu, X. O., Qin, L., ... Iyer, V. (2024). From classification to clinical insights: Towards analyzing and reasoning about mobile and behavioral health data with large language models. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 8*(2), 1−25.

Farruque, N., Goebel, R., Sivapalan, S., & Zaïane, O. R. (2024). Depression symptoms modelling from social

media text: An LLM driven semi-supervised learning approach. *Language Resources and Evaluation, 58*(3), 1013−1041.

Flint, J. (2023). The genetic basis of major depressive disorder. *Molecular psychiatry, 28*(6), 2254−2265.

Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society B: Biological Sciences, 360*(1456), 815−836.

Gao, J., Zeng, M., Li, Y., Wang, F., Zheng, R., Liu, J., ... Li, M. (2024). *DP-BERT: A pre-trained deep language model for depression prediction using microarray data. 2024 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)* (pp.919−924). https://doi.org/10.1109/BIBM62325.2024.10822456

Gerczuk, M., Triantafyllopoulos, A., Amiriparian, S., Kathan, A., Bauer, J., Berking, M., & Schuller, B. W. (2023). Zero-shot personalization of speech foundation models for depressed mood monitoring. *Patterns, 4*(11), 100873.

Gu, Y., Zhou, Y., Chen, Q., Zhou, N., Zhou, J., Zhou, A., & He, L. (2024). Enhancing depression-diagnosis-oriented chat with psychological state tracking. *ArXiv*. https://doi.org/10.48550/arXiv.2403.09717

Guo, Y., & Guo, Y. (2024). A knowledge graph and large language model-based framework for depression detection. *2024 International Conference on Image Processing, Computer Vision and Machine Learning (ICICML)* (pp.670−673). https://doi.org/10.1109/ICICML63543.2024.10958051

Gupta, A. K., Dhamaniya, A., & Gupta, P. (2024). RADIANCE: Reliable and interpretable depression detection from speech using transformer. *Computers in Biology and Medicine, 183*, 109325.

Haque, O. S., & Waytz, A. (2012). Dehumanization in medicine: Causes, solutions, and functions. *Perspectives on Psychological Science, 7*(2), 176−186.

He, L., Chan, J. C. W., & Wang, Z. (2021). Automatic depression recognition using CNN with attention mechanism from videos. *Neurocomputing*, *422*, 165−175.

Hur, J. K., Heffner, J., Feng, G. W., Joormann, J., & Rutledge, R. B. (2024). Language sentiment predicts changes in depressive symptoms. *Proceedings of the National Academy of Sciences, 121*(39), e2321321121.

Insel, T. R., & Cuthbert, B. N. (2015). Brain disorders? precisely. *Science, 348*(6234), 499−500.

Jain, B., Goyal, G., & Sharma, M. (2024). Evaluating emotional detection and classification capabilities of GPT-2 and GPT-Neo using textual data. *2024 14th International Conference on Cloud Computing, Data Science & Engineering (Confluence)* (pp. 12−18). https://doi.org/10.1109/Confluence60223.2024.10463396

Jarvers, I., Ecker, A., Donabauer, P., Kampa, K., Weißenbacher, M., Schleicher, D., ... Ludwig, B. (2024). MINI-KID interviews with adolescents: A corpus-based language analysis of adolescents with depressive disorders and the possibilities of continuation using Chat GPT. *Frontiers in Psychiatry, 15*, 1425820.

Jiang, Z., Seyedi, S., Griner, E., Abbasi, A., Rad, A. B., Kwon, H., ... Clifford, G. D. (2024). Multimodal mental health digital biomarker analysis from remote interviews using facial, vocal, linguistic, and cardiovascular patterns. *IEEE Journal of Biomedical and Health Informatics, 28*(3), 1680−1691.

Jin, Z., Bi, D., Hu, J., & Zhao, K. (2024). Evaluating the efficacy of AI-based interactive assessments using large language models for depression screening. *MedRxiv*. https://doi.org/10.1101/2024.07.19.24310543

Joyce, D. W., Kormilitzin, A., Smith, K. A., & Cipriani, A. (2023). Explainable artificial intelligence for mental health through transparency and interpretability for understandability. *NPJ Digital Medicine, 6*(1), 6.

Juarto, B. (2024). Sentiment analysis for mental health using boosting, bagging, and DeBERTa from social media data. *2024 IEEE International Conference on Control & Automation, Electronics, Robotics, Internet of Things, and Artificial Intelligence (CERIA)* (pp.1−6). https://doi.org/10.1109/CERIA64726.2024.10915163

Kendall, K. M., Van Assche, E., Andlauer, T. F. M., Choi, K. W., Luykx, J. J., Schulte, E. C., & Lu, Y. (2021). The genetic basis of major depression. *Psychological Medicine, 51*(13), 2217−2230.

Kerasiotis, M., Ilias, L., & Askounis, D. (2024). Depression detection in social media posts using transformer-based models and auxiliary features. *Social Network Analysis and Mining, 14*(1), 196.

Kifayathullah, M., Sekar, R., R, A., & K, V. (2025). Personalized mental health assistance: Integrating emotion prediction with GPT-based chatbot. *2025 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS)* (pp. 1−6). https://doi.org/10.1109/SCEECS64059.2025.10940203

King, C. A., Knox, M. S., Henninger, N., Nguyen, T. A., Ghaziuddin, N., Maker, A., & Hanna, G. L. (2006). Major depressive disorder in adolescents: Family psychiatric history predicts severe behavioral disinhibition. *Journal of Affective Disorders, 90*(2−3), 111−121.

Kube, T., Schwarting, R., Rozenkrantz, L., Glombiewski, J. A., & Rief, W. (2020). Distorted cognitive processes in major depression: A predictive processing perspective. *Biological Psychiatry, 87*(5), 388−398.

Leow, J. J. D., Chua, H. N., Jasser, M. B., Issa, B., & Wong, R. T. K. (2025). Comparison of depression detection between LLMs and zero-shot learning using DAD dataset. *2025 21st IEEE International Colloquium on Signal Processing & Its Applications (CSPA)* (pp.295−300). https://doi.org/10.1109/CSPA64953.2025.10933098

Liu, X. Q., Wang, X., & Zhang, H. R. (2024). Large multimodal models assist in psychiatry disorders prevention and diagnosis of students. *World Journal of*

Psychiatry, 14(10), 1415.

Long, Y., Ma, Z., Mei, Y., & Su, Z. (2024). AffirmativeAI: Towards LGBTQ+ friendly audit frameworks for large language models. *ArXiv.* https://doi.org/10.48550/arXiv.2405.04652

Lorenzoni, G., Velmovitsky, P. E., Alencar, P., & Cowan, D. (2024). GPT-4 on clinic depression assessment: An LLM-based pilot study. *2024 IEEE International Conference on Big Data (BigData)* (pp.5043−5049). https://doi.org/10.1109/BigData62323.2024.10825184

Lorge, I., Joyce, D. W., Taylor, N., Nevado-Holgado, A., Cipriani, A., & Kormilitzin, A. (2025). Detecting the clinical features of difficult-to-treat depression using synthetic data from large language models. *Computers in Biology and Medicine, 194*, 110246.

Ma, Z., Mei, Y., & Su, Z. (2023, November 11−15). Understanding the benefits and challenges of using large language model-based conversational agents for mental well-being support [Paper presentation]. *American Medical Informatics Association 2023 Annual Symposium,* New Orleans, LA, United States.

MacKenzie, L. E., Uher, R., & Pavlova, B. (2019). Cognitive performance in first-degree relatives of individuals with vs without major depressive disorder: A meta-analysis. *JAMA Psychiatry, 76*(3), 297−305.

McCoy, T. H., Castro, V. M., & Perlis, R. H. (2025). Estimating depression severity in narrative clinical notes using large language models. *Journal of Affective Disorders, 381*, 270−274.

Meng, Y., Speier, W., Ong, M. K., & Arnold, C. W. (2021). Bidirectional representation learning from transformers using multimodal electronic health record data to predict depression. *IEEE Journal of Biomedical and Health Informatics, 25*(8), 3121−3129.

Menne, F., Dörr, F., Schräder, J., Tröger, J., Habel, U., König, A., & Wagels, L. (2024). The voice of depression: Speech features as biomarkers for major depressive disorder. *BMC Psychiatry, 24*(1), 794.

Mennen, A. C., Norman, K. A., & Turk-Browne, N. B. (2019). Attentional bias in depression: Understanding mechanisms to improve training and treatment. *Current Opinion in Psychology, 29*, 266−273.

Mittermaier, M., Raza, M. M., & Kvedar, J. C. (2023). Bias in AI-based models for medical applications: Challenges and mitigation strategies. *NPJ Digital Medicine, 6*(1), 113.

Nadeem, M., Sohail, S. S., Madsen, D. Ø., Alzahrani, A. A., Del Ser, J., & Muhammad, K. (2025). A multi-modal assessment framework for comparison of specialized deep learning and general-purpose large language models. *IEEE Transactions on Big Data, 11*(3), 1001−1012.

Nushida, T., Kang, X., Matsumoto, K., Yoshida, M., & Zhou, J. (2025). An automated depression diagnosis system utilizing a knowledge base created with GPT. *2025 IEEE 17th International Conference on Computer Research and Development (ICCRD)* (pp.329−333). https://doi.org/10.1109/ICCRD64588.2025.10963224

Ogunleye, B., Sharma, H., & Shobayo, O. (2024). Sentiment informed sentence BERT-Ensemble algorithm for depression detection. *Big Data and Cognitive Computing, 8*(9), 112.

Oh, J., Kim, M., Park, H., & Oh, H. (2023). Are you depressed? Analyze user utterances to detect depressive emotions using DistilBERT. *Applied Sciences, 13*(10), 6223.

Ohse, J., Hadžić, B., Mohammed, P., Peperkorn, N., Danner, M., Yorita, A., ... Shiban, Y. (2024). Zero-Shot Strike: Testing the generalisation capabilities of out-of-the-box LLM models for depression detection. *Computer Speech & Language, 88*, 101663.

Omar, M., Soffer, S., Charney, A. W., Landi, I., Nadkarni, G. N., & Klang, E. (2024). Applications of large language models in psychiatry: A systematic review. *Frontiers in Psychiatry, 15*, 1422807.

Palmer, A., & Schwan, D. (2022). Beneficent dehumanization: Employing artificial intelligence and carebots to mitigate shame‐induced barriers to medical care. *Bioethics*, *36*(2), 187−193.

Perlis, R. H., Goldberg, J. F., Ostacher, M. J., & Schneck, C. D. (2024). Clinical decision support for bipolar depression using large language models. *Neuropsychopharmacology, 49*(9), 1412−1416.

Priyadarshana, Y. H. P. P., Liang, Z., & Piumarta, I. (2024). Transferring large language models for depression detection through multi-party conversation analysis. *2024 IEEE International Conference on E-health Networking, Application & Services (HealthCom)* (p.1). https://doi.org/10.1109/HealthCom60970.2024.10880743

Qasim, A., Mehak, G., Hussain, N., Gelbukh, A., & Sidorov, G. (2025). Detection of depression severity in social media text using transformer-based models. *Information, 16*(2), 114.

Rabie, E. M., Hashem, A. F., & Alsheref, F. K. (2025). Recognition model for major depressive disorder in Arabic user-generated content. *Beni-Suef University Journal of Basic and Applied Sciences, 14*(1), 1−16.

Raj, A., Ali, Z., Chaudhary, S., Bali, K. K., & Sharma, A. (2024). Depression detection using BERT on social media platforms. *2024 IEEE International Conference on Artificial Intelligence in Engineering and Technology (IICAIET)* (PP.228−233). https://doi.org/10.1109/IICAIET62352.2024.10730329

Rizwan, M., Mushtaq, M. F., Akram, U., Mehmood, A., Ashraf, I., & Sahelices, B. (2022). Depression classification from tweets using small deep transfer learning language models. *IEEE Access, 10*, 129176−129189.

Sabour, S., Zhang, W., Xiao, X., Zhang, Y., Zheng, Y., Wen, J., ... Huang, M. (2023). A chatbot for mental health support: Exploring the impact of Emohaa on reducing mental distress in China. *Frontiers in Digital Health, 5*,

1133987.

Sadeghi, M., Egger, B., Agahi, R., Richer, R., Capito, K., & Rupp, L. H. (2023). Exploring the capabilities of a language model-only approach for depression detection in text data. *2023 IEEE EMBS International Conference on Biomedical and Health Informatics (BHI)* (pp.1−5).

Sadeghi, M., Richer, R., Egger, B., Schindler-Gmelch, L., Rupp, L. H., Rahimi, F., ... Eskofier, B. M. (2024). Harnessing multimodal approaches for depression detection using large language models and facial expressions. *NPJ Mental Health Research, 3*(1), 66.

Saraswat, P., & Beniwal, R. (2024). BERT-based RNN for effective detection of depression with severity levels from text data. *2024 IEEE Symposium on Wireless Technology & Applications (ISWTA)* (pp.52−56). https://doi.org/10.1109/ISWTA62130.2024.10651873

Schuller, B. W. (2018). Speech emotion recognition: Two decades in a nutshell, benchmarks, and ongoing trends. *Communications of the ACM, 61*(5), 90−99.

Senn, S., Tlachac, M., Flores, R., & Rundensteiner, E. (2022). Ensembles of BERT for depression classification. *2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)* (pp. 4691−4694). https://doi.org/10.1109/EMBC48229.2022.9871120

Shah, S. M., Gillani, S. A., Baig, M. S. A., Saleem, M. A., & Siddiqui, M. H. (2025). Advancing depression detection on social media platforms through fine-tuned large language models. *Online Social Networks and Media, 46*, 100311.

Sharma, A., Lin, I. W., Miner, A. S., Atkins, D. C., & Althoff, T. (2023). Human−AI collaboration enables more empathic conversations in text-based peer-to-peer mental health support. *Nature Machine Intelligence, 5*(1), 46−57.

Shen, Z., & Paik, I. (2023). Predicting depression on Twitter with word embedding by pretrained language model. *2023 12th International Conference on Awareness Science and Technology (iCAST)* (pp.247−252). https://doi.org/10.1109/iCAST57874.2023.10359279

Shin, D., Kim, H., Lee, S., Cho, Y., & Jung, W. (2024). Using large language models to detect depression from user-generated diary text data as a novel approach in digital mental health screening: Instrument validation study. *Journal of Medical Internet Research, 26*, e54617.

Smith, K. M., Renshaw, P. F., & Bilello, J. (2013). The diagnosis of depression: Current and emerging methods. *Comprehensive Psychiatry, 54*(1), 1−6.

Sood, P. (2024). *Harnessing large language models for mental health: From sentiment analysis to depression screening* [Unpublished master's thesis]. Stevens Institute of Technology.

Stasak, B., Epps, J., & Goecke, R. (2019). Automatic depression classification based on affective read sentences: Opportunities for text-dependent analysis. *Speech Communication, 115*, 1−14.

Tao, Y., Yang, M., Shen, H., Yang, Z., Weng, Z., & Hu, B. (2023). Classifying anxiety and depression through LLMs virtual interactions: A case study with ChatGPT. *2023 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)* (pp.2259−2264). https://doi.org/10.1109/BIBM58861.2023.10385305

Taylor, N., Kormilitzin, A., Lorge, I., Nevado-Holgado, A., Cipriani, A., & Joyce, D. W. (2024). Model development for bespoke large language models for digital triage assistance in mental health care. *Artificial Intelligence in Medicine, 157*, 102988.

Teferra, B. G., Rueda, A., Pang, H., Valenzano, R., Samavi, R., Krishnan, S., & Bhat, V. (2024). Screening for depression using natural language processing: Literature review. *Interactive Journal of Medical Research, 13*(1), e55067.

Tlachac, M. L., Reisch, M., & Heinz, M. (2023). Mobile communication log time series to detect depressive symptoms. *2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)* (pp.1−4). https://doi.org/10.1109/EMBC40787.2023.10341154

Townson, S. (2023). Manage AI bias instead of trying to eliminate it. *MIT Sloan Management Review, 64*(2), 1−3.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... Polosukhin, I. (2017). Attention is all you need. *ArXiv.* https://doi.org/10.48550/arXiv.1706.03762

Verma, S., Vishal, Joshi, R. C., Dutta, M. K., Jezek, S., & Burget, R. (2023). AI-enhanced mental health diagnosis: Leveraging transformers for early detection of depression tendency in textual data. *2023 15th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT)* (pp.56−61). https://doi.org/10.1109/ICUMT61075.2023.10333301

Wang, L., & Zhang, Q. (2024). Dual-diagnostic method for depression patients based on BERT model and ViT model for audio and text analysis. *2024 IEEE 2nd International Conference on Image Processing and Computer Applications (ICIPCA)* (pp.1−6). https://doi.org/10.1109/ICIPCA61593.2024.10709143

Wang, X., Liu, K., & Wang, C. (2023). Knowledge-enhanced pre-training large language model for depression diagnosis and treatment. *2023 IEEE 9th International Conference on Cloud Computing and Intelligent Systems (CCIS)* (pp.532−536). https://doi.org/10.1109/CCIS59572.2023.10263217

Wei, Y., Guo, L., Lian, C., & Chen, J. (2023). ChatGPT: Opportunities, risks and priorities for psychiatry. *Asian Journal of Psychiatry, 90,* 103808.

Whiting, P. F., Rutjes, A. W., Westwood, M. E., Mallett, S., Deeks, J. J., Reitsma, J. B., ... QUADAS-2 Group. (2011). QUADAS-2: A revised tool for the quality assessment of diagnostic accuracy studies. *Annals of Internal Medicine, 155*(8), 529−536.

World Health Organization. (2017). *Depression and other*

common mental disorders: Global health estimates. https://coilink.org/20.500.12592/thw4fb

Wu, W., Zhang, C., & Woodland, P. C. (2023). Self-supervised representations in speech-based depression detection. *2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp.1−5). https://doi.org/10.1109/ICASSP49357.2023.10094910

Xin, C., & Zakaria, L. Q. (2024). Integrating Bert with CNN and Bilstm for explainable detection of depression in social media contents. *IEEE Access, 12*, 161203−161212.

Xu, S., Yan, Y., Ding, Y., Li, F., Zhang, S., Tang, H., ... Chen, J. (2025). Identifying psychiatric manifestations in outpatients with depression and anxiety: A large language model-based approach. *Npj Mental Health Research,4*(1), 63. https://doi.org/10.1038/s44184-025-00175-1

Yang, B., Cao, M., Zhu, X., Wang, S., Yang, C., Ni, R., & Liu, X. (2024). MMPF: Multimodal purification fusion for automatic depression detection. *IEEE Transactions on Computational Social Systems, 11*(6), 7421−7434.

Yang, K., Zhang, T., Kuang, Z., Xie, Q., Huang, J., & Ananiadou, S. (2024). MentaLLaMA: Interpretable mental health analysis on social media with large language models. *ACM Web Conference 2024*, 4489−4500.

Zhang, J., & Guo, Y. (2024). Multilevel depression status detection based on fine-grained prompt learning. *Pattern Recognition Letters, 178*, 167−173.

Zhang, L., Zhao, J., He, L., Jia, J., & Meng, X. (2023). An improved global−local fusion network for depression detection telemedicine framework. *IEEE Internet of Things Journal, 10*(22), 20230−20240.

Zhang, X., Cui, W., Wang, J., & Li, Y. (2024). Chat, summary and diagnosis: A LLM-enhanced conversational agent for interactive depression detection. *2024 4th International Conference on Industrial Automation, Robotics and Control Engineering (IARCE)* (pp.343−348).

Zhao, W. X., Zhou, K., Li, J., Tang, T., Wang, X., Hou, Y., ... Wen, J. R. (2023). A survey of large language models. *ArXiv*. https://doi.org/10.48550/arXiv.2303.18223

Zhou, X., Jin, K., Shang, Y., & Guo, G. (2018). Visually interpretable representation learning for depression recognition from facial images. *IEEE Transactions on Affective Computing*, *11*(3), 542−552.

The application of foundation models in depression screening and diagnosis

XIE Yu[1], ZHENG Hongxin[1], LIU Yizi[1], YU Honggang[2], YANG Chenghe[2]

([1] *School of Education Science, Anhui Normal University, Wuhu* 241000, *China*)

([2] *Anhui Branch of China Telecom Co., Ltd., Hefei* 230001, *China*)

Abstract: Depression is a common mental disorder that significantly impairs patients' social functioning and quality of life. In recent years, foundation models, with their powerful semantic understanding capability and multimodal data-processing capacity, have shown notable potential in the early screening and auxiliary diagnosis of depression. The construction of foundation model-based systems for depression screening and diagnosis typically involves four stages: data preprocessing, model selection, model training, and model evaluation. In these applications, foundation models primarily operate through contextualized semantic representation, attention mechanisms, multimodal behavioral capture, and predictive processing. Despite these advantages, their application still faces challenges such as algorithmic bias, insufficient diagnostic specificity, hallucination phenomena, privacy and security concerns, and ethical risks. In the future, the integration of foundation models into psychological intervention frameworks should be strengthened, with an emphasis on clinical translation pathways, in order to construct a more refined, dynamic, and culturally adaptive digital phenotype of depression, and to achieve the digital and intelligent transformation of mental health services.

Keywords: foundation models, depression, early screening, auxiliary diagnosis

附录

附表 1　大模型在抑郁症筛查和诊断中的应用

序号	研究	国家	样本量	研究对象	数据类型	基础大模型	性能评估方法
1	张冬瑜 等, 2025	中国	11431	社交媒体帖子	文本数据、图像数据	GPT-3.5-Turbo、Flan-T5	准确率、F1 分数
2	Al Masud et al., 2025	孟加拉	1602	大学生	文本数据	BERT	准确率、召回率、精确率、F1 分数
3	Bendebane et al., 2025	阿尔及利亚	26280	社交媒体帖子	文本数据	BERT	准确率、召回率、精确率、F1 分数
4	Beniwal & Saraswat, 2024	印度	10295	社交媒体帖子	文本数据、图像数据	BERT	准确率、召回率、精确率、F1 分数
5	Carstensen et al., 2024	美国	105	参与抑郁症筛查的患者	文本数据	Llama 3、Gemma 2	准确率
6	Danner et al., 2023	英国	464	心理学专业学生	文本数据、图像数据、语音数据	BERT、GPT-3.5、ChatGPT-4	精确率、召回率、F1 分数
7	Elyoseph et al., 2024b	以色列	2460	心理健康专业人士和普通大众	文本数据	ChatGPT-3.5、ChatGPT-4、Claude、Bard	单因素方差分析、最小显著差异法
8	El-Ramly et al., 2021	埃及	7000	社交媒体帖子	文本数据	BERT	准确率、精确率、召回率、F1 分数
9	Englhardt et al., 2024	美国	90	参与研究的学生	行为数据	GPT-3.5、GPT-4、PaLM 2	准确率、事实性、忠实度
10	Farruque et al., 2024	加拿大	4567	社交媒体帖子	文本数据	BERT	准确率、召回率、F1 分数
11	Gao et al., 2024	中国	9799	抑郁症患者和健康人群	基因数据	BERT	准确率、精确率、召回率、F1 分数、AUC
12	Gerczuk et al., 2023	德国	143	抑郁症患者和健康人群	语音数据、文本数据	Wav2vec	MAE、相关系数、Gini 指数
13	Gu et al., 2024	中国	1339	患有抑郁症或可能存在抑郁症状的患者	文本数据	ChatGLM3	BLEU-2、ROUGE-L、METEOR、DIST-2
14	Guo & Guo, 2024	中国	524	医生和患者的咨询对话	文本数据	EmoLLM	准确率、精确率、召回率、F1 分数
15	Gupta et al., 2024	印度	542	抑郁症患者的对话记录	语音数据	ViT	准确率、精确率、召回率、F1 分数
16	Hur et al., 2024	美国	467	线上招募被试	文本数据	GPT-3.5、GPT-4	相关性系数、RMSE
17	Jain et al., 2024	印度	13826	社交媒体帖子	文本数据	GPT-2、GPT-Neo-125M	准确率、精确率、召回率、F1 分数
18	Jarvers et al., 2024	德国	53	青少年	文本数据	BERT、ChatGPT	准确率、召回率、F1 分数
19	Jiang et al., 2024	美国	73	线上招募的被试	语音数据、视频数据、文本数据	DinoV2、WavLM、LLAMA-65B	准确率、AUC
20	Jin et al., 2024	中国	20	成年人	文本数据	ChatGPT	相关系数、AUC
21	Juarto, 2024	印尼	82715	社交媒体帖子	文本数据	BERT	精确率、召回率、F1 分数

续表

序号	研究	国家	样本量	研究对象	数据类型	基础大模型	性能评估方法
22	Kerasiotis et al., 2024	希腊	3553	社交媒体帖子	文本数据	BERT	精确率、召回率、F1 分数
23	Kifayathullah et al., 2025	印度	/	社交媒体帖子	文本数据	GPT-4o mini	准确率
24	Leow et al., 2025	马来西亚	26370	社交媒体帖子	文本数据	BERT、BART	准确率、精确率、召回率、F1 分数
25	Lorenzoni et al., 2024	加拿大	189	临床访谈文本	文本数据	GPT-4	准确率、精确率、召回率、F1 分数
26	Lorge et al., 2025	英国	100	成年抑郁症患者	文本数据	GPT-3.5、BERT	精确率、召回率、F1 分数
27	McCoy et al., 2025	美国	15000	电子病例	文本数据	GPT-4o	相关系数、召回率、特异性
28	Meng et al., 2021	美国	43967	电子病例	文本数据	BRLTM	AUC
29	Nadeem et al., 2025	印度	232047	社交媒体帖子	文本数据	GPT-3.5、GPT-4、Google Bard	准确率、精确率、召回率、F1 分数
30	Nushida et al., 2025	日本	60	社交媒体帖子	文本数据	GPT-4o	准确率、精确率、召回率、F1 分数
31	Ogunleye et al., 2024	英国	13804	社交媒体帖子	文本数据	BERT	准确率、精确率、召回率、F1 分数
32	Oh et al., 2023	韩国	140467	社交媒体帖子	文本数据	BERT	准确率、精确率、召回率、特异性、F1 分数
33	Ohse et al., 2024	德国	82	参与访谈的被试	文本数据	BERT Llama2-13B GPT-3.5、GPT-4	精确率、召回率、特异性、F1 分数
34	Priyadarshana et al., 2024	日本	/	对话文本和社交媒体帖子	文本数据	Llama、GPT-3、GPT-4	召回率
35	Qasim et al., 2025	墨西哥	24000	社交媒体帖子	文本数据	BERT	精确率、召回率、F1 分数
36	Rabie et al., 2025	埃及	5500	社交媒体帖子	文本数据	BERT	准确率、精确率、召回率和 F1 分数
37	Raj et al., 2024	斐济	7732	社交媒体帖子	文本数据	BERT	准确率、精确率、召回率、F1 分数、AUC
38	Rizwan et al., 2022	巴基斯坦	73355	社交媒体帖子	文本数据	ESG、ESD、XDL、ABV	准确率、精确度、召回率、特异性、F1 分数
39	Sadeghi et al., 2023	德国	275	参与访谈的个体	文本数据	GPT-3.5-Turbo、DepRoBERTa	RMSE、MAE
40	Sadeghi et al., 2024	德国	275	参与访谈的个体	文本数据	GPT-3.5-Turbo、DepRoBERTa	RMSE、MAE
41	Saraswat & Beniwal, 2024	印度	16632	社交媒体帖子	文本数据	BERT、LSTM、GRU	准确率、召回率、精确率、F1 分数
42	Senn et al., 2022	美国	189	参与访谈的个体	文本数据	BERT	准确率、精确率、召回率、F1 分数
43	Shah et al., 2025	巴基斯坦	40000	社交媒体帖子	文本数据	GPT-3.5、LLaMA2	准确率、召回率、精确率、F1 分数
44	Shen & Paik, 2023	日本	1600000	推特用户	文本数据	BERT, CNN, LSTM,	准确率, 精确率, 召回率, F1 分数

续表

序号	研究	国家	样本量	研究对象	数据类型	基础大模型	性能评估方法
45	Shin et al., 2024	韩国	91	写日记的 APP 用户	文本数据	GPT-3.5 和 GPT-4	准确率、召回率、精确率、F1 分数、特异性
46	Sood, 2024	美国	1415	参与访谈的个体	文本数据	BERT	精确率、召回率、灵敏度和 F1 分数
47	Tao et al., 2023	中国	139	抑郁症与焦虑症患者	语音数据、文本数据	ChatGPT	准确率、召回率、精确率、F1 分数
48	Verma et al., 2023	印度	35622	社交媒体帖子	文本数据	BERT	准确率、召回率、精确率、F1 分数
49	Wang & Zhang, 2024	中国	/	抑郁症患者	文本数据、图像数据	BERT、ViT	准确率、精确率、召回率、F1 分数
50	Wang et al., 2023	中国	/	抑郁症患者	文本数据	BERT	安全性、可用性、流畅性
51	Xin & Zakaria, 2024	马来西亚	46022	社交媒体用户及心理健康语料库用户	文本数据	BERT	准确率、召回率、精确率和 F1 分数
55	Xu et al., 2025	中国	1160	门诊患者	语音数据、文本数据	Qwen2	准确率、精确率、召回率、F1 分数、AUC
53	Yang, Cao, et al., 2024	中国	189	临床访谈的受试者	语音数据、文本数据	EfficientNet-B7、BiLSTM	F1 值、准确率、精确率和召回率
54	Zhang et al., 2024	中国	1339	抑郁症患者的对话记录	文本数据	ChatGLM-6B	BLEU-2 、 ROUGE-L 、METEOR、DIST-2、准确率、召回率、F1 分数
55	Zhang & Guo, 2024	中国	189	临床访谈的受试者	文本数据	T5、BERT	准确率、精确率、召回率、F1 分数、MAE