

大学生心理压力智能评估:基于融合文本与影像的多模态模型的设计及验证

谢辉荣, 胡潮滨, 梁国华, 韩红喆, 黄牧, 冯前进
南方医科大学生物医学工程学院, 广东 广州 510515

摘要:目的 提出一种融合社交媒体文本与图像信息的多模态大学生心理压力自动评估模型,以提升评估的客观性与准确性,服务高校心理健康智能化建设。方法 基于深度学习技术,设计包含文本情绪建模、图像情绪建模及多模态融合预测3大模块的评估框架。通过Bi-LSTM提取文本情感特征、U-Net提取图像语义线索,并采用特征拼接策略实现跨模态语义协同,完成轻度、中度、重度3类心理压力的自动识别。研究以广东省多所高校1577名学生的社交平台数据为基础构建多模态标注数据集,清洗后随机抽取252个样本用于模型训练与测试。结果 在3分类任务中,模型在测试集上表现出色,准确率达92.86%,F1分数为0.9276,展现出良好的稳定性与一致性。混淆矩阵分析亦显示模型能有效区分不同压力等级。结论 本研究所构建的多模态心理压力评估模型有效融合了非结构化社交行为数据,提升了心理状态识别的科学性与实用性,为智能心理服务系统建设提供了理论支持和技术路径。
关键词:大学生心理压力;多模态数据融合;自动评估;深度学习;社交媒体情感分析

Design and validation of a multimodal model integrating text and imaging data for intelligent assessment of psychological stress in college students

XIE Huirong, HU Chaobin, LIANG Guohua, HAN Hongzhe, HUANG Mu, FENG Qianjin
School of Biomedical Engineering, Southern Medical University, Guangzhou 510515, China

Abstract: Objective We propose a multimodal model integrating social media text and image data for automated assessment of psychological stress in college students to support the development of intelligent mental health services in higher education institutions. **Methods** Based on deep learning technology, we designed an evaluation framework comprising a text sentiment modeling module, an image sentiment modeling module, and a multimodal fusion prediction module. Text sentiment features were extracted using Bi-LSTM, and image semantic cues were extracted via U-Net. A feature concatenation strategy was used to enable cross-modal semantic collaboration to achieve automatic identification of 3 psychological stress levels: mild, moderate, and severe. We constructed a multimodal annotated dataset using social platform data from 1577 students across multiple universities in Guangdong Province. After data cleaning, 252 samples were randomly selected for model training and testing. **Results** In the 3-classification task, the model demonstrated outstanding performance on the test set, and achieved an accuracy of 92.86% and an F1 score of 0.9276, exhibiting excellent stability and consistency. Confusion matrix analysis further revealed the model's ability to effectively distinguish between different pressure levels. **Conclusion** The multimodal psychological stress assessment model developed in this study effectively integrates unstructured social behavior data to enhance the scientific rigor and practical applicability of psychological state recognition, and thus provides support for developing intelligent psychological service systems.

Keywords: college students' psychological stress; multimodal data fusion; automated assessment; deep learning; social media sentiment analysis

学生的心理健康是我国教育发展战略的核心议题之一,不仅是衡量教育现代化与人才培养质量的关键指标,还是教育强国建设的重要基石。为了切实加强学生心理健康工作,构建高质量的学生心理健康服务体系,近年来国家密集出台相关政策,从教育部等十七部门联合印发《全面加强和改进新时代学生心理健康工作专项行动计划(2023~2025年)》指出要构建“筛查预警、健康教育、咨询服务、干预处置”的四位一体的心理健康工作体系,到中共中央、国务院印发的《教育强国建设规划纲要

(2024~2035年)》明确要求“普及心理健康教育,建立全国学生心理健康监测预警系统”,均对提升学生心理健康工作的科学化、精准化和体系化水平的提出了明确的要求。

大学生群体处在生理发育、心理成熟以及社会适应能力构建的关键时期,心理健康状况不仅会影响个体的成长与幸福感,还直接关联到高素质人才培养的质量以及国家未来的创新与发展潜力。因此,高校大学生心理健康工作显得十分重要。大学生正处于从青春期末向成年期过渡的关键人生阶段,其世界观、人生观以及价值观尚未完全定型,心理层面有高度可塑性与脆弱性。这使得他们在面对各种挑战时更易受到影响。随着高等教育的普及化以及就业形势变得日益严峻,大学生普遍面临学业、就业、人际关系等多种压力的冲击。这些压力影响他们的心理健康,还可能对其学业表现和未来

收稿日期:2025-07-27

基金项目:国家自然科学基金(52305023);广东省自然科学基金(2024A1515011979)

Supported by National Natural Science Foundation of China (52305023).

作者简介:谢辉荣,硕士,讲师,E-mail: 450282452@qq.com

通信作者:冯前进,博士,教授,E-mail: 1271992826@qq.com

职业发展产生长期影响。值得关注的是,当多重压力持续作用且缺少有效疏导时,大学生大多时候会陷入慢性应激状态,从而引发一系列心理健康问题。现有研究显示,长期处于高压状态会削弱个体的认知功能与情绪调节能力,还会降低学习效率与社会适应能力,容易诱发焦虑、抑郁等心理障碍,甚至导致自伤自杀等极端行为出现^[1]。关注大学生心理健康,及时识别并有效干预其心理压力状态,是高校心理健康教育的关键要点,更是落实“立德树人”根本任务的必然要求^[2-4]。

当下高校广泛采用的评估方法主要依靠以标准化量表为核心的传统模式,虽说操作简便,然而存在一定的局限性。在实际操作方面,量表仅能呈现特定时间点的心理状态,没办法实施实时动态监测预警,评估结果有滞后特性,难以契合“体系化、常态化”的评估要求。其数据来源较为单一,很容易受到社会期许效应以及主观认知偏差的干扰,并且高度依赖人工分析致使成本较高,难以支持高校开展大规模、常态化的精准筛查^[5]。在分析深度上,它无法达到“科学化、精准化”的评估标准,传统量表的分析框架着重于症状的统计学描述,而非心理机制的解剖,这致使评估结果大多时候只能对群体进行粗略的风险划分,缺少对个体心理问题深层病理机制的剖析,无法达成精准有效的分型分级以及个性化的干预方案^[6]。这种“测不准、测不全、测不深”的方法困境,成为限制当前大学生心理健康工作推进的最大障碍。

互联网的普遍推广以及社交媒体的大量应用,给解决上述问题给予了全新的思考方向。据中国互联网络信息中心所发布的第54次报告说明,我国大学生当中,日均上网时长超过2h的比例达到了15.3%。《青少年互联网使用情况调查报告》则指出,社交媒体已然成为大学生用以记录生活、表达自身观点以及进行社交互动的主要平台,其产生的如文本、语音、视频等行为数据,可动态地呈现出大学生的心理状态。而这些非结构化的网络内容,为借助数据分析来进行心理压力评估提供了新的数据源头以及分析渠道^[7]。

随着人工智能以及大数据技术持续发展,心理健康评估领域开拓出了全新路径。运用机器学习技术与自然语言处理技术对个体心理健康状况给予建模,呈现出更为出色的效能。Ophir等^[7]依靠人工神经网络模型,依据Facebook的8万多条社交文本信息,提出了一种创新的自杀风险语言自动评估方法。夏先益^[8]以及张风云^[9]分别借助文本挖掘和ON-LSTM技术,也达成了心理健康评估与情感的自动分类。随着深度学习技术的进步,基于深度神经网络的方法,构建图像与图像描述之间的映射关系,识别个体心理状态取得了一定进展,借助剖析视频图像中的面部特征点与微表情,成功实现了抑郁症识别^[10]。然而单一模态数据难以全面捕捉被评估者的动态心理状态,容易致使分析结果出现偏差。

典型的“强颜欢笑”现象,就指出情绪表达的跨模态不一致性,而融合文本、图像/表情、语音等多模态信息,可更全面且准确地挖掘个体的真实情绪与心理状态^[11]。随后,章荪等^[12]提出的时序多模态情感分析模型,以及Zhu等^[13]结合深度学习与多模态数据构建的大学生动态心理健康评价模型,均呈现出了巨大潜力。尽管多模态数据融合在心理健康评估领域呈现出良好的应用前景,不过在大学生心理压力识别中,针对文本与图像等异质模态之间的语义协同机制建模仍存在不足^[14]。现有方法在特征融合的稳定性和表达能力方面仍需要提升,怎样高效整合多模态信息以提高评估的准确性与泛化能力,依旧是当前急需攻克的难题。

为应对诸多挑战,本文提出设计一个大学生心理压力自动评估模型。该模型融合文本与图像模态,借助深度学习提取社交媒体文本及图像深层特征,达成跨模态语义协同,从而突破单模态局限,实现压力状态隐性识别与自动分类。希望通过这种方式响应国家心理健康智能监测体系建设的需求,为高校提供及时且精准的大学生心理监测技术支撑,解决传统评估“被动滞后”核心难题,服务健康中国2030与教育强国建设人才培养目标。

1 资料和方法

1.1 资料

本研究以广东省多所高校在校大学生为实验对象,采用“问卷+表达任务”的两阶段采集模式,旨在模拟大学生在真实社交媒体环境中的自我表达行为。所有被试在参与研究前均签署知情同意书,并授权研究团队在遵循数据匿名化与隐私保护原则的前提下,采集其公开的网络行为数据。具体而言,受试者在匿名完成压力评估问卷后,需撰写一段“最符合当前情绪状态、且最可能发布于社交平台的文字”,并可自愿上传一张“最能代表此刻心情”的图像。已有研究表明^[15-17],个体在不同心理压力水平下,其社交媒体内容在情绪语义与视觉特征上存在显著差异。因此,该设计不仅增强了数据的生态效度,也为多模态心理建模提供了更具行为真实性的基础。

1.2 心理压力评估

心理压力,也就是心理应激,是个体在遭遇外部挑战、需求或者威胁之时所产生的一种带有紧张感或者不安感的心理反应。在大学生这个群体当中,心理压力呈现出高度的普遍性以及阶段性特点,大学生正处在成长的过渡阶段,很容易受到学业、人际交往、就业等诸多外部因素的影响,一旦压力超过了个体的心理韧性阈值,就有可能引发焦虑、抑郁等负面情绪反应,甚至会对其正常的学习和生活造成影响。

传统的心理压力评估主要依靠知觉压力量表以及抑郁-焦虑-压力量表等自评量表。这种评估方式虽说方便专业人员对心理状态做初步诊断,不过因为指标单

一、效率不高、时效性欠佳等存在测量误差,很难达成大规模、常态化的持续监测^[5]。随着人工智能不断发展,基于多模态数据的心理压力评估方法渐渐受到研究者的关注。这种方法凭借整合文本、语音、图像、生理信号等多种信息来源,可更加全面地呈现个体的心理状态,有较高的客观性以及应用潜力。然而多模态数据的处理也面临着不少挑战,如数据多源、特征维度复杂、融合策略不明确等问题。针对这些问题,本文设计了一种基于社交媒体多模态数据的压力评估模型。借助采集大学生群体在社交平台上的文本和图像信息,运用深度学习技术提取其语义特征与情感表达模式,构建心理压力等级自动识别模型,达成心理状态的持续监测^[18, 19]。

1.3 数据分类及处理

依据《大学生压力评估问卷》^[20-22]的总体得分(满分120分)分布,采用三分位法设定压力等级的分界阈值(表1),具体而言,得分 ≤ 45 为无压力或轻度压力(Stress Level 0),46~65为中度压力(Stress Level 1), ≥ 66 为重度压力(Stress Level 2)。该分级方式基于样本得分的整体分布特征,兼顾统计学合理性与心理学解释性。需要说明的是,在心理学理论框架中,“无压力”与“轻度压力”均属于个体可通过自我调节维持心理健康的范围,对学习与生活的影响较小,因此在建模分析阶



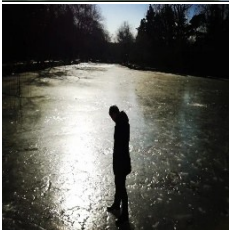
段将二者合并为同一等级(Stress Level 0),以增强模型的稳定性与可解释性。本研究共回收问卷1577份,经多轮筛选后,最终保留1216份有效样本,有效率为77.1%。

数据筛选流程包括3个步骤:首先,问卷中嵌入隐性验证题项,用于识别随意填答或应付式响应,若受试者在该题项上出现明显逻辑错误,则该问卷被自动标记为无效;其次,系统自动剔除作答时间过短(< 120 s)、选项重复率过高($> 80\%$)或存在明显逻辑矛盾的问卷;最后,由两名具备心理测评背景的研究人员对疑似异常样本进行二次审阅,确保数据的真实性与一致性。为进一步提升数据可靠性,我们从最终样本中随机抽取252例(训练集210例,测试集42例)用于模型训练与验证,确保研究结果具有代表性与可推广性。

再者,图像数据的形态是“社交媒体公开图像”而非“实验室标准人脸”,我们在视觉建模路径上刻意回避了依赖面部关键点或微表情的传统情感计算范式,转而采用轻量级 U-Net 网络框架提取全局语义与局部情绪线索的耦合特征,模型始终与“非受控、非生理、非高清”的图像属性保持同频,重点分析“情绪表达特征”而非“生理结构特征”,既避免了因采集规格不一带来的域偏移,也规避了高精度人脸识别可能引发的隐私争议。

表1 数据标注示例

Tab.1 Instances of annotation of text and image data from social media

Image	Text	Label
	毫无压力/一切发生皆有利于我/开心就好/一点点压力,正好够推动我向前/向往自由的生活,与家人一起出去旅游,好好地感受世界	Stress Level 0
	都会好起来的/迷茫中.../燃尽/顺其自然/压力就是动力!/最近课好多好忙好累压力好大啊/有点累/都会好起来的	Stress Level 1
	弃学中, 误Q/迷茫 没有动力 悲观 低迷/压力好大/生活好难, 我好烦/生活给我一巴掌 我说没有上次响	Stress Level 2

经过数据清洗后,采用随机抽样共取252个样本。其中抽取210个样本数据用于模型的训练,42个独立样本用于模型测试。训练集与测试集的比例为5:1,且保持与总体一致的压力等级分布。这表明该数据集的构

建为模型开发与泛化能力验证提供了有效支撑。为确保模型输入的数据合理性与统一性,首先对原始社交数据进行系统性的清洗和预处理,具体处理流程如下:(1) 问卷样本筛选:剔除压力评估量表中填写不规范者,如作

答时间过短、答案较为极端等样本;(2)用户内容筛选:剔除在研究周期内发布信息量过少或信息密度不足的用户数据;(3)模态数据预处理:针对文本与图像两类模态数据分别进行标准化处理。文本部分删除HTML标签、停用词、无效符号,统一编码与长度(padding至100 tokens),再将词语映射为索引形式以便模型输入。图部分统一图像尺寸为 256×256 ,并将像素值归一化至 $[0,1]$ 区间。最终,文本数据构造成整数张量 $X_{text} \in Z^{B \times T}$,图像数据构造成浮点张量 $X_{img} \in R^{B \times 3 \times 256 \times 256}$,其中 B 为batch大小, T 为最大文本长度。

1.4 多模态融合技术

随着多模态融合技术持续发展,心理学研究者开始尝试把多模态融合技术运用到心理压力评估。目前在心理压力评估应用里经常使用的多模态融合技术包含特征融合和决策融合这两种模式。多模态融合技术指把来自不一样传感器以及信息源的数据开展整合^[23],以此来提取更为全面且准确的信息,最终达成对个体心理压力的综合评估。大学生于社交媒体平台内,借助如动态、评论、短语等文本,以及图像,如自拍、配图、截图等,来表达观点、分享生活,这为心理压力识别提供了真实的数据源。

然而,文本、图像这类单一模态信息存在一定局限性,有可能出现语义模糊、缺少情境解释性等问题。凭借对文本与图像信息进行跨模态关联建模,可有效利用模态之间的互补优势,提高心理状态识别的准确度以及鲁棒性^[24]。这种融合机制已然成为当前多模态情感计算以及心理建模领域的关键研究方向^[25,26]。

1.5 深度学习建模基础

在心理压力评估这个领域当中深度学习给多模态融合提供了可靠的技术方面的支持^[27,28]。在开展文本模态特征提取工作的时候一般会采用跟循环神经网络相关的模型来达成文中情绪倾向和语义的准确关联,例如双向长短时记忆网络提取序列特征。对于图像模态而言,大多采用卷积神经网络,比如VGG、ResNet来提取局部视觉特征以及全局情感线索^[29]。此外可在多模态融合阶段运用注意力机制、特征拼接、权重加权等多种不同方式来形成统一的心理状态表征。

本研究提出的自动评估模型是依据上述深度学习框架设计的,分别针对文本和图像模态进行特征提取以及情感分类,并且采用最大值融合策略综合判断个体的心理压力水平以此实现高效且无创的心理状态智能识别。

1.6 方法

为实现对大学生社交行为数据中隐含的心理压力状态的智能化评估,本研究构建一种基于文本与图像多模态融合的学生心理压力自动分类模型。该模型采用深度学习网络对社交平台数据进行展示性特征提取和情感编码,将文本与图像中的情绪线索进行一致化融

合,从而实现对心理压力级别的智能化评估。全模型由3大功能模块组成:文本情绪建模,图像情绪建模,以及多模态融合与心理状态预测构建(图1)。

1.6.1 U-Net式深层卷积网络 图像数据中包含了情绪表达的观察信息,包括色调、槽构、平面布局等观察量。为精准抽取图像情绪特征。因此,本研究将图像内容作为另一模态的关键输入,提取其潜在的压力表达模式。针对图像模态,采用U-net架构(图1A),提取高层次语义特征^[30-33]。

a. 多尺度下采样模块

定义下采样模块为两层 3×3 卷积+BN+LeakyReLU:

$$x^{(l)} = \sigma(BN(Conv_3(\sigma(BN(Conv_3(x^{(l-1)})))))) \quad (1)$$

其中,激活函数为 $\sigma(x) = \max(0.01x, x)$ 。每经过一次pooling降维。通道加倍。

b. Bottom 模块:非对称卷积建模上下文

为增强图像纵向信息整合能力,引入非对称卷积层以捕获上下文语义:

$$f_{context} = Conv_{10 \times 3}(x) \quad (2)$$

该结构有助于挖掘心理表达相关的区域性与结构性模式。最终输出特征经展平后得到一维向量 $z_{img} \in R^C$,其中 $C \approx 4096$ 。

1.6.2 双向LSTM编码器 文本信息常蕴含个体心理状态的显性与隐性线索,尤其在社交平台环境下,文字往往体现了当下的心理负荷水平。因此,本研究设计了一套面向“压力表达”的文本建模子系统。为了建模文本的情绪表达特征,采用词嵌入层+双向长短时记忆网络(Bi-LSTM)结构^[34-36],充分挖掘上下文情绪语义(图1B)。

a. 词嵌入表示

设输入文本索引序列为 $x_i \in Z^{B \times T}$,词嵌入层参数 $E \in Z^{B \times T}$,其中 V 为词汇表大小, d 为嵌入维度。经词嵌入层映射后,文本序列表示为:

$$e_i = Embed(x_i) \in R^{B \times T \times d} \quad (3)$$

b. Bi-LSTM编码

使用双向LSTM对序列进行建模,其前向与后向的隐状态分别为 \vec{h}_i 和 \overleftarrow{h}_i ,合并输出为:

$$h_i = \vec{h}_i \oplus \overleftarrow{h}_i \in R^{2H}, \quad (4)$$

选取最后时刻的合并状态 h_T 作为全局特征,并映射为统一维度 H :

$$z_{text} = W_{text} \cdot h_T + b_{text}, \quad (5)$$

其中,向量 $z_{text} \in R^H$ 表示文本情绪特征。

1.6.3 多模态融合与心理状态预测 将文本特征 $z_{text} \in R^H$ 与图像特征 $z_{img} \in R^C$ 拼接融合^[37](图1C):

$$z_{fusion} = z_{text} \oplus z_{img} \in R^{H+C} \quad (6)$$

融合特征经两层全连接网络(MLP)进行分类,输出心理健康状态的类别概率:

$$\hat{y} = \text{Softmax}(W_2 \cdot \text{ReLU}(W_1 \cdot z_{fusion} + b_1) + b_2) \quad (7)$$

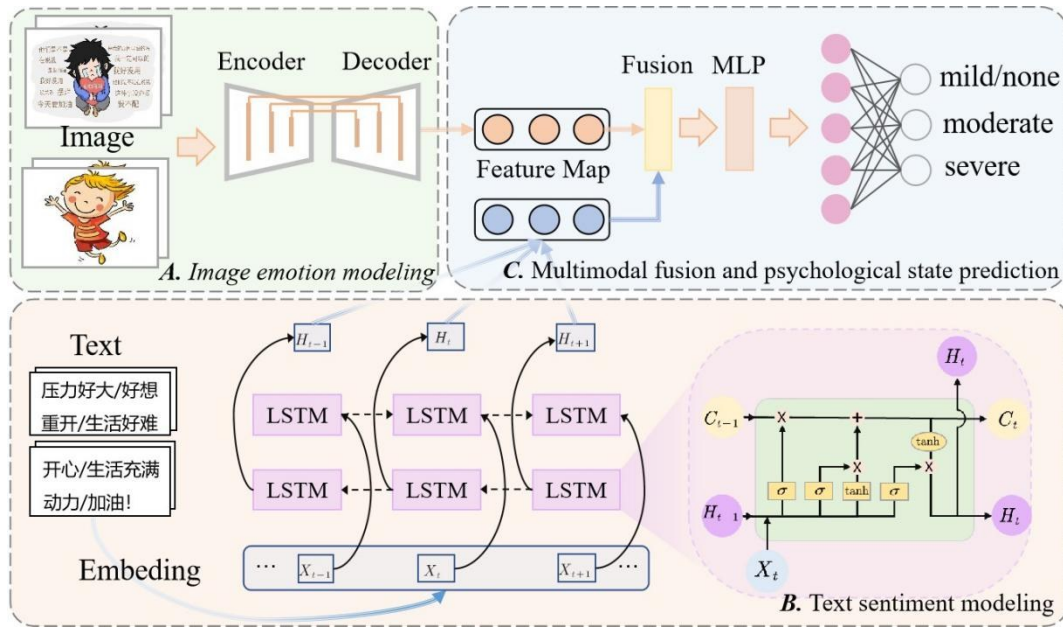


图1 网络流程图

Fig. 1 Network flowchart. A: U-Net image encoder. B: Bi-LSTM text encoder. C: Multimodal fusion and psychological-state prediction.

损失函数采用交叉熵定义为:

$$L = -\sum_{i=1}^K \log(\hat{y}_i) \quad (8)$$

其中 K 为类别数, 本文为心理健康三分类 (Stress Level 0, Stress Level 1, Stress Level 2)。

1.7 评价指标

为了全面评估多模态模型的性能, 本实验采用了4种常见的定量评价指标: 准确率 (Acc)、精确率 (Pre)、召回率 (Rec) 和 F1 分数 (F1)。

Acc 衡量整体分类预测的正确率:

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \quad (9)$$

其中, TP 为真阳性, TN 为真阴性, FP 为假阳性, FN 为假阴性。

Pre 衡量模型预测为某类样本中实际为正例的比例:

$$Pre = \frac{TP}{TP + FP} \quad (10)$$

Rec 衡量模型对正类样本的识别能力:

$$Rec = \frac{TP}{TP + FN} \quad (11)$$

F1-score 是精确率与召回率的调和平均, 综合反映模型精度与覆盖率:

$$F1 = 2 \cdot \frac{Pre \cdot Rec}{Pre + Rec} \quad (12)$$

1.8 实验设置

实验基于 Python 3.8、PyTorch 2.1 作为框架, 使用 NVIDIA GeForce GTX 2080Ti GPU 的 Linux 服务器。使用 Adam 优化器通过随机梯度下降对网络训练 400 个 epoch, 初始学习率为 0.001。

2 结果

2.1 多模态融合实验结果

在本研究所设定的三分类任务中, 所提出的压力自动评估模型在测试集上取得了显著的性能表现。Acc、Pre、Rec 和 F1-score 四项指标上分别达到了 0.9286、0.9282、0.9286 和 0.9276, 表现出高度均衡的分类能力。

为更直观地展现各评价维度的综合性能绘制雷达图。各项指标数值集中于 0.93 附近, 反映出模型在分类任务中的稳定性与一致性 (图 2)。不同于某些模型存在“偏科”现象 (如高精度低召回), 本模型兼顾了查准率与查全率, 体现了良好的实用价值。

测试集的混淆矩阵可以直观展示该模型在各类别上的识别效果。具体而言, 对于 Stress Level 0 共有 13 个。其中 11 个样本能被模型精准识别, 然而 1 个样本被误判为 Stress Level 1, 还有 1 个 Stress Level 2。这些被误判样本的文本或图像内容, 大多呈现出模糊情绪特征, 例如“向往自由”“最近有点累”这类表述。这可能是表达不充分、语义模糊等原因导致模型判断出现偏差 (图 3)。通过分类出来的 Stress Level 1 共有 15 个, 有 1 个样本被误分类为 Stress Level 0, 其余显示正确分类。这表明该模型说明模型对 Stress Level 1 这一情绪区间的判别能力较强, 识别准确率较高。被准确分类为 Stress Level 2 的则有 14 个样本, 并且没有出现任何误判断的情况。这一结果凸显了模型在高压情绪状态识别中的高度敏感性, 同时也说明强烈的负面情绪表达在多模态特征空间中具有较好的可区分性。

此外, 从模型分类表现来看, 三类压力之间的误

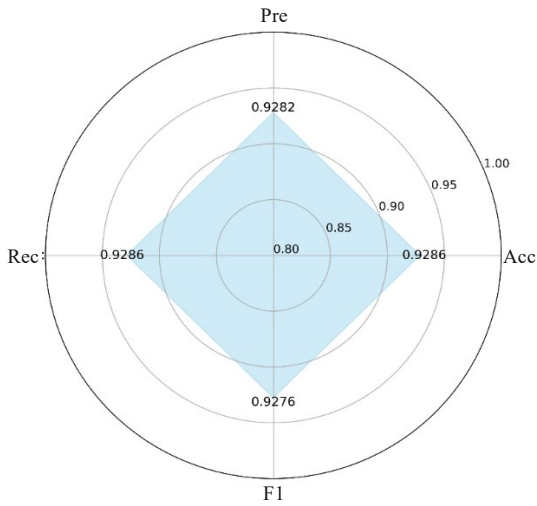


图2 模型测试结果雷达图
Fig.2 Radar chart of model test results.

	Predicted 0	Predicted 1	Predicted 2
True 0	11	1	1
True 1	1	14	0
True 2	0	0	14

图3 测试结果混淆矩阵
Fig.3 Confusion matrix of the test result.

判现象主要集中在 Stress Level 0 与 Stress Level 1 这两个类别之间。这在一定程度上说明情绪过渡存在模糊区域。大学生的心理压力状态并非是单一、割裂的,而是具有一定的连续性。在心理状态识别领域,边界模糊是一个普遍面临而具有挑战性的难题。结合真实语境,该模型在 Stress Level 0 与 Stress Level 1 间容易出现轻微交叉误判的情况。这一现象也反映了大学生心理状态具有复杂性和主观性的特点。因此,心理工作者可以重点关注处于这一压力过渡区间的群体,提前做好预防干预工作。

该模型在小样本测试集下仍然表现出卓越的泛化能力,得益于多模态融合机制在特征层对情感表征的补充和强化。实验结果充分验证了本文所提出的融合架构在心理压力评估任务中的有效性与稳健性,也为高校智能心理支持系统的构建提供了技术路径与经验支撑。

3 讨论

本研究构建了一种融合社交媒体文本与图像信息的多模态心理压力评估模型,并在真实高校人群数据中进行验证。结果显示,该模型在三分类心理压力识别任

务中准确率达92.86%,F1分数为0.9276,尤其在识别中高压状态方面表现出良好稳定性与区分能力。该结果初步证实,融合文本情绪与图像语义信息可有效提升心理压力评估的客观性与准确性,为高校心理健康智能监测提供了新的技术路径。

本研究在以下方面具有一定特色与优势。首先,在数据层面,本研究基于真实社交媒体行为数据构建多模态心理标注数据集,突破了传统自陈量表“静态、主观、延迟”的局限,体现出更强的生态效度与时效性。其次,在建模策略上,采用Bi-LSTM与U-Net分别提取文本与图像深层语义特征,并通过特征拼接实现跨模态融合,兼顾了模型的可解释性与计算效率,区别于以往依赖黑箱式端到端训练的多模态模型^[13, 14]。此外,本研究在样本量有限条件下仍表现出良好泛化能力,提示融合策略在提升小样本学习效率方面具有潜力。

从机制角度看,文本与图像作为社交媒体中最常见的表达载体,分别承载了个体显性(如情绪词、压力描述)与隐性(如色调、构图、场景)心理线索。通过深度学习模型对二者进行协同建模,有助于捕捉个体情绪表达的多维特征,提升对心理状态连续性与复杂性的识别能力。这一思路与当前多模态情感计算领域“语义+感知”融合的发展趋势相一致^[11, 18],也为理解大学生心理压力表达的跨模态特征提供了数据支持。

然而,本研究仍存在一定局限性。首先,样本来源集中于广东省高校,地域与文化背景相对单一,模型的跨地区、跨人群泛化能力尚需在更广泛样本中进一步验证。其次,部分弱表达样本(如情绪模糊、图文不一致)存在误判现象,提示模型在细粒度情绪识别与边界模糊状态区分方面仍有提升空间。再次,尽管模型在分类性能上表现良好,但其内部机制仍较为复杂,缺乏对关键特征的显式解释,未来应引入可解释性技术(如注意力可视化、SHAP分析)提升模型透明度与临床可用性。最后,本研究尚未对模型在实际心理干预场景中的适用性与用户接受度进行评估,后续应结合高校心理咨询系统开展部署验证与用户体验研究。

综上所述,本研究提出的多模态心理压力评估模型在高校社交媒体数据中展现出良好识别性能,初步验证了融合文本与图像信息进行心理状态自动识别的可行性。未来研究应进一步拓展样本来源、引入时序建模与因果推断机制,并加强与高校心理健康服务系统的集成应用,推动该技术从“模型验证”走向“系统落地”,为构建全国学生心理健康监测预警体系提供技术支撑与实践经验。

Declaration of interests: The authors declare no competing interests.

参考文献:

[1] Fan YY, Liu J, Zeng YY, et al. Factors associated with non-suicidal self-injury in Chinese adolescents: a meta-analysis[J]. Front

- Psychiatry, 2021, 12: 747031.
- [2] Huang JP, Nigatu YT, Smail-Crevier R, et al. Interventions for common mental health problems among university and college students: a systematic review and meta-analysis of randomized controlled trials[J]. *J Psychiatr Res*, 2018, 107: 1-10.
- [3] Bruffaerts R, Mortier P, Kiekens G, et al. Mental health problems in college freshmen: Prevalence and academic functioning[J]. *J Affect Disord*, 2018, 225: 97-103.
- [4] Cage E, Jones E, Ryan G, et al. Student mental health and transitions into, through and out of university: student and staff perspectives [J]. *J Furth High Educ*, 2021, 45(8): 1076-89.
- [5] Pedrelli P, Nyer M, Yeung A, et al. College students: mental health problems and treatment considerations[J]. *Acad Psychiatry*, 2015, 39(5): 503-11.
- [6] Hickey BA, Chalmers T, Newton P, et al. Smart devices and wearable technologies to detect and monitor mental health conditions and stress: a systematic review[J]. *Sensors (Basel)*, 2021, 21(10): 3461.
- [7] Ophir Y, Tikochinski R, Asterhan CSC, et al. Deep neural networks detect suicide risk from textual facebook posts[J]. *Sci Rep*, 2020, 10(1): 16685.
- [8] 夏先益. 基于文本挖掘的在线论坛用户心理健康自动评估[D]. 南昌: 江西财经大学, 2019.
- [9] 张凤云. 基于ON-LSTM的文本情绪分析方法研究[D]. 郑州: 郑州大学, 2020.
- [10] Cole EJ, Phillips AL, Bentzley BS, et al. Stanford neuromodulation therapy (SNT): a double-blind randomized controlled trial[J]. *Am J Psychiatry*, 2022, 179(2): 132-41.
- [11] Baker J, Ngo H, Efthimiou TN, et al. Electrical stimulation of smiling muscles reduces visual processing load and enhances happiness perception in neutral faces[J]. *Commun Psychol*, 2025, 3(1): 94.
- [12] 章 菀, 尹春勇. 基于多任务学习的时序多模态情感分析模型[J]. *计算机应用*, 2021, 41(6): 1631-9.
- [13] Zhu QJ, Xiong JC, Peng LL. College students' mental health evaluation model based on tensor fusion network with multimodal data during the COVID-19 pandemic[J]. *Biotechnol Genet Eng Rev*, 2024, 40(3): 1821-35.
- [14] Fu ZW, Liu F, Xu Q, et al. LMR-CBT: learning modality-fused representations with CB-Transformer for multimodal emotion recognition from unaligned multimodal sequences[J]. *Front Comput Sci*, 2023, 18(4): 184314.
- [15] Kalinkara Y, Talan T. Psychological balances in the digital world: dynamic relationships among social media addiction, depression, anxiety, academic self-efficacy, general belongingness, and life satisfaction[J]. *J Genet Psychol*, 2025, 186(2): 85-113.
- [16] Chandra Guntuku S, Buffone A, Jaidka K, et al. Understanding and measuring psychological stress using social media[J]. *Proc Int AAAI Conf Web Soc Medium*, 2019, 13: 214-25.
- [17] Wolfers LN, Utz S. Social media use, stress, and coping[J]. *Curr Opin Psychol*, 2022, 45: 101305.
- [18] Khoo LS, Lim MK, Chong CY, et al. Machine learning for multimodal mental health detection: a systematic review of passive sensing approaches[J]. *Sensors (Basel)*, 2024, 24(2): 348.
- [19] Du C, Liu C, Balamurugan P, et al. Deep learning-based mental health monitoring scheme for college students using convolutional neural network[J]. *Int J Artif Intell Tools*, 2021, 30(6n08): 2140014.
- [20] Asad MM, Erum D, Churi P, et al. Effect of technostress on psychological well-being of post-graduate students: a perspective and correlational study of higher education management[J]. *Int J Inf Manag Data Insights*, 2023, 3(1): 100149.
- [21] Taylor JM. Psychometric analysis of the ten-item perceived stress scale[J]. *Psychol Assess*, 2015, 27(1): 90-101.
- [22] 樊蓓蓓, 张春华. 大学生心理健康的标准及评估(英文)[J]. *中国临床康复*, 2006, 46: 223-5.
- [23] Zhao F, Zhang CC, Geng BC. Deep multimodal data fusion[J]. *ACM Comput Surv*, 2024, 56(9): 1-36.
- [24] Mukta MSH, Ahmad J, Zaman A, et al. Attention and meta-heuristic based general self-efficacy prediction model from multimodal social media dataset[J]. *IEEE Access*, 2024, 12: 36853-73.
- [25] Deng H, Yang ZG, Hao TY, et al. Multimodal affective computing with dense fusion transformer for inter- and intra-modality interactions[J]. *IEEE Trans Multimed*, 2022, 25: 6575-87.
- [26] Zhao YX, Cao XY, Lin JL, et al. Multimodal affective states recognition based on multiscale CNNs and biologically inspired decision fusion model[J]. *IEEE Trans Affect Comput*, 2023, 14(2): 1391-403.
- [27] Li WB, Tan RC, Xing Y, et al. A multimodal psychological, physiological and behavioural dataset for human emotions in driving tasks[J]. *Sci Data*, 2022, 9(1): 481.
- [28] Zhu LN, Zhu ZC, Zhang CW, et al. Multimodal sentiment analysis based on fusion methods: a survey[J]. *Inf Fusion*, 2023, 95: 306-25.
- [29] University DP, Saleem Abdullah SM, Abdulazeez AM, et al. Facial expression recognition based on deep learning convolution neural network: a review[J]. *J Soft Comput Data Min*, 2021, 2(1): 53-65.
- [30] 耿亿霖, 臧琳, 毛飞跃, 等. 基于U-Net神经网络的CALIPSO产品漏检层次分类[J]. *光学学报*, 2024, 44(24): 97-106.
- [31] Fu LY, Li SW. A new semantic segmentation framework based on UNet[J]. *Sensors (Basel)*, 2023, 23(19): 8123.
- [32] Wang X, Jing SH, Dai HF, et al. High-resolution remote sensing images semantic segmentation using improved UNet and SegNet [J]. *Comput Electr Eng*, 2023, 108: 108734.
- [33] Abdollahi A, Pradhan B, Alamri AM. An ensemble architecture of deep convolutional Segnet and Unet networks for building semantic segmentation from high-resolution aerial images[J]. *Geocarto Int*, 2022, 37(12): 3355-70.
- [34] Behera RK, Jena M, Rath SK, et al. Co-LSTM: Convolutional LSTM model for sentiment analysis in social big data[J]. *Inf Process Manag*, 2021, 58(1): 102435.
- [35] Lindemann B, Müller T, Vietz H, et al. A survey on long short-term memory networks for time series prediction[J]. *Procedia CIRP*, 2021, 99: 650-5.
- [36] Shanmuganathan V, Suresh A. LSTM-Markov based efficient anomaly detection algorithm for IoT environment[J]. *Appl Soft Comput*, 2023, 136: 110054.
- [37] Wu J, Zhu TL, Zhu JH, et al. A optimized BERT for multimodal sentiment analysis[J]. *ACM Trans Multimedia Comput Commun Appl*, 2023, 19(2s): 1-12.