

基于多模态情感交互的学生心理健康支持系统

史静怡¹, 杨鹏飞^{1*}, 黄嘉阳¹, 贾瑞¹, 姚炫竹¹, 许喆¹, 常志奇¹, 戴逸飞¹, 魏萍²

¹西安电子科技大学计算机科学与技术学院, 陕西 西安

²西安电子科技大学心理健康教育中心, 陕西 西安

收稿日期: 2025年4月1日; 录用日期: 2025年4月23日; 发布日期: 2025年4月30日

摘要

心理健康是社会普遍关注的问题, 我国抑郁症患者群体人数持续扩大, 发病群体呈现年轻化趋势, 且高校学生群体占比持续升高, 利用人工智能技术赋能高校心理健康工作刻不容缓。本文针对现有高校心理健康工作存在集中评测精度不足、隐患排查时效性差、传统面谈覆盖面窄等问题, 提出了基于多模态情感交互的学生心理健康支持系统。该系统依托校园行为大数据, 构建学生异常情绪及行为监测与预警机制; 通过自主研发的大语言模型, 实现学生情绪的动态识别与智能评估, 并结合心理学理论, 动态适配个性化疏导策略, 支持多角色的情感陪伴与心理支持。系统在实际应用中表现出良好的效果, 心理状态评测准确率超过85%, 显著提升了心理服务的精准性与响应效率, 为校园心理健康教育体系的智能化与科学化建设提供了有力的技术支撑。

关键词

心理健康评估, 心理状态监测, 多模态情感交互, 个性化疏导, 大语言模型应用

Student Mental Health Support System Based on Multimodal Emotional Interaction

Jingyi Shi¹, Pengfei Yang^{1*}, Jiayang Huang¹, Rui Jia¹, Xuanzhu Yao¹, Zhe Xu¹, Zhiqi Chang¹, Yifei Dai¹, Ping Wei²

¹School of Computer Science and Technology, Xidian University, Xi'an Shaanxi

²Mental Health Education Center, Xidian University, Xi'an Shaanxi

Received: Apr. 1st, 2025; accepted: Apr. 23rd, 2025; published: Apr. 30th, 2025

Abstract

Mental health is a common concern in society. The number of depression patients in my country

*通讯作者。

文章引用: 史静怡, 杨鹏飞, 黄嘉阳, 贾瑞, 姚炫竹, 许喆, 常志奇, 戴逸飞, 魏萍. 基于多模态情感交互的学生心理健康支持系统[J]. 嵌入式技术与智能系统, 2025, 2(2): 78-95. DOI: 10.12677/etis.2025.22007

continues to expand, the incidence group shows a trend of younger age, and the proportion of college students continues to increase. It is urgent to use artificial intelligence technology to empower college mental health work. Aiming at the problems of insufficient centralized evaluation accuracy, poor timeliness of hidden danger investigation, and narrow coverage of traditional interviews in existing college mental health work, this paper proposes a student mental health support system based on multimodal emotional interaction. Relying on campus behavior big data, the system builds a monitoring and early warning mechanism for students' abnormal emotions and behaviors; through the independently developed large language model, it realizes the dynamic recognition and intelligent evaluation of students' emotions, and combines psychological theory to dynamically adapt personalized counseling strategies to support emotional companionship and psychological support for multiple roles. The system has shown good results in practical applications, with an accuracy rate of more than 85% in psychological state evaluation, which significantly improves the accuracy and response efficiency of psychological services, and provides strong technical support for the intelligent and scientific construction of the campus mental health education system.

Keywords

Mental Health Assessment, Mental State Monitoring, Multimodal Emotional Interaction, Personalized Counseling, Large Language Model Application

Copyright © 2025 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

随着社会的快速发展, 学生群体正在面临着激烈学业竞争、巨大就业压力、复杂人际关系等多重心理挑战, 其心理健康问题愈发凸显。最新研究数据显示, 我国大学生抑郁症状检出率高达 24.71% [1]; 《中国国民心理健康发展报告(2021~2022)》进一步揭示, 大学生群体中抑郁和焦虑的风险检出率分别为 21.48% 和 45.28% [2]。心理危机不仅影响大学生个体的成长与发展, 也给家庭和学校带来诸多挑战, 更有可能导致极端行为的发生, 进而对国家的人才培养质量和社会稳定性造成不可忽视的负面影响。

校园心理健康教育具有“育心、启智、养德、导行”的重要思政引领作用, 是应对学生心理问题、建设教育强国的重要保障。当前, 心理健康教育工作仍面临精准性不足、响应滞后、个性化支持匮乏等挑战, 亟需从心理状态监测、心理状态评估与疏导这两个关键环节进行优化提升。在心理状态监测方面, 现有方法大多停留在静态数据分析, 缺乏持续性、系统性监测, 难以揭示学生深层次的心理状态与问题根源, 且评估结果单一, 难以有效支撑个性化疏导。在心理状态评估与疏导方面, 现有评估手段多依赖单一问卷或测试, 缺乏多维度、多模态的数据支撑, 评估结果难以全面刻画学生的个体心理特征。心理干预多依赖统一课程和普适性讲解, 忽视个体差异与多样化需求, 且心理咨询资源有限, 服务隐蔽性不足, 学生因羞耻心或担忧评价往往不愿敞开心扉, 影响干预效果。

近年来, 情感计算技术凭借在情绪识别与个性化支持方面的优势, 广泛应用于医疗、教育等领域。针对上述问题, 构建以情感交互技术为支撑的系统化心理健康支持体系成为关键突破口。情感数字人作为情感交互的应用载体, 能够感知学生情绪变化, 融合多模态数据, 精准解析心理状态, 并以无评判、私密的方式提供个性化心理支持, 消除传统心理咨询中隐蔽性不足的问题, 提升疏导效果。

基于此, 本文以情感数字人为核心, 探索其在校园心理健康教育中的应用路径与实践方法。在监测环节, 系统融合多源异构数据, 实时感知学生行为模式与心理状态, 动态识别情绪波动与潜在风险, 构

建全面、细粒度的心理变化监测机制。在评估与疏导环节，情感数字人基于学生个体差异，深入解析情绪演化过程与潜在心理需求，生成精准评估结果，并通过拟人化情感交互方式建立信任关系，提供具备情感共鸣与心理支持价值的个性化疏导服务。通过构建智能化心理监测系统与评估-疏导一体化服务机制，可以有效提升心理问题筛查的覆盖率、评估的精准性与干预的适配性，推动高校心理健康教育体系智能化、科学化转型，助力学生心理健康保障与社会可持续发展。本文主要创新点如下：

第一，提出了基于校园大数据的实时心理状态监测方法。该方法能够在不干扰学生正常学习与生活的前提下，持续采集其生活行为数据，实时监测心理状态变化并进行风险预警。避免了传统调查方式中由于有意隐瞒、环境干扰等导致的信息偏差问题，提升了心理健康风险识别的及时性与隐蔽性检测能力。

第二，提出了基于大语言模型的心理状态评估和情感支持方法。即以情感数字人为核心，基于自主研发的大模型，通过模拟多角色、多场景的情感交互，实现了学生心理状态的智能评估，结合历史数据提供个性化、动态适配的心理疏导服务，并生成个性化的智能报告。突破了传统依赖单一数据维度或经验判断的局限，打破了时空资源限制，增强了心理支持的情感共鸣与持续陪伴效果。

2. 方案设计

本文围绕学生心理健康问题的动态监测与精准干预，实现了基于多模态情感交互的学生心理健康支持系统，具体包含了两个功能模块：基于校园大数据的实时心理状态监测模块、基于大语言模型的心理状态评估和情感支持模块。系统总体架构如图 1 所示。

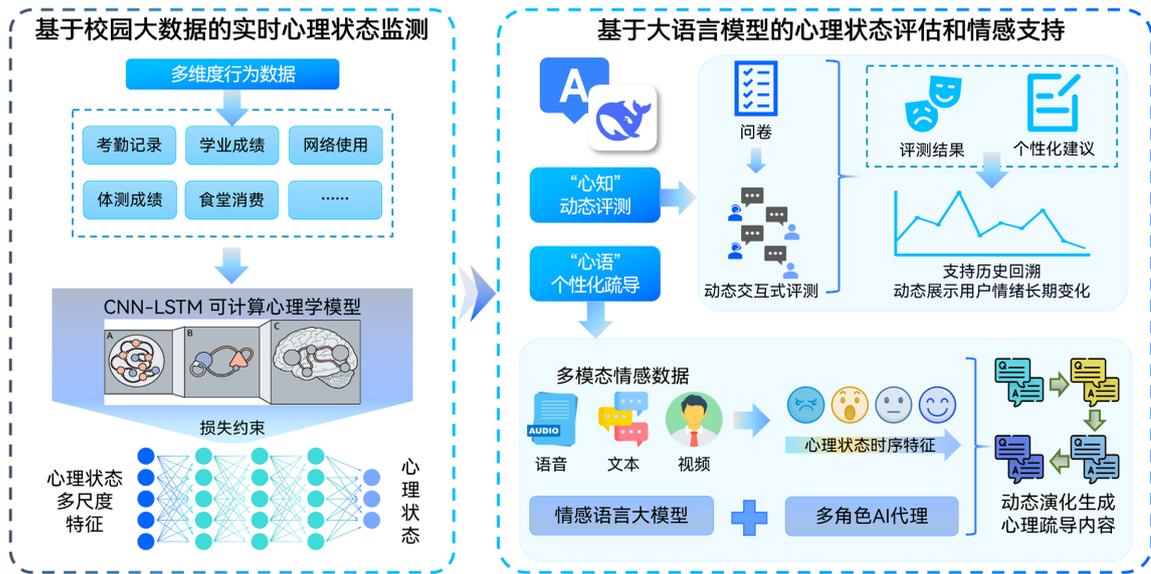


Figure 1. System overall architecture diagram
图 1. 系统总体架构图

基于校园大数据的实时心理状态监测模块旨在实现学生心理状态的连续感知与动态风险识别。本文首先对采集到的校园行为数据(如消费记录、考勤签到、课程成绩等)进行预处理与特征构建；随后引入位置编码与 Transformer Encoder 对行为序列进行建模，提取心理相关的时序特征；采用自监督学习策略，在时序数据中注入随机掩码噪声，增强模型对特征结构与时间依赖的学习能力。最终输出个体在不同时段的情绪状态标签，用于生成动态心理健康画像并驱动下游疏导模块的响应机制。

基于大语言模型的心理状态评估和情感支持模块旨在实现心理评估与个性化疏导的深度融合。本文

基于 DeepSeek: R1-14B 模型进行多轮微调, 数据涵盖心理测评文本、情绪问答对话、心理咨询记录等; 引入多模态融合机制, 支持文本、语音、视频(表情与姿态)输入, 避免用户的刻意隐瞒现象; 构建多角色 AI 代理, 分别训练老师、家人、朋友三种角色, 并引入情感共鸣调控机制提升语言共情力。

最终, 通过上述两个模块的协同运行, 本文构建了集智能化心理监测、评估疏导于一体的服务机制。基于情感数字人技术, 推动心理健康服务智能化升级, 显著提升了校园心理健康教育的覆盖率、响应速度与服务质量。

3. 基于校园大数据的实时心理状态监测

基于校园大数据的实时心理状态监测模块旨在构建一套覆盖高校学生数据采集、特征提取、心理健康状态监测及危机分级预警的全流程智能化监测体系, 其总体结构如图 2 所示。该模块依托信息化平台与大数据分析技术, 从多源异构的学生数据中提取有效信息, 进行心理健康状态的评估与预警, 为早期干预提供支持。

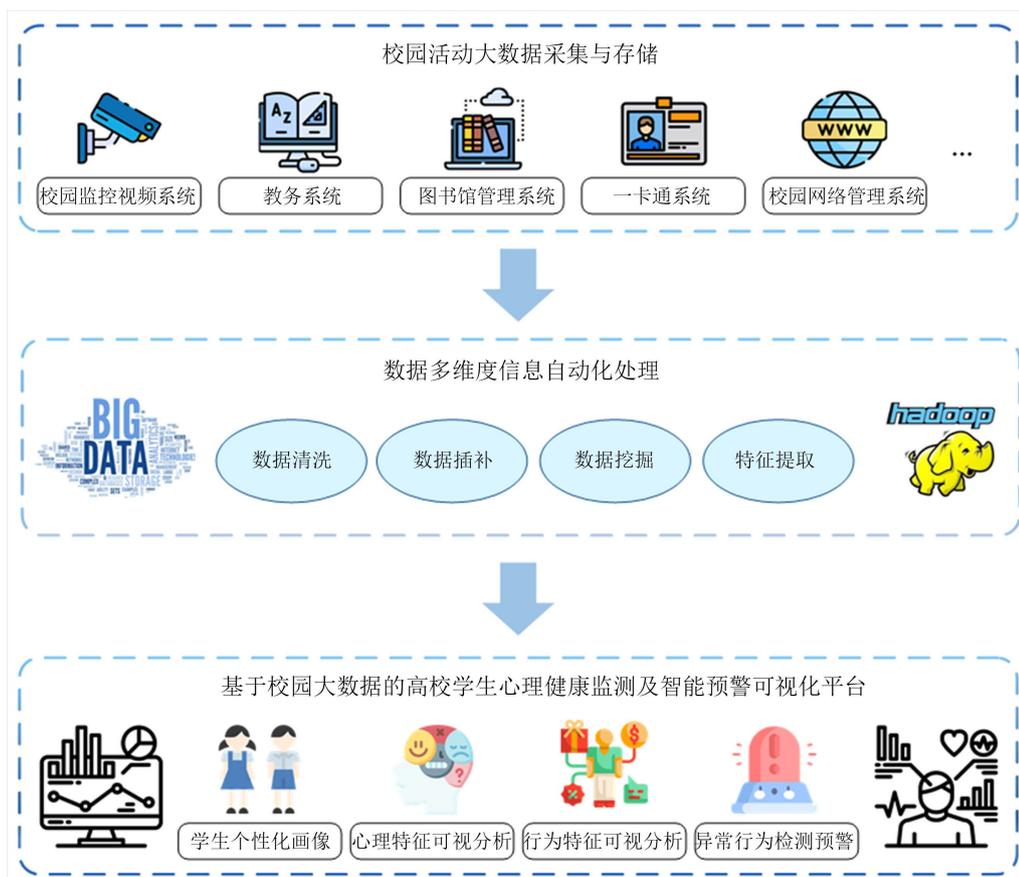


Figure 2. Real-time psychological state monitoring process based on campus big data

图 2. 基于校园大数据的实时心理状态监测流程

3.1. 数据采集

该模块通过整合学生在校内的多源异构数据, 包括食堂消费记录、图书馆打卡情况、学业成绩波动、体测结果变化以及网络使用习惯等, 经过数据清洗和预处理, 构建了一个包含静态特征和动态特征的综合数据集, 形成动态更新的个人行为画像, 最终得到学生在校的相关数据如表 1 所示。

Table 1. Student data sheet
表 1. 学生在校相关数据表

数据	数据条数
本科生基本信息表	34,273
食堂消费记录表	3,141,047
图书馆打卡记录表	366,855
考试成绩表	2,727,288
学生体测成绩表	37,100
校园网使用记录表	6,123,024
学生心理健康信息表	996
贫困生申请记录表	11,543
奖学金申请记录表	27,621

3.2. 特征提取

学生特征可大致分为静态特征和动态特征。静态特征主要为不会随时间发生变化的特征，如学生性别、民族和专业等；同时，由于学生家庭经济状况、成绩和身体素质在短时间内基本不会发生巨大变化，本文暂将反映学生成绩和身体素质的特征也视为静态特征。动态特征则主要为学生消费、图书馆使用、网络使用情况等相关特征。最终构建的完整特征体系如表 2 所示。

Table 2. Table of static and dynamic characteristics of students
表 2. 学生静态特征与动态特征表

学生静态特征			学生动态特征		
特征来源	特征名	特征类别	特征来源	特征名	特征类别
基本信息	性别	分类特征	食堂消费记录	用餐总次数	数值特征
	是否少数民族			用餐总金额	
	专业			早餐用餐次数/金额	
	年级			午餐用餐次数/金额	
	是否贫困生			晚餐用餐次数/金额	
体测成绩	体测总分	数值特征	图书馆打卡记录	总打卡次数	数值特征
	BMI			上午打卡次数	
	肺活量			下午打卡次数	
	50 米成绩			晚上打卡次数	
学业成绩	初修成绩均分	数值特征	校园网使用情况	校园网使用次数	数值特征
	初修成绩标准差			校园网使用上行/下行流量数	
	初修次数			上午校园网使用次数	
	重修次数			上午校园网使用上行/下行流量数	
	重考次数			下午校园网使用次数	
	获得奖学金次数			下午校园网使用上行/下行流量数	
	获得奖学金总额			晚上校园网使用次数	
	平均获得奖学金金额			晚上校园网使用上行/下行流量数	
	深夜校园网使用次数		深夜校园网使用上行/下行流量数		

3.3. 心理健康状态监测

本文提出了一种融合静态和时序特征的自监督学习方法。在数据处理方面，将采样粒度细化到每日，构建更精细的时间序列数据；在模型构建上，设计了基于门控机制的特征融合模块，将学生的静态特征(如年级、性别、专业等)与动态行为序列进行自适应融合；同时，借助自监督学习策略，通过在时序数据中注入随机噪声并重建原始序列和静态特征的方式，充分利用大量未标注数据来提升模型的特征表示能力。

使用有标签的训练集 $D_{\text{train}} = \{(X^{(i)}, s^{(i)}, y^{(i)}) | 1 \leq i \leq N_{\text{train}}\}$ 、测试集 $D_{\text{test}} = \{(X^{(i)}, s^{(i)}, y^{(i)}) | 1 \leq i \leq N_{\text{test}}\}$ ，其中， $X^{(i)} \in \mathbb{R}^{d_t \times T}$ 为按天数采样的时间序列， T 为当月天数， d_t 为每个时间点的特征数量。 $X^{(i)} = [\mathbf{x}_1^{(i)}, \dots, \mathbf{x}_T^{(i)}, \dots, \mathbf{x}_T^{(i)}]$ ，其中 $\mathbf{x}_t^{(i)} \in \mathbb{R}^{d_t}$ 为第 t 天的观测数据。 $s^{(i)} \in \mathbb{R}^{d_s}$ 为静态特征， d_s 是静态特征维度。 $y^{(i)}$ 表示学生的心理状态($y^{(i)} = 1$ 表示存在心理危机风险， $y^{(i)} = 0$ 表示心理状态正常)。无标签数据集定义为 $D_{\text{unlabel}} = \{(X^{(i)}, s^{(i)}) | 1 \leq i \leq N_{\text{unlabel}}\}$ 。本节研究的问题可以形式化为：设计一个判别函数 $G(\cdot)$ ，任意给定一个学生一个月内的动态时序特征 $X^{(i)}$ 和静态特征 $s^{(i)}$ ，判断其心理状态 $\hat{y}^{(i)} = G(X^{(i)}, s^{(i)})$ 。该判别函数需要同时考虑时序特征中的动态变化模式和静态特征中的背景信息，从而实现对学生心理状态的准确预测。

本节的关键挑战在于如何有效建模时序数据中的动态变化模式，以及如何将静态特征与时序特征进行融合。如图 3 所示，本文提出的模型主要包含以下关键步骤。首先，通过门控机制将每个时间点的动态特征与静态特征进行融合，得到融合特征表示 $\bar{\mathbf{x}}_t^{(i)}$ 。其次，将静态特征通过线性变换编码为 $\bar{\mathbf{x}}_0^{(i)}$ ，并将其作为时间序列的第一个位置，这种设计类似于自然语言处理中[CLS]标记的作用。接着，对每一天的特征表示进行位置编码，并输入到由自注意力机制和前馈神经网络组成的编码器中，得到各个时间点的特征表示 $\mathbf{z}_t^{(i)}$ ， $0 \leq t \leq T$ 。其中， $\mathbf{z}_0^{(i)}$ 作为整体序列的表示，将被输入到分类器中得到最终的分类结果。最后，所有时间点的特征表示 $\mathbf{z}_t^{(i)}$ ， $0 \leq t \leq T$ 也将用于自监督学习任务，以提升模型的特征提取能力。

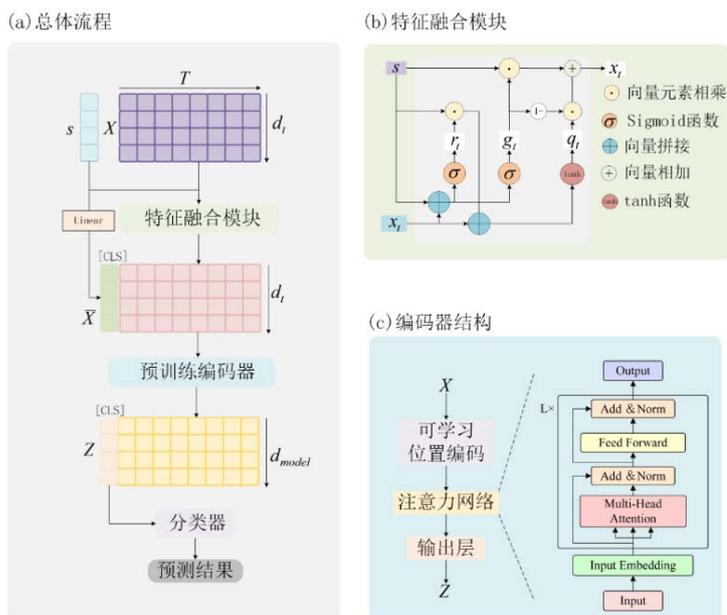


Figure 3. Schematic diagram of the self-supervised learning model architecture that integrates static and temporal features
图 3. 融合静态和时序特征的自监督学习模型架构示意图

3.3.1. 特征融合

在学生心理状态预测任务中，动态特征和静态特征各自承载着不同维度的信息价值。动态特征捕捉了学生行为模式随时间的变化，这些变化往往能反映学生心理状态的波动。而静态特征如性别、年级、学院等则构成了学生的基本画像，为行为解读提供了背景框架。两类异质特征的有效融合是模型性能的关键。若简单地将这两类特征直接拼接，则难以充分挖掘它们之间的复杂关系。在实际场景中，两类特征的信息密度和重要性在不同时间点存在显著差异，且它们之间存在复杂的交互关系。

针对这些考虑，本文使用了一种基于门控机制的特征融合方法，该方法能够自适应地调整两类特征的贡献比例。具体地，在 t 时刻：

$$g_t = \text{sigmoid}\left(W_g \left[\mathbf{x}_t^{(i)}, \mathbf{s}^{(i)} \right] + b_g\right) \quad (1)$$

$$r_t = \text{sigmoid}\left(W_r \left[\mathbf{x}_t^{(i)}, \mathbf{s}^{(i)} \right] + b_r\right) \quad (2)$$

$$q_t = \tanh\left(W_q \mathbf{x}_t^{(i)} + b_q + r_t \odot U_q \mathbf{s}^{(i)}\right) \quad (3)$$

$$\bar{\mathbf{x}}_t^{(i)} = g_t \odot \mathbf{s}^{(i)} + (1 - g_t) \odot q_t \quad (4)$$

在这个机制中， g_t 作为控制原始动态特征的保留程度， r_t 调节静态特征对候选状态 q_t 的影响强度。这种设计使模型能够根据不同时间点的特征表现，动态调整融合策略。当某个时间点出现异常行为模式时，模型可以通过门控机制赋予该动态特征更高的权重，而当动态特征不足以提供有效信息时，模型则可以更多地依赖静态特征进行判断。

此外，受到自然语言处理领域的启发，本文将每个学生的静态特征 \mathbf{s} 通过线性变换编码为特殊的 [CLS] 标记，并将其作为时间序列的起始位置，起到类似 [CLS] 标记的作用：

$$\bar{\mathbf{x}}_0^{(i)} = W_{\text{cls}} \mathbf{s} + b_{\text{cls}} \quad (5)$$

这样，编码器的完整输入为：

$$\bar{\mathbf{X}}^{(i)} = \left[\bar{\mathbf{x}}_0^{(i)}, \bar{\mathbf{x}}_1^{(i)}, \dots, \bar{\mathbf{x}}_T^{(i)} \right] \in \mathbb{R}^{d_t \times (T+1)} \quad (6)$$

这种设计使静态特征能够在自注意力机制中与所有时间点的动态特征进行全面交互，从而构建更为丰富的特征表示。通过这种双重融合机制，模型既能捕获局部时间点的特征交互，又能建立全局层面的特征关联，从而更全面地理解学生行为模式与心理状态之间的复杂关系。

3.3.2. 编码器

输入 $\bar{\mathbf{X}}^{(i)}$ 首先经过一个线性变换层，将每个时间点的特征映射到 d_{model} 维度空间，以满足后续注意力机制的计算需求：

$$\hat{\mathbf{X}}^{(i)} = W_{\text{proj}} \bar{\mathbf{X}}^{(i)} + b_{\text{proj}} \quad (7)$$

其中， $W_{\text{proj}} \in \mathbb{R}^{d_{\text{model}} \times d_t}$ ， $b_{\text{proj}} \in \mathbb{R}^{d_{\text{model}}}$ 。

本文采用可学习的位置编码矩阵 $P \in \mathbb{R}^{(T+1) \times d_{\text{model}}}$ ，其中每个元素都是可训练的参数。这种设计使得模型可以更灵活地学习序列中的位置信息。将位置编码与变换后的序列相加，得到带有位置信息的输入矩阵：

$$E_0^{(i)} = \hat{\mathbf{X}}^{(i)} + P \quad (8)$$

设编码器的层数为 L ，对于第 l 层 ($l=1, 2, \dots, L$)，其输入为 $E_{l-1}^{(i)}$ 。每一层首先通过三个不同的线性变

换得到查询矩阵 Q_l 、键矩阵 K_l 和值矩阵 V_l ：

$$Q_l = W_{Q,l} E_{l-1}^{(i)} \quad (9)$$

$$K_l = W_{K,l} E_{l-1}^{(i)} \quad (10)$$

$$V_l = W_{V,l} E_{l-1}^{(i)} \quad (11)$$

其中, $W_{Q,l}, W_{K,l} \in \mathbb{R}^{d_k \times d_{\text{model}}}$, $W_{V,l} \in \mathbb{R}^{d_v \times d_{\text{model}}}$ 。设置 $d_k = d_v = d_{\text{model}}/H$, H 为注意力头数。

对于第 h 个注意力头, 其计算过程为:

$$\text{head}_h = \text{Attention}(Q_l, K_l, V_l) = V_l \text{softmax}\left(\frac{Q_l^T K_l}{\sqrt{d_k}}\right) \quad (12)$$

将 H 个头的结果拼接并经过线性变换:

$$\hat{E}_l^{(i)} = W_l^O [\text{head}_1; \dots; \text{head}_H] \quad (13)$$

其中, $W_l^O \in \mathbb{R}^{d_{\text{model}} \times H d_v}$ 是一个线性变换权重矩阵, 用于将拼接后的结果转换回与输入维度 d_{model} 相同的空间。

接着进行残差连接和层归一化:

$$\tilde{E}_l^{(i)} = \text{LayerNorm}\left(\hat{E}_l^{(i)} + E_{l-1}^{(i)}\right) \quad (14)$$

最后, 利用前馈神经网络的激活函数为模型引入非线性变换的能力, 进一步挖掘更复杂的更抽象的特征。前馈神经网络由两个全连接层组成, 中间有一个激活函数。将计算结果与输入进行残差链接并进行层归一化得到一层多头注意力输出:

$$E_l^{(i)} = \text{LayerNorm}\left(W_{\text{fc},2} \text{ReLU}\left(W_{\text{fc},1} \tilde{E}_l^{(i)} + b_{\text{fc},1}\right) + b_{\text{fc},2} + \tilde{E}_l^{(i)}\right) \quad (15)$$

经过多个这样的多头注意力机制层和前馈神经网络层的处理后获得最终输出 $E_L^{(i)}$, 在本节中记为 $\mathbf{Z}^{(i)} = [\mathbf{z}_0^{(i)}, \mathbf{z}_1^{(i)}, \dots, \mathbf{z}_T^{(i)}]$ 。

3.3.3. 分类器

对于有标签的学生, 其对应样本为 $(X^{(i)}, s^{(i)}, y^{(i)})$, 将 $X^{(i)}$ 和 $s^{(i)}$ 输入到骨干网络中, 得到 $\mathbf{Z}^{(i)} = [\mathbf{z}_0^{(i)}, \mathbf{z}_1^{(i)}, \dots, \mathbf{z}_T^{(i)}]$, 从中取出 $\mathbf{z}_0^{(i)}$ 作为总体特征表示, 将其输入到由隐含层全连接神经网络构成的分类器中, 得到 $\hat{y}^{(i)}$ 作为分类结果。

$$\hat{y}^{(i)} = p\left(\mathbf{z}_0^{(i)}\right) = \text{sigmoid}\left(W_{\text{cls},2}\left(\text{ReLU}\left(W_{\text{cls},1} \mathbf{z}_0^{(i)} + b_{\text{cls},1}\right)\right) + b_{\text{cls},2}\right) \quad (16)$$

3.3.4. 模型训练过程

在模型训练过程中, 首先开展自监督学习阶段, 旨在通过重建任务增强模型对数据内部结构和特征分布的理解能力。在该阶段, 原始数据首先被注入随机掩码噪声, 以构造具有挑战性的训练样本, 从而模拟实际应用中可能出现的缺失或干扰情况。带噪声的数据被输入至特征融合模块与预训练编码器中, 提取包含时间序列模式的高维特征表示。随后, 模型利用多层感知机对动态特征进行重建, 以恢复被扰动的数据, 并从特征表示中还原静态特征, 从而确保编码器提取的表示不仅包含时间演化信息, 也保留了个体的背景属性。

为此，训练过程设计了包含动态特征重建误差与静态特征重建误差的联合损失函数，通过对二者的加权求和，引导模型学习数据的时序依赖与结构特征。该阶段训练结果为后续有监督学习奠定了良好的特征表达基础，有助于提升整体模型的鲁棒性与泛化能力。具体流程见图 4。

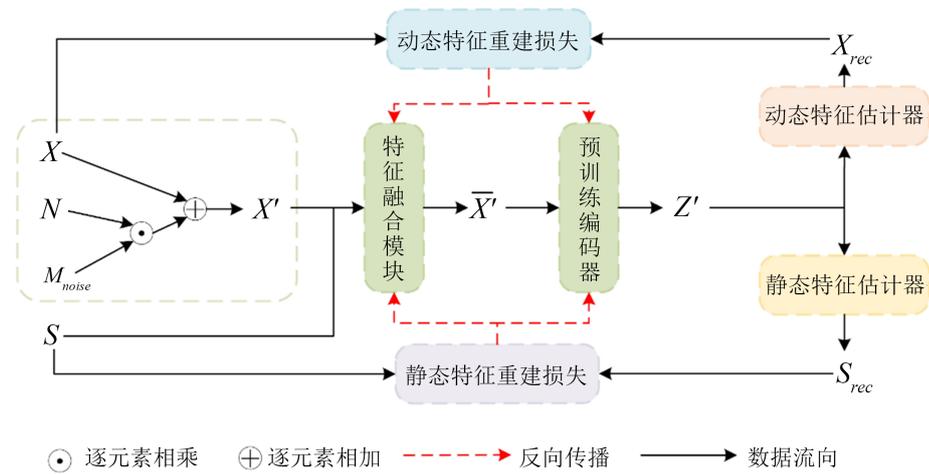


Figure 4. Schematic diagram of the self-supervised training process
图 4. 自监督训练过程示意图

3.4. 危机预警

在对学生心理状态进行初步判别后，系统基于二分类模型识别出存在心理健康风险的学生。一旦检测到潜在问题，系统将按照预设流程，生成预警信息并推送至辅导员及相关管理人员，以便及时介入并提供精准支持。系统根据预设的心理危机分级标准，生成不同级别的预警信息，确保多方协同介入，实现心理问题的提前预测与干预。

4. 基于大语言模型的心理状态评估和情感支持

在心理健康支持体系中，开展科学、动态、个性化的心理状态评估与情感支持，是实现早期识别、精准干预的关键路径。为突破传统测评方式在数据维度、交互模式与反馈响应上的局限，本文基于 DeepSeek 大语言模型架构，融合心理学知识体系与自然交互需求，构建了一个面向校园场景的智能化心理状态感知与疏导模型，提出心理状态评估与情感支持一体化的新范式。

通过融合学生历史对话内容与实时采集的多模态数据特征(包括文本、语音、视频)，模型能够动态识别个体情绪变化与深层心理需求；同时引入多角色 AI 代理机制，自动匹配情绪状态与心理疏导方案，生成贴合学生沟通偏好的个性化干预内容，显著提升心理支持的及时性、共情性与实效性。

4.1. 数据集构建

为支撑心理状态评估与情绪支持的多层次功能实现，本文围绕任务需求设计并构建了两类核心数据集：一类为面向大语言模型微调的心理对话数据集，用于增强模型在心理测评与情感疏导中的语言理解与生成能力；另一类为多模态辅助识别数据集，用作 CNN-LSTM 情绪识别模型的输入，以实现基于语音、视频等非语言信号的心理状态感知，构建多源信息融合的心理健康评估机制。

4.1.1. 大语言模型微调数据集

本数据集主要用于提升语言模型在校园心理健康场景下的交互自然性与专业性，数据构建涵盖两类

子任务：心理测评问答、多角色情感支持对话。

1. 心理测评数据集构建

本文对国际公认的心理量表，如 PHQ-9、DASS-21、SCL-90 进行了结构性转换。借助提示词工程，将量表条目改写为符合人机自然交互风格的提问表达。在完整保留评估维度与专业内涵的同时，显著提升了语言的亲和力与使用体验，克服了传统测评在交互形式上的限制。

2. 多角色情感支持数据集构建

本部分数据集基于开源语料如 PsyQA、CPsyCoun，结合本文设计的“角色档案”与“场景档案”，构建多轮、多角色的情绪疏导对话。在生成过程中，引入思维链提示策略，并严格参考认知行为疗法 (Cognitive Behavioral Therapy, CBT) 等心理干预原则，以保证对话的逻辑性、专业性与情绪连贯性。具体流程见图 5。

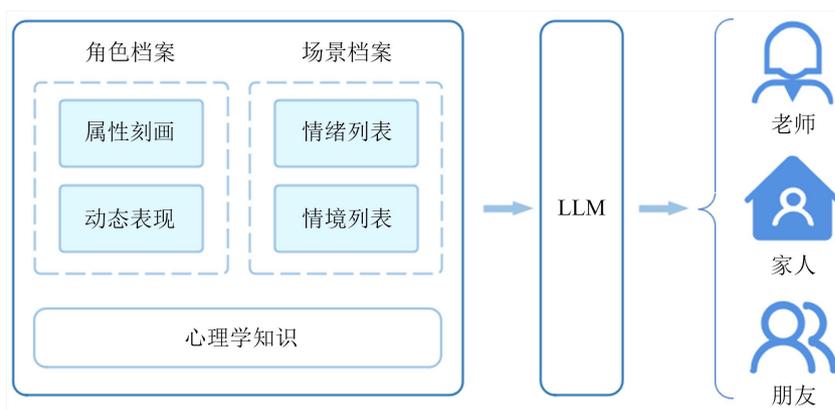


Figure 5. Multi-role emotional support dataset construction process

图 5. 多角色情感支持数据集构建流程

(1) 角色档案

对话式 AI 角色的关键在于深度地理解和模仿人类的交流，致力于创造逼真、可信且引人入胜的虚拟对话伙伴。所以在角色建模的过程中，本文将角色构成分为“属性刻画”与“动态表现”两大类，以构建稳定的且具有鲜明人格的对话代理。属性刻画包括身份、三观、经历；动态表现则体现在互动中的语言特征与性格。具体内容见表 3。

Table 3. Character files

表 3. 角色档案

角色定义	类别	具体内容
属性刻画	身份	姓名、性别、年龄、职业等
	三观	世界观、人生观、价值观
	经历	有意义的经历
动态表现	语言特征	口头禅、文风特点、常用词句等
	性格	温柔、冷漠等

(2) 场景档案

在多角色个性化疏导的设计中，仅依赖角色建模仍难以全面覆盖用户在实际交互中的多样化心理需

求。所以本文引入了场景档案的构建，旨在通过情绪与情境两个维度，系统地定义学生可能面临的心理状态与对话背景，为后续对话内容的生成与适配提供丰富且细致的上下文支持。具体内容见表 4。

Table 4. Scene files

表 4. 场景档案

场景定义	具体内容
情绪列表	快乐、悲伤、焦虑、愤怒、厌倦、困惑 同情、平静、欣赏、满足、尴尬、渴望
情境列表	学习、工作、家人、朋友、社交、恋爱 身心健康、兴趣爱好、日常生活、个人安全、未来

本文构建的多角色引入了“老师”、“家人”、“朋友”三种典型的角色，作为虚拟代理进行对话式的情感支持。通过角色属性和语言风格的设定，以及情绪、情境类别的组合，系统化生成多轮的对话样本，从而确保数据集在内容、风格与心理支持策略上的多样性与专业性。

4.1.2. 多模态辅助识别数据集

为支持基于非语言信息的心理状态识别，本文设计采集了文本、语音与视频三类模态的数据，主要作为 CNN-LSTM 识别模型的输入，以实现对学生情绪状态的综合感知。

数据采集基于统一交互平台完成，确保多模态数据在时间轴上的精确同步。文本模态通过记录用户输入的对话内容，完成分词、清洗与语义编码处理；语音模态由麦克风采集，提取梅尔频率倒谱系数 (MFCC)、基频、能量等声学特征；视频模态通过摄像头捕捉面部表情与身体姿态，提取关键点位置信息与表情参数。

为确保模态之间信息的准确融合，系统对采集的多模态数据执行统一时间戳标记，进行精确的时间对齐处理。所有数据经标准化预处理流程，包括去噪、异常值剔除与归一化等操作，进一步提升特征提取的稳定性与输入的一致性。实验数据均来源于标准化采集流程，场景设计中充分考虑多样性与标签有效性，以支持后续模型训练的泛化能力。

4.2. 模型微调

为了进一步提高模型的专业适配性，本文基于上述构建的数据集，采用 LoRA (Low-Rank Adaptation) 技术对基础大语言模型 DeepSeek: R1-14B 进行高效微调。LoRA 通过在保持原有模型参数不变的前提下，仅对少量新增的低秩矩阵进行训练，从而显著降低了微调过程中的计算与存储开销，适合在资源受限的环境中部署和优化。同时集成了高效训练框架 Unsloth，以进一步加速训练流程。

通过将 LoRA 与 Unsloth 结合使用，本文在保证模型性能的基础上，实现了心理任务领域下的轻量化、高效率模型适配，为后续部署于校园的心理服务系统提供了坚实的模型基础。

4.3. 动态交互式评测

在完成对标准化心理量表的对话式转化后，本文进一步实现了面向真实应用场景的动态交互式评测。与传统的纸质问卷、人工统计评分的流程不同，本文借助了深度微调的大语言模型，通过自然语言的引导性提问与实时反馈机制，构建了智能化、自适应的交互式评测路径。

在评测过程中，模型以提示词为引导，以问卷条目为核心，结合情感识别，将每一条量表问题转化为自然的对话单元，并在用户作答后进行内容解析与情感识别。根据用户回答中出现的频率词汇和情绪表达，模型能够在对应症状维度进行即时得分。例如，当用户在睡眠质量相关的问题中使用了“经常惊

醒”、“多梦”等描述，模型就会在相应的指标上进行加分。各评测项目归属于不同的心理症状类别，如焦虑、抑郁、压力等，模型在对话的过程中逐步累积各维度得分，最终依据设定的分数阈值判断风险等级(例如，焦虑量表得分超过 10 分即判定为轻度焦虑)。

为了进一步提升评测的敏感性与精准性，本文设计了基于初筛结果的追问机制。当用户在某一指标上表现出了较高的风险倾向时，模型能够切换到针对性更强的细化问卷，展开更深层次的评测。具体示例见图 6。这种即时调整的策略有效地缩短了发现心理问题的时间差，提升了整体评测的效率与深度。

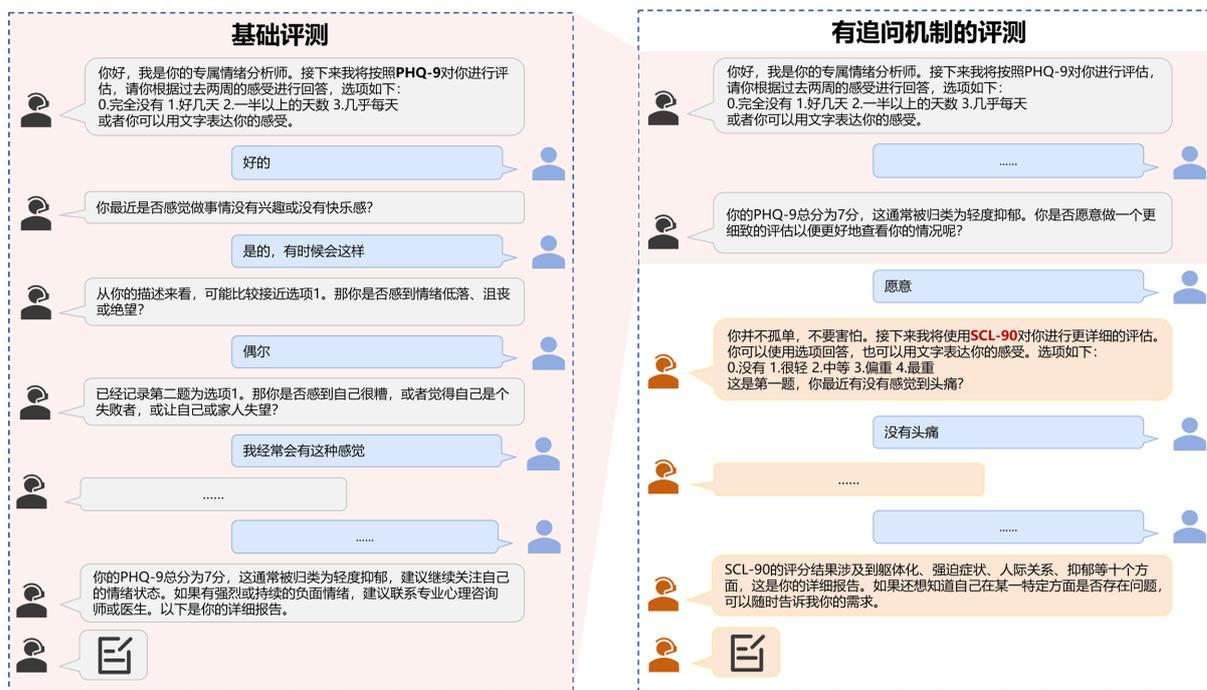


Figure 6. Comparison between basic evaluation and evaluation with follow-up mechanism

图 6. 基础评测与有追问机制评测的对比

此外，系统支持大规模在线评测与实时数据统计。可在短时间内完成对新生群体的心理普查，快速筛查出潜在的高风险个体，并对其心理状态变化进行持续跟踪，为后续心理干预提供数据支撑。

4.4. 智能化报告生成

在完成心理状态评测后，本文进一步设计了结构化、个性化的反馈机制。旨在为用户提供积极的心理暗示，提升他们的体验感，减少因为评测结果而引发的负面情绪。

系统支持 24 小时不间断评测服务，会主动关注深夜进行测试的用户。在特殊时段，模型在保持专业评测的同时，会根据用户的回答情绪适度提供共情反馈。在用户完成全部问题后，模型会自动生成鼓励性的话语以结束对话，如“恭喜你顺利完成了这次评测，下面一起来看看你的专属心理体检报告吧！”若检测到用户情绪低落，模型则会调整语气与表达方式，避免使用过于欢快的表述，以防用户产生反感或情绪抵触。

在正式反馈阶段，本文将传统的评测分数转化为具象化、游戏化的个性化心理体检报告。为了打破冰冷数据带来的疏离感，系统会根据用户的综合情绪状态自动生成相应风格的卡通角色形象。如压力值高时角色形象呈现出凌乱疲惫的状态，抑郁倾向明显时则呈现低落内敛，心理状态良好时则展现活泼轻松。这种具象化表达方式可以拉近系统与用户之间的距离，给予用户柔性化、正向的心理暗示，缓解因

心理评测带来的潜在不安情绪。

在报告生成后，系统还会进一步基于个体的评测结果提供针对性、个性化的建议。例如，针对“难以开始工作”的用户推荐他试试番茄工作法；对于有“社交焦虑”表现的用户则推送给他破冰对话技巧清单；对存在“入睡困难”的用户可以推荐白噪音助眠资源等，具体示例见图7。对于心理风险等级较高的用户，报告中会标注紧急心理求助的热线信息，确保用户在需要时能够及时地获得专业的帮助。这为构建友好、可持续的心理健康支持体系提供了有力支撑。



Figure 7. Intelligent reporting
图7. 智能化报告

4.5. 多角色个性疏导

为了实现对学生的个性化疏导，本文设计并实现了系统化、分阶段的智能情感支持体系。本模块整体架构由三个核心阶段构成：情绪识别、策略推理与回复生成。模型在充分理解用户心理状态的基础上，采用相应的对话策略生成支持性回复，在多轮对话的过程中帮助用户降低情绪困扰。各阶段相互协作、逐步推进，形成智能化心理疏导流程。完整的模块架构及处理流程如图8所示。

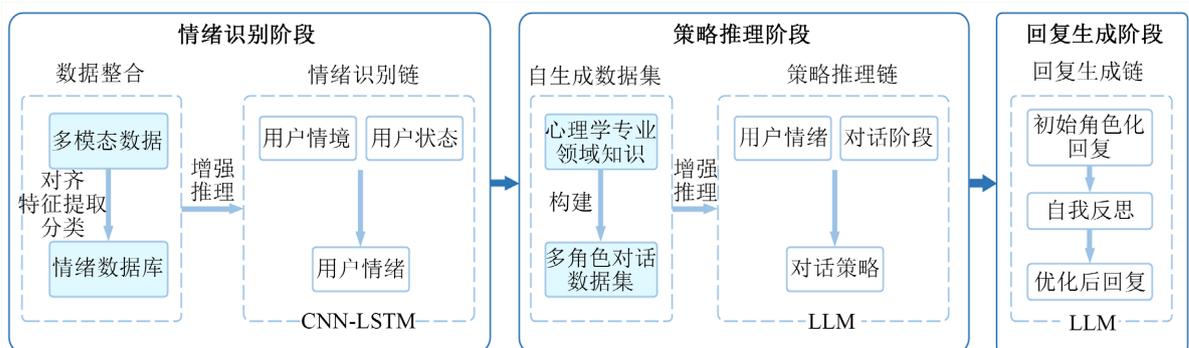


Figure 8. The overall framework of multi-role personality guidance
图8. 多角色个性疏导的整体架构

4.5.1. 情绪识别阶段

在前述多模态数据采集与特征提取的基础上, 本文进一步构建了系统化的学生情绪数据库, 并提出了一种基于卷积神经网络与长短时记忆网络(CNN-LSTM)的多模态情绪识别方法, 用于实现学生心理状态的实时、精准识别。

该方法融合了 CNN 在局部特征提取方面的优势与 LSTM 在建模时间动态特征方面的能力, 通过特征级融合与跨模态注意力机制对多源特征进行权重动态调整, 突出与心理状态高度相关的关键因素。最终通过全连接分类器输出用户的情绪状态, 实现即时且高精度的情绪识别。具体流程见图 9。

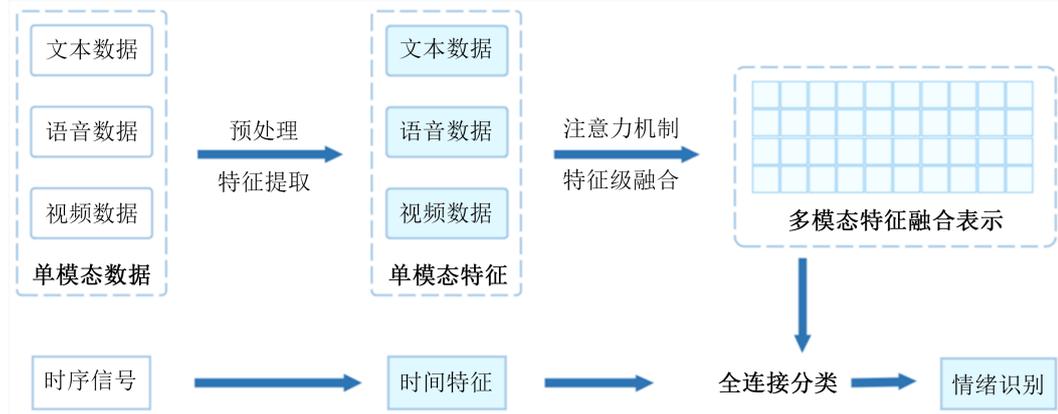


Figure 9. Multimodal emotion recognition process

图 9. 多模态情绪识别流程

整体模型结构包括局部特征提取、特征融合、时序建模与注意力加权四个阶段, 具体描述如下:

首先, 对于每种模态提取得到的特征表示 $\mathbf{x}_t^{(m)}$ (其中 m 表示模态类型, t 表示时间步), 通过一维卷积核 \mathbf{W}_c 进行局部特征提取:

$$\mathbf{f}_t^{(m)} = \text{ReLU}(\mathbf{W}_c * \mathbf{x}_t^{(m)} + \mathbf{b}_c) \quad (17)$$

其中, $\mathbf{f}_t^{(m)}$ 为卷积后的模态局部特征, $*$ 表示卷积操作, \mathbf{b}_c 为偏置项。

随后, 将所有模态在相同时间步上的特征拼接形成融合特征输入:

$$\mathbf{F}_t = [\mathbf{f}_t^{(1)}, \mathbf{f}_t^{(2)}, \dots, \mathbf{f}_t^{(M)}] \quad (18)$$

该融合特征序列被输入至双向长短时记忆网络(Bi-LSTM), 以捕捉跨时间的心理状态动态演化特征:

$$\mathbf{h}_t = \text{BiLSTM}(\mathbf{F}_t) \quad (19)$$

为增强模型对关键时刻与关键模态的关注能力, 本文在时序建模基础上引入注意力机制, 计算注意力权重 α_t 并将加权聚合隐藏状态表示为:

$$\alpha_t = \frac{\exp(\mathbf{v}^\top \tanh(\mathbf{W}_a \mathbf{h}_t + \mathbf{b}_a))}{\sum_i \exp(\mathbf{v}^\top \tanh(\mathbf{W}_a \mathbf{h}_i + \mathbf{b}_a))} \quad (20)$$

$$\mathbf{H}_{att} = \sum_t \alpha_t \mathbf{h}_t \quad (21)$$

最终, 将注意力加权后的融合特征表示 \mathbf{H}_{att} 输入至全连接层, 通过 Softmax 函数输出情绪分类结果:

$$\hat{y} = \text{Softmax}(\mathbf{W}_o \mathbf{H}_{att} + \mathbf{b}_o) \quad (22)$$

通过上述结构,模型实现了从局部模态特征提取到时序建模再到情绪状态分类的完整识别链,兼具细节表达与动态理解能力。随后将识别结果作为输入反馈给大语言模型,为后续内容生成阶段提供情绪感知支持。

4.5.2. 策略推理阶段

在学生心理疏导对话的过程中,策略选择的合理性直接影响交互的有效性与支持的精准性。本文在前述情绪识别阶段给出用户情绪的基础上,结合当前与用户的对话阶段(如初始接触、情绪表达、疏导建议等),共同构成策略推理的语境依据。

例如,在对话初期阶段,若用户情绪状态低落,系统将优先选用以“建立信任、传递支持感”为核心的引导型策略,生成体现倾听、认同与鼓励的响应内容,以降低用户的心理防御,促进情感表达。而在对话中后期,若检测到情绪逐渐趋于稳定,模型则可适时引入认知重构、问题解决等具有引导性的策略,以协助用户理性分析问题与形成积极的应对机制。

在准确捕捉学生情绪的基础上,经过学习由心理学专业领域知识指导生成的多角色对话数据集,并掌握其中基于认知行为疗法(CBT)推理范式的训练后,情感支持大语言模型具备匹配当前用户情绪特征的对话策略样本及相关知识。通过构建由对话阶段到对话策略的推理链,模型能够根据实时对话阶段与用户情绪状态,自适应生成最合适的对话策略。例如,在对话初期,模型更倾向于通过建立信任感主动引导交流。策略推理阶段对筛选和编码后的对话策略样本进行规范化处理,确保其准确反映策略要点及与学生情绪的契合度,并为后续语言生成环节提供高质量输入。

4.5.3. 回复生成阶段

在完成用户情绪识别与对话策略推理后,进入回复生成阶段。此阶段以大语言模型为核心,在多因素驱动下生成具备情绪感知与策略引导能力的响应内容。具体生成过程充分融合三类信息:一是识别出用户当前的情绪状态;二是当前所处的对话阶段及匹配策略;三是用户预设的情感支持角色(老师、朋友、家人)。

模型生成初步的角色化回复,回复生成过程中引入语言模型的自我反思机制,结合情感共鸣调控,动态检查生成内容是否真正贴合学生的情绪状态,语言表达是否符合其沟通偏好等要求。根据反思结果,进一步优化初步回复,例如将表达生硬的语句调整为更加亲和、安抚性的表述。最终输出的优化回复能够灵活地适应学生的个性化心理需求,有效提升情感疏导的针对性与干预成效。

5. 系统应用与测试

5.1. 学生异常情绪及行为监测与预警系统

本文基于校园大数据的实时心理状态监测设计并实现了学生异常情绪及行为监测与预警系统,如图10所示。该系统通过动态化捕捉学生的多维度行为数据,融合时间序列建模与异常检测技术,实现对潜在异常情绪与行为的精准识别与实时预警,旨在为校园心理健康管理提供智能化、系统化支持。

目前,该系统已与智慧校园平台实现集成,覆盖超过两万名学生样本,累计采集特征数据规模过亿。与常规心理筛查结果进行对比分析显示,本系统的心理异常识别准确率达到85%以上,验证了其实用性与有效性。

为了验证本系统中采用的时序模型在心理危机检测中的有效性,本文选择了以下几种针对时序数据的基线方法进行对比。

DTW-1NN 是基于动态时间规整(DTW)距离的最近邻算法,是处理时间序列分类的经典方法。

LSTM 是长短期记忆网络,专门设计用于捕捉序列数据中的长期依赖关系。

TSIN [3]该方法提出了一种基于双流 Informer 网络的学生心理健康预测模型，通过时间编码器捕捉时间序列的周期性趋势，行为编码器分析行为特征间的依赖关系，并利用中间融合模块和门控机制合并两个编码器输出。

TS2VEC [4]是一种基于对比学习的时间序列表示学习方法，通过最大化相似样本间的相似度来学习时间序列的表示。

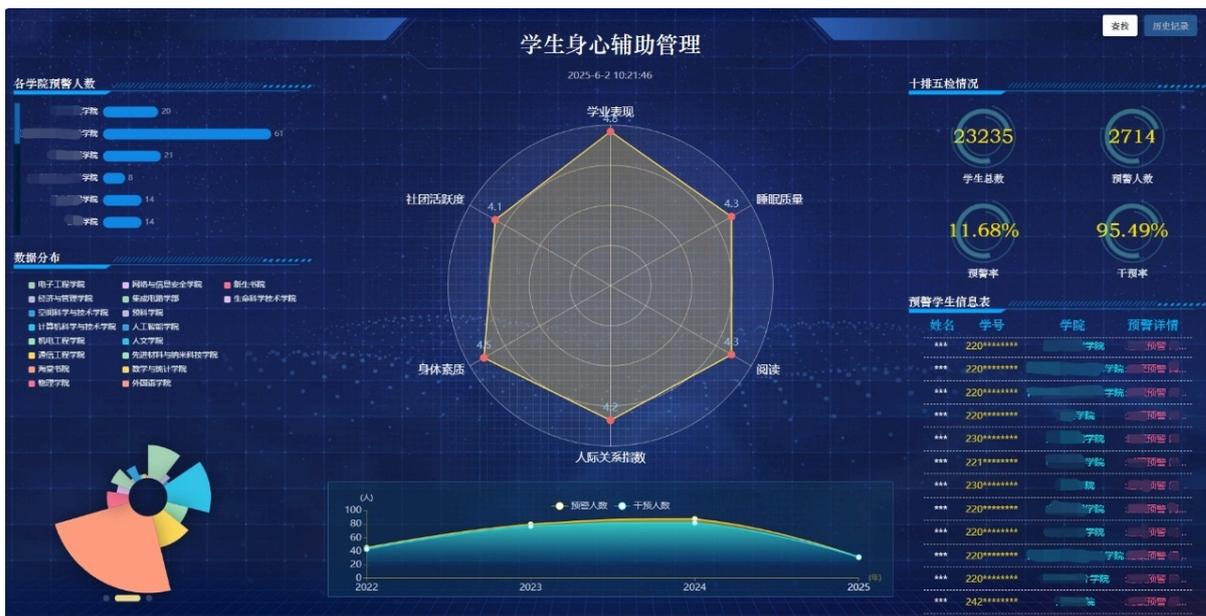


Figure 10. Students' abnormal emotions and behaviors monitoring and early warning system
图 10. 学生异常情绪及行为监测与预警系统

Table 5. Performance comparison of different methods
表 5. 不同方法的性能比较

方法	Accuracy	Precision	Recall	F1 Score
DTW-1NN	0.7828	0.7792	0.5933	0.6617
LSTM	0.8138	0.7857	0.6600	0.7174
TSIN	0.8533	0.7793	0.7533	0.7661
TS2VEC	0.8425	0.8088	0.7333	0.7692
Ours	0.8783	0.8510	0.8000	0.8247

表 5 展示了不同方法在测试集上的性能表现，本方法在所有评价指标上均优于其他方法。相比于传统的 DTW-1NN 方法，本方法的准确率提升了 9.55 个百分点，F1 分数提升了 16.3 个百分点；相比于深度学习方法 LSTM，准确率提升了 6.45 个百分点，F1 分数提升了 10.73 个百分点；相比于 TSIN 方法，准确率提升了 2.5 个百分点，F1 分数提升了 5.86 个百分点；相比于自监督时序学习方法 TS2VEC，准确率提升了 3.58 个百分点，F1 分数仍有 5.55 个百分点的提升。这些结果表明，本方法在捕捉学生行为的时序模式方面和无标签数据有效利用方面具有显著优势。

5.2. 心知心语平台

本系统旨在构建多终端、全场景的沉浸式心理疏导环境，为用户提供灵活多样的对话体验选择。系

统支持 Web 端、移动端和 MR 端独立运行，并实现数据互通，构建立体化的心理健康服务网络，为学生提供长期、动态的心理状态追踪服务，并建立个性化的心理健康档案。

5.2.1. Web 端与移动端

Web 平台和移动端均提供轻量化、便捷化的访问方式，采用极简交互设计，突出对话内容本身，确保交互过程的流畅性和专注度。为学生提供 7 × 24 小时不间断的心理健康支持，显著提升服务的即时性和可及性。如图 11 所示，分别为系统的登录界面、模块选择界面、聊天界面以及报告界面。

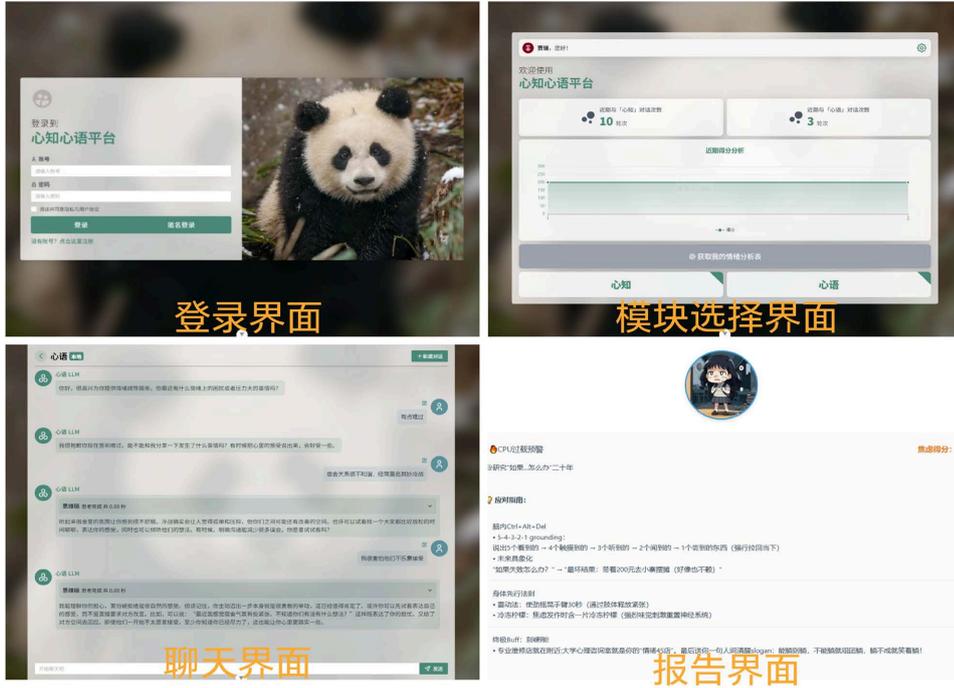


Figure 11. Interface display
图 11. 界面展示

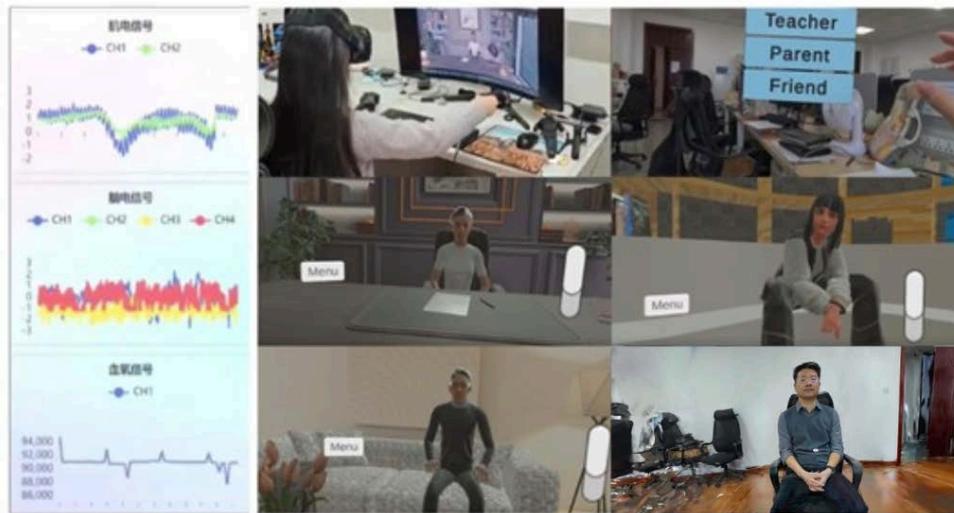


Figure 12. MR interactive display
图 12. MR 交互展示

5.2.2. MR 端

系统针对 VR 头显和 VisionPro 设备进行了深度适配, 依托混合现实(MR)技术构建沉浸式心理疏导环境。通过高精度 3D 建模和实时渲染技术, 创造逼真的虚拟对话场景, 增强心理干预的深度与效果。系统支持自然语言交互、手势识别等多种交互方式, 提供与虚拟角色面对面的深度交流体验, 如图 12 所示。未来计划进一步拓展虚拟场景的选择范围, 通过精心设计的视觉元素、环境音效和交互细节, 创造更具疗愈效果的对话空间。

6. 结束语

在人工智能时代背景下, 将 AI 技术引入心理健康服务体系, 为心理问题的早期识别、精准评估与个性化干预提供了全新思路。本文围绕智能化心理监测、评估疏导全过程, 构建了融合校园大数据与情感数字人的智能心理支持系统, 显著提升了学生心理状态识别的实时性与准确性, 切实减轻了高校学生管理工作的压力。为高校心理健康教育体系的智能化、科学化转型提供了可行路径与技术支撑, 具有良好的推广前景与现实意义。

基金项目

本研究得到了陕西省重点研发重点产业创新链(群)项目(2024GX-ZDCYL-02-15)、陕西省杰出青年科学基金(2025JC-JCQN-079)的支持。

参考文献

- [1] 王蜜源, 韩芳芳, 刘佳, 等. 大学生抑郁症状检出率及相关因素的 meta 分析[J]. 中国心理卫生杂志, 2020, 34(12): 1041-1047.
- [2] 傅小兰, 张侃, 陈雪峰, 等. 中国国民心理健康发展报告(2021~2022) [M]. 北京: 社会科学文献出版社, 2023.
- [3] Xu, J.M., Ding X.F., Ke, H.Y., *et al.* (2023) Student Behavior Prediction of Mental Health Based on Two-Stream Informer Network. *Applied Sciences*, **13**, Article 2371. <https://doi.org/10.3390/app13042371>
- [4] Yue, Z., Wang, Y., Duan, J., Yang, T., Huang, C., Tong, Y. and Xu, B. (2022) TS2Vec: Towards Universal Representation of Time Series. *Proceedings of the 36th AAAI Conference on Artificial Intelligence*, 22 February-1 March 2022, 8980-8987. <https://doi.org/10.1609/aaai.v36i8.20881>