

基于多模态数据与大模型赋能的 学生情感智能分析与监测系统设计

物联网工程专业 21 物联网 2 班黄月胜 指导老师：黎嘉文

摘要：随着人工智能与深度学习技术的深入发展，多模态情感分析在教育领域的应用潜力日益显现。传统单模态情感识别方法在捕捉学生复杂情绪状态方面存在局限，而多模态分析通过整合面部表情、语音信息与生理信号，能显著提升情感识别的准确性。当前教育信息化背景下，学生心理健康监测需求迫切，但现有方法存在时效性差、主观性强、难以规模化等问题，限制了其在校园环境的普及。

针对上述挑战，本文提出并实现了一套基于 ESP32 与 ESP32S3 硬件平台，结合轻量化多模态融合算法及大语言模型的学生情感智能分析与监测系统。该系统旨在利用低成本、高集成度的嵌入式技术，融合面部、语音、心率等多源数据，为教育工作者、家长及学生提供实时、准确、便捷的情感监测与支持工具。

在硬件层面，采用 ESP32 与 ESP32S3 双主板分布式架构。ESP32 主板集成 ESP32CAM、心率传感器等，实现面部表情识别、生理数据采集及基础反馈；ESP32S3 主板集成数字麦克风、音频功放、显示屏等，实现基于百度文心一言 API 的智能对话功能。软件层面，基于 Node.js 构建服务端，采用 SQLite 进行数据存储，开发了面向教师、学生、家长的多角色 Web 应用界面。算法层面，设计了基于 Deepface 的面部情绪识别、基于文心一言 API 的语音情感分析，设计了动态权重决策级多模态融合算法，并引入基于数据量的多模型情绪趋势预测方法。同时，利用提示词优化了大模型在情感支持对话任务中的表现。

设计完成后，完成了系统硬件平台的搭建与调试，并对软件系统进行了全面的功能测试与验证，包括白盒测试与黑盒测试。测试结果表明，系统各模块运行稳定，功能符合设计要求，能够有效整合多模态数据进行学生情感状态分析与监测，验证了该设计的可行性与有效性。

关键词：多模态情感分析；学生情感监测；ESP32；大语言模型；数据融合

1 绪论

1.1 背景和意义

青少年心理健康发展是我国教育现代化进程中的重要议题。教育部等十七部门联合印发的《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）》明确指出，要建立动态监测机制，加强人工智能等新技术在心理健康管理中的应用创新^[1]。随着深度学习与人工智能技术的快速发展，多模态情感分析系统因其在心理健康监测、人机交互等领域的广泛应用前景而受到广泛关注^[2]。传统的仅基于文本或面部表情的单模态情感分析方法往往难以全面捕捉用户的真实情绪状态，而多模态情感分析能够整合面部表情、语音、生理信号等多种信息，通过不同模态间的互补与验证，显著提升情感识别的准确性^[3]。特别是在教育领域，学生的情绪状态直接影响其学习效率与心理健

康, 如何实时、准确地监测并分析学生的情感变化, 已成为当前智能教育研究的重要方向之一^[4]。

当前, 我国教育信息化建设持续推进, 智慧校园的普及对个性化学习与心理健康管理提出了更高要求^[5]。然而, 传统的学生情绪监测主要依赖问卷调查或教师主观观察, 不仅时效性差, 且容易受到主观因素干扰, 难以实现大规模、精准化的情感状态评估^[6]。此外, 现有的商业化情感计算系统通常依赖高性能计算设备, 成本较高, 难以在普通校园环境中推广^[7]。针对这一问题, 本设计提出了一种基于 ESP32 和 ESP32S3 硬件与轻量化多模态融合算法的学生情感智能分析系统, 通过结合面部表情识别、语音情感分类和心率分析, 在保证实时性的同时降低计算资源需求, 为教师、家长和学生提供高效的情绪监测与干预工具。

本系统的实施具有重要的现实意义。对教师而言, 系统提供的班级情绪分布、异常预警及情绪-成绩相关性分析功能, 可帮助其及时发现学生的心理波动, 优化教学策略; 对家长而言, 通过实时获取子女的情绪变化趋势, 能够更主动地参与心理健康管理; 对学生而言, 系统集成的大模型情绪助手可提供即时情感支持, 缓解焦虑等负面情绪。此外, 系统的硬件设计和多模态融合方法采用低成本方案, 便于在资源有限的校园环境中部署, 为推动情感计算技术在教育领域的普惠化应用提供了可行的技术路径。

1.2 国内外现状

1.2.1 国外现状

近年来, 随着情感计算技术的快速发展, 多模态情感识别系统在国际学术领域已成为备受关注的研究热点^[8]。欧美等发达国家在该领域的研究起步较早, 已形成较为完善的技术体系和应用场景^[9]。在技术应用层面, 国外学者普遍采用深度学习与多模态融合策略, 通过整合面部表情、语音信号和生理特征等多种信息, 显著提升了情感识别的准确性^[10]。美国麻省理工学院媒体实验室开发的 Affective Computing 系统^[11]率先将计算机视觉与语音情感分析相结合, 实现了对用户情绪的实时监测, 该系统已成功应用于心理健康评估和人机交互等领域。韩国电子通信研究院提出的 Modality-Aware Affect Network 模型^[12], 通过轻量化网络结构设计, 在保证实时性的同时降低了计算资源消耗, 为嵌入式设备上的情感识别提供了重要技术参考。

在研究方向上, 国际学术界不仅关注多模态情感识别的技术实现, 更注重其与实际应用场景的结合。英国剑桥大学情感智能研究中心通过构建大规模多模态情感数据集, 深入探究了不同模态信息在情绪表达中的互补关系, 并开发出基于注意力机制的特征融合算法, 显著提升了跨模态情感表征能力^[13]。GAN 等人^[14]则将情感计算技术与教育场景深度融合, 其研发的智能教室系统通过分析学生的面部表情和语音特征, 为教师提供实时课堂情绪反馈, 有效提升了教学互动质量。这些跨学科的研究方法为多模态情感识别在教育领域的应用奠定了重要基础。

随着边缘计算和物联网技术的发展, 国外学者也开始探索轻量化多模态情感识别系统的部署方案。韩国蔚山国立科学技术研究所开发的便携式情感监测设备^[15], 通过优化神经网络模型和传感器数据融合算法, 实现了在低功耗嵌入式平台上的实时情绪分析。印度德里大学则着重研究多模态数据的隐私保护问题, 提出基于联邦学习的分布式情感识别框架^[16], 为教育场景中学生隐私数据的安全处理提供了新思路。

然而，国外在多模态情感识别领域的研究仍面临挑战。现有系统多依赖高性能计算设备，难以在资源有限的普通教育环境中大规模推广。尽管如此，国外在多模态情感识别领域形成的技术积累和应用经验，对推动我国智能教育领域的情感计算研究仍具有重要的借鉴价值。

1.2.2 国内现状

我国教育场景下的多模态情感识别研究始于 21 世纪初，早期主要集中于单一模态的情感分析技术，如基于面部表情或语音的情感识别^[17]。这一阶段的研究受限于当时的人工智能技术水平和硬件计算能力，识别准确率和实时性都难以满足实际教学需求。相较于同期国外已较为成熟的多模态情感计算系统，我国在该领域的研究明显滞后，处于技术探索的初级阶段^[18]。

随着 2010 年后深度学习技术的突破性发展，国内学者开始尝试将多模态融合技术引入教育情感识别领域。程二丽^[19]开发的课堂情感分析系统，采用改进的 LBP 算法实现人脸检测进行表情识别，首次实现了对课堂教学中学生五种基本情感状态的实时监测。这一创新标志着我国教育情感识别研究从理论走向教室的转变。然而，该系统仍存在情感识别太过单一的问题，难以起到泛化作用。

2018 年以来，物联网技术和边缘计算的兴起为我国教育情感识别研究注入了新的活力。张新新^[20]研发的基于轻量化神经网络的教育机器人系统，通过混合注意力机制实现了表情和姿态的双模态识别，在嵌入式设备上取得了较好的实时性能。这一时期的研究开始关注实际应用场景的需求，但受限于传感器精度和算法复杂度，多数系统仍难以实现精准的多模态数据同步和高效的特征融合。

近年来，随着大语言模型等人工智能技术的快速发展，我国教育情感识别研究进入了新的阶段。詹勇鑫^[7]提出的基于二元模态表示学习与融合网络的多模态情感分析方法，通过跨模态注意力机制实现了教学视频和文本评论的情感倾向分析，为课堂评价提供了新的技术手段。同时，国内多所高校和研究机构也在积极探索多模态情感识别技术在教育领域的应用，如基于视觉-语音双模态的在线学习情绪监测系统、融合生理信号的学习专注度评估系统等。

尽管取得了诸多进展，当前国内教育场景下的多模态情感识别研究仍面临诸多挑战。在技术层面，现有系统的多模态数据融合算法复杂度高，难以在嵌入式设备上实现实时处理；在应用层面，情感识别结果与教学实践的有机结合不够深入，缺乏有效的反馈干预机制；在硬件层面，数据采集设备的稳定性、兼容性和成本问题仍有待解决。特别是在基础教育领域，适合大规模部署的低成本、高可靠解决方案仍然稀缺。

随着国家教育数字化战略的推进和新一代人工智能发展规划的实施，多模态情感识别技术正在成为智能教育发展的重要支撑。未来研究需要着力突破轻量化算法设计、多模态数据高效融合、边缘计算部署等关键技术瓶颈，推动教育情感识别系统向实用化、普及化方向发展。这些挑战也为本研究的系统设计提供了明确的技术攻关方向和创新空间。

本毕业设计以 ESP32 和 ESP32S3 开发板为核心硬件平台，采用双主板设计方案，融合多模态数据采集与大模型技术，构建一套面向教育场景的学生情感智能分析与监测系统。系统通过集成摄像头、语音模块和心率传感器等多源感知设备，基于轻量化的多模态情感融合算法，实现对学生的实时情感状态分析。在软件层面，系统采用分层架构设计，为教师、家长和学生分别提供定制化的情感可视化界面和分析功能，同时结合文心一言 API 实现情感对话辅助功能，旨在提升教育教学过程

中的情感交互质量和心理健康监测水平。

1.3 设计内容

立足于当前教育信息化发展需求与学生心理健康管理的现实挑战，本设计以物联网技术为核心支撑，设计并实现了一套基于 ESP32 和 ESP32S3 开发板的多模态情感智能监测与分析系统。该系统通过集成化的硬件终端采集学生的面部表情、语音信号及生理指标，构建了面向教师、家长和学生的智能化情感分析平台，为校园心理健康管理提供了一套完整的解决方案，具体设计内容如下：

(1) 本系统采用双主板设计，主板 1 以 ESP32 开发板作为主控芯片，承担情绪监测与反馈功能。系统整合了 ESP32CAM 摄像头模块进行面部图像采集，通过 WiFi 将图像数据传输至电脑端 Python 程序，基于 opencv、matplotlib 和 deepface 进行面部情绪识别分析，能够判断人脸表情所反映的情绪状态。主板 1 还整合了心率检测传感器采集生理参数，并配备 WS2812 彩色灯珠和震动马达提供情绪反馈，ASRPRO 模块负责语音唤醒与播报。系统通过 WiFi 网络实现数据传输，ESP32CAM 将图像信息和处理结果通过 HTTP 协议与电脑端 Python 程序交互，并通过串口与 ESP32 主控制板进行通信。

(2) 主板 2 以 ESP32S3 作为控制核心，专注于大模型对话功能实现。该部分整合了 INMP441 I2S 数字声音传感器用于声音数据采集，MAX98357 I2S 音频放大器模块与喇叭组合提供语音输出能力，并配备 1.8 寸 128×160 TFT 显示屏用于 emoji 表情显示，构建完整的人机交互界面。主板 2 通过 WiFi 连接互联网，访问基于文心一言的 API 服务，实现智能对话处理，为用户提供自然流畅的语音交互体验，并通过显示屏呈现相应的 emoji 表情，增强交互的情感表达能力。

(3) 软件系统基于 Node.js 平台开发，采用分层架构设计。系统分为教师端、学生端和家长端三个模块，教师端可查看班级情绪分布、历史趋势及异常预警；学生端提供个人情绪分析和大模型对话功能；家长端则关注子女情绪变化。

(4) 本系统采用的语音处理流程，通过 Node.js 的 fs.watch 实时监控新增 WAV 文件，触发百度语音识别 API 进行文本转换，转写结果后调用文心一言 API 进行语义情感分析，回传设计好的即 7 种情绪的置信度 JSON 格式数据。既保证了语音识别的准确性，又降低了网络传输负载。

(5) 本系统采用现代化的前端技术栈，构建了一个响应式、交互性强的 Web 应用界面。在框架选择上，采用轻量级的 Tailwind CSS 作为样式框架，实现了高度可定制的 UI 组件系统。采用模块化的组件设计思想，将功能按照教师、学生、家长三个角色进行解耦，每个角色都有独立的功能模块和数据展示界面。在数据可视化方面，系统集成了图表组件，用于展示情绪分析的趋势数据和统计信息。

(6) 本系统实现的关键算法包括多模态数据融合算法，通过动态权重方式整合面部、语音和心率数据；大模型交互接口调用文心一言 API，配合预设的提示词（Prompt）实现情感化对话；情绪预测模块根据记录数量采用不同算法，少于 50 条使用移动平均法，50-200 条采用线性模型，超过 200 条则应用随机森林预测。

本设计通过物联网终端与服务端的协同工作，构建了一个完整的情绪监测与分析系统，为校园心理健康管理提供了可行的技术实现方案。

1.4 论文结构安排

本论文围绕上述研究内容，共划分为七章展开论述，其具体的结构安排如下：

第一章：绪论。开篇先介绍了学生情感智能分析与监测系统的研究背景和意义，阐述其在促进教育信息化和提升学生心理健康管理中的重要作用。接着对国内外在多模态情感分析技术的研究现状进行综述，分析当前研究的进展与不足。最后提出本论文的研究内容和整体结构安排。

第二章：学生情感智能分析与监测系统总体方案设计。本章从实际应用角度分析了技术实现可能性、经济合理性及操作便捷性。通过对教育场景的观察，确立了系统的功能边界与性能标准，并基于此提出了模块化、开放性和交互友好的设计理念，最终形成了硬件与服务端结合的系统架构。

第三章：系统硬件设计。本章详细阐述了情感智能分析与监测系统的硬件设计方案。首先，介绍系统的双主板架构设计，随后分别介绍 ESP32 主控板模块和 ESP32S3 大模型交互板模块，包括主控芯片和各传感器模块的选择与连接。最后设计了系统通信与数据处理方案，确保多模态数据的有效采集与传输。

第四章：学生情感智能分析系统算法设计与实现。本章主要介绍系统核心算法的设计与实现过程。首先，设计多模态情感指数算法，包括基于 Deepface 的人脸情绪指数算法和基于文心一言的语音情绪指数算法，以及多模态融合决策算法。接着，设计多模型融合的情绪指数预测算法，实现情绪趋势的预测与分析。最后，介绍硬件与服务端交互子系统设计 and 情绪阈值控制机制。

第五章：学生情感智能分析软件端设计。本章主要介绍系统软件端的设计和开发。首先，介绍开发环境与技术栈选择。接着，详细设计系统的功能模块，包括教师端、学生端和家长端的差异化功能设计。最后，进行系统数据库的概念设计、逻辑设计和表设计，确保数据的有效存储与管理。

第六章：系统调试与分析。本章对学生情感智能分析与监测系统进行全面的调试与分析。首先，进行硬件系统的安装和调试，确保各模块正常工作。接着，对软件系统进行白盒测试和黑盒测试，验证情感分析、多模态融合和预测算法的有效性。最后，对系统核心功能进行测试，包括登录注册、大模型情感分析、数据导出和 AI 聊天助手等功能，确保系统的可靠性和稳定性。

第七章：总结和展望。本章对全文进行总结，梳理研究成果和实现功能。同时，分析论文研究中存在的不足和局限性，提出改进方向和未来研究展望。通过对学生情感智能分析与监测系统未来发展方向的探讨，为教育信息化和情感计算领域的研究提供借鉴和参考。

2 学生情感智能分析与监测系统总体方案设计

本章基于学生情感状态识别的实际需求，从可行性论证、需求解析到系统设计逐层递进，构建了一套完整的轻量化情感分析与监测系统方案。在可行性层面，重点阐释了基于 ESP32 硬件平台与多模态数据融合的技术实现路径，兼顾了部署成本与操作效率的平衡；在需求分析阶段，系统化梳理了多模态数据同步采集、高精度实时情感计算、动态情绪可视化及家校协同干预等功能模块，并确立了性能与安全的量化指标；最终通过感知层、网络层、应用层的模块化架构设计，实现了硬件资源优化与算法效能的协同提升。

2.1 系统可行性分析

2.1.1 技术可行性

本系统设计的情感智能分析系统采用 ESP32 和 ESP32S3 嵌入式技术、多传感器融合技术、Wi-Fi 无线通信技术、Node.js 后端开发技术以及大模型 API 集成技术等，这些技术均属于当前物联网和人工智能领域的主流技术方案^[21]。系统采用双主板架构，分别基于 ESP32 和 ESP32S3 微控制器开发。Deepface^[22]作为轻量级面部表情识别库，已在多个边缘计算场景中得到验证；百度语音识别和文心一言 API 提供了成熟的自然语言处理服务，为系统的情感分析奠定了技术基础^[23]。硬件平台选用 ESP32 和 ESP32S3 芯片，它们具备充足的运算能力和丰富的外设接口，完全满足多传感器数据采集需求。Web 前端采用响应式设计，基于现代浏览器技术实现跨平台访问。从整体技术架构来看，该系统所采用的技术路线具有充分的技术可行性。

2.1.2 经济可行性

本系统采用分层建设方案，硬件部分基于开源硬件平台设计，核心传感器模块均为商业化成熟产品，具有较高的性价比。软件层面充分利用开源技术栈，Node.js 后端和 Web 前端开发无需额外授权费用。云端服务采用按量付费模式，通过百度智能云平台提供的语音识别和大模型服务，有效控制了 API 调用成本。与传统的心理健康评估方式相比，本系统能够实现自动化情绪监测，显著降低人工评估成本。因此，本系统具有充分的经济可行性。

2.1.3 操作可行性

系统采用“终端设备+Web 应用”的交互模式，硬件终端设计为即插即用型设备，用户仅需完成简单配置即可投入使用。Web 应用界面根据教师、学生和家長三类用户角色进行了针对性设计，采用直观的数据可视化方案和简洁的操作流程。教师端提供班级整体的情绪报告可视化分析功能，家长端可实现子女情绪状态的即时查询，学生端能够通过对话式大模型完成情绪疏导。本系统无需用户具备过多的专业知识，操作门槛显著低于专业心理评估工具。因此，本系统具有充分的操作可行性。

综上所述，本情感智能分析系统在技术实现、经济成本和操作便捷性三个方面均具备充分的可行性。通过该系统的实施，能够为校园心理健康管理提供智能化解决方案，实现对学生情绪状态的实时监测和科学干预，具有良好的应用前景。

2.2 系统需求分析

2.2.1 总体需求分析

当前，学校对学生心理健康状态的监测主要依靠传统的问卷调查和心理教师评估，这种方式存

在反馈滞后、主观性强等问题^[24]。为提高学生心理健康管理的科学性和时效性，本情感智能分析系统的设计与实现需要满足以下总体需求：

(1) 系统需实现对学生情绪的实时监测与智能分析。通过多模态数据采集设备，系统应当能够持续捕捉学生的面部表情、语音特征和心率变化等生理指标，并运用人工智能技术对这些数据进行综合分析。在技术实现上，系统需要具备 7 种基本情绪（愤怒、厌恶、恐惧、快乐、悲伤、惊讶、平静）的识别能力，并将分析结果通过可视化界面直观呈现。

(2) 系统应建立完善的数据管理与预警机制。对于教师用户，系统需要提供班级整体情绪状态的可视化展示，包括情绪分布热力图、历史变化趋势图等。同时，当检测到学生出现异常情绪时，系统应当能够及时生成预警信息。对于家长和学生用户，系统需要提供个性化的情绪报告和查询功能。

(3) 考虑到实际应用场景，系统设计要注重操作的便捷性和隐私保护。软件界面需要针对不同用户角色进行差异化设计，确保教师、家长和学生都能快速掌握使用方法。同时，系统需要建立严格的数据访问权限控制机制，保护学生隐私信息的安全。

为了更精准地剖析系统需求，本文绘制了系统用例图。这些用例图详细划分了系统的不同功能模块，具体包括单片机硬件系统用例，如图 2.1 所示、情感智能分析系统学生端用例，如图 2.2 所示、情感智能分析系统教师端用例，如图 2.3 所示以及情感智能分析系统家长端用例，如图 2.4 所示。系统的整体设计围绕这些用例图展开，确保功能实现与需求契合。

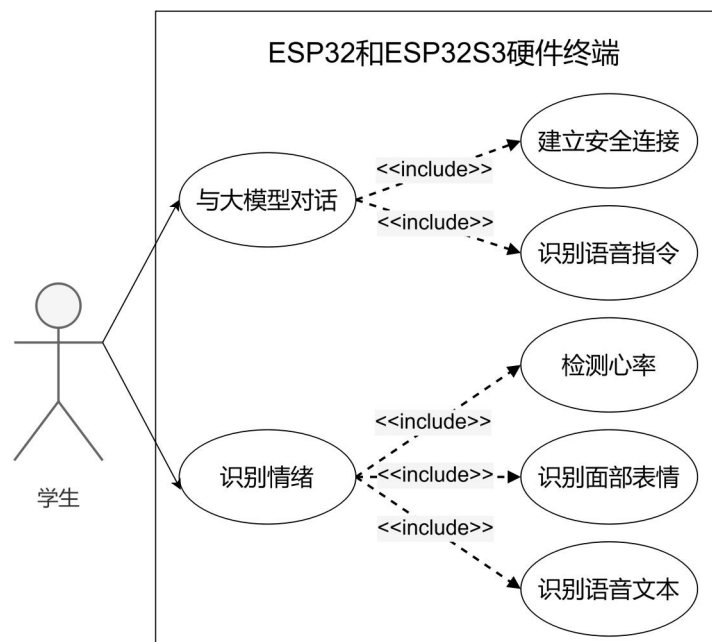


图 2.1 单片机硬件系统用例图

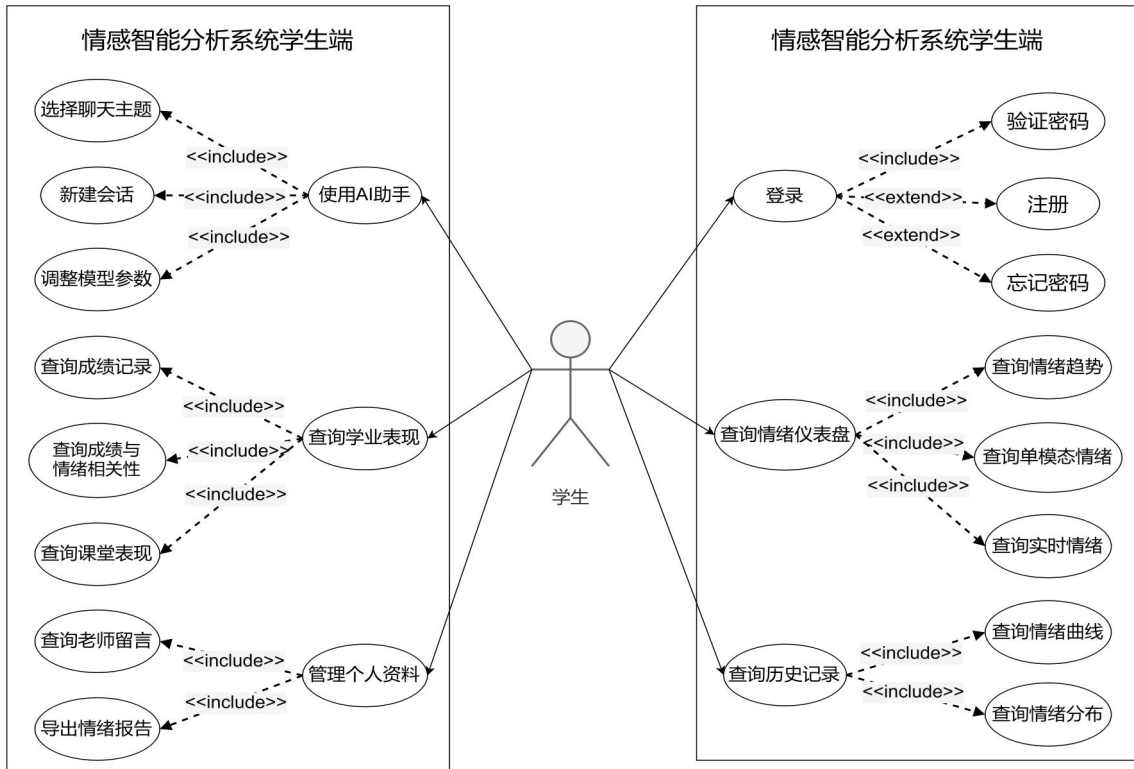


图 2.2 情感智能分析系统学生端用例图

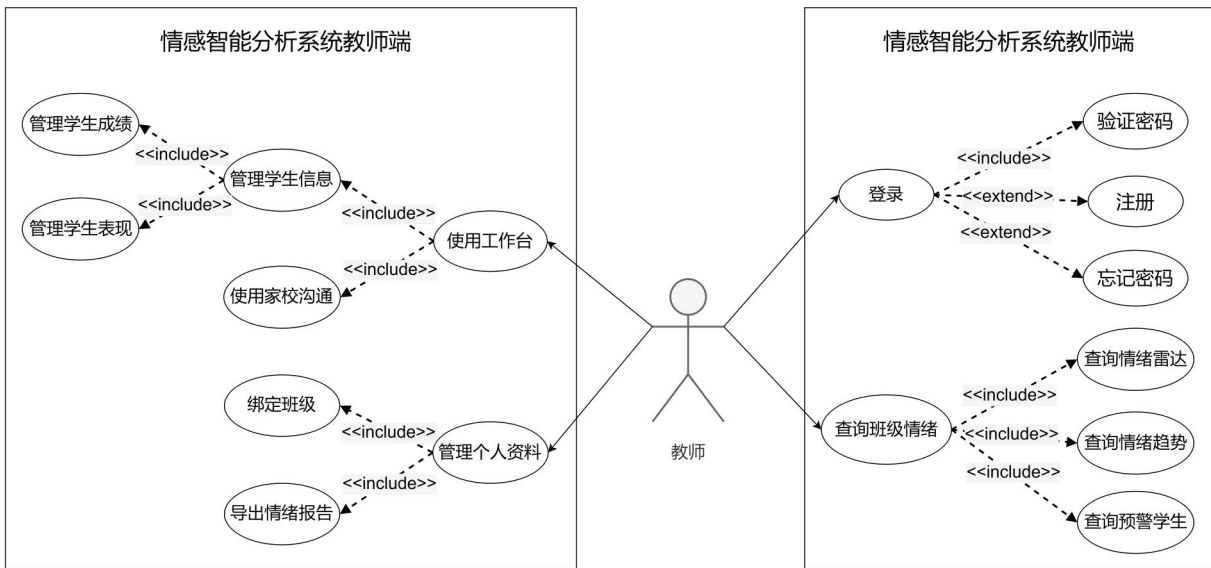


图 2.3 情感智能分析系统教师端用例图

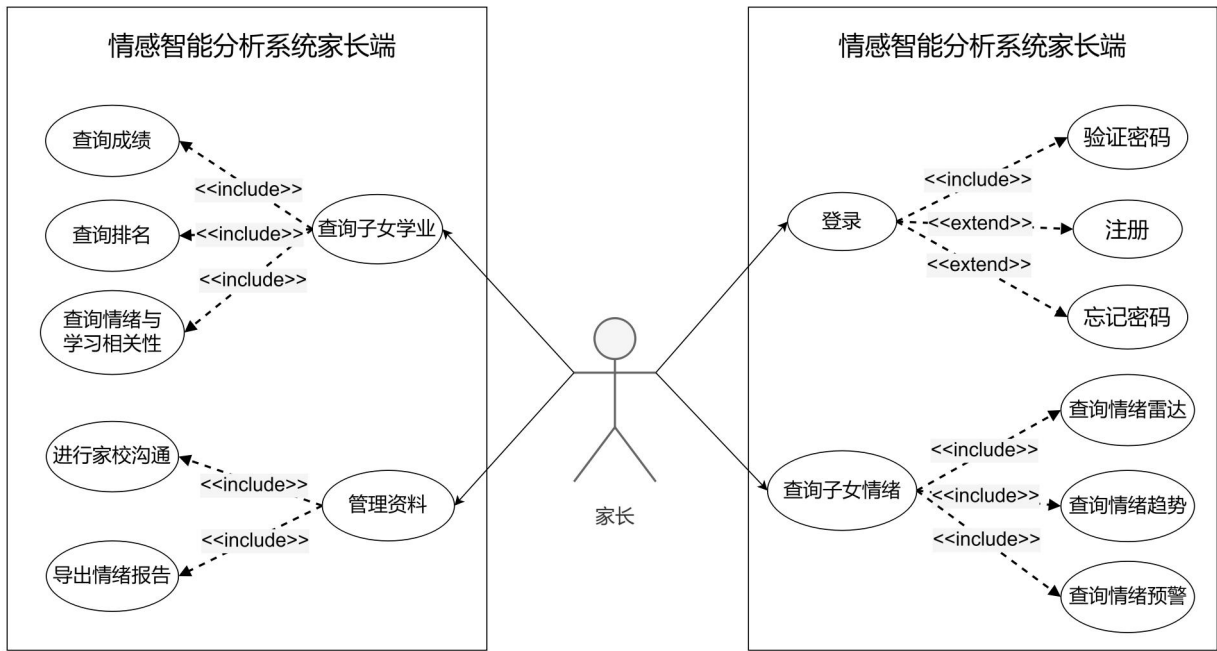


图 2.4 情感智能分析系统家长端用例图

2.2.2 功能性需求分析

参照教育智能化与学生心理健康管理的实际需求，结合多模态情感分析与大语言模型的技术发展趋势，学生情感智能分析与监测系统应集成以下核心功能：

1、多模态情感数据采集与传输

系统应具备利用多种传感器实时捕获学生情感状态参数的能力，包括通过 ESP32CAM 采集面部表情数据、通过 INMP441 麦克风采集语音数据以及利用心率传感器获取生理数据。系统应建立稳定的 WiFi 传输机制，确保数据能够安全可靠地上传至服务端进行分析处理。用户可以通过 Web 应用界面实时查看情感分析结果，了解个人或班级的情绪状态。

2、多模态情感融合与分析

系统应设计完善的多模态情感融合算法，能够将面部表情、语音内容和心率数据进行综合分析，通过动态权重机制适应不同情境下的情感评估需求。系统应能识别七种基本情绪（愤怒、厌恶、恐惧、快乐、悲伤、惊讶、平静），并提供情绪强度的量化评估，为教学管理提供科学依据。

3、情绪状态可视化与监测

系统应提供多维度的情绪数据可视化界面，包括实时情绪仪表盘、历史情绪趋势图、情绪分布雷达图等，使用户能够直观了解情绪变化。系统应能对情绪数据进行预警分析，当检测到学生情绪异常波动时，及时向教师或家长发出提醒，支持提前干预。

4、大模型情绪对话与支持

系统应整合文心一言大模型 API，构建针对学生心理需求的情感对话助手。该功能应支持语音和文本双模式交互，能够识别学生的情绪状态，并提供相应的情感支持和建议，帮助学生缓解负面情绪，促进心理健康。

5、多角色协同管理

系统应建立教师、学生和家長三类用户角色的差异化功能模块，教师端需具备班级情绪概览、

学生管理、课程安排及工作台等功能；学生端需提供个人情绪监测、历史记录、AI 助手及个人中心等功；家长端则应支持孩子情绪监护、学业表现查看及家校沟通等功能。各角色间应建立信息共享与沟通机制，促进家校协同育人。

2.2.3 非功能性需求分析

除了确保符合功能性要求外，系统还需关注并满足以下非功能需求：

1、实时性需求

系统应具备良好的实时处理能力，情感分析结果的响应时间应控制在 2 秒以内，以确保用户能够获得即时的情绪反馈。大模型对话功能的响应时间应控制在 5 秒以内，保证良好的交互体验。

2、准确性需求

系统的情感识别准确率应不低于 50%，特别是在识别明显的情绪变化时，应具备较高的敏感度和精确度。多模态融合算法应能有效处理不同模态间的互补与冲突，提升整体识别性能。

3、可扩展性需求

系统架构应采用模块化设计，便于未来增加新的情感识别模型或扩展功能模块。数据库设计应考虑长期数据积累的需求，支持大规模学生情感数据的存储与分析。

4、安全性需求

系统应建立严格的用户权限管理机制，确保学生的情感数据只能被授权人员访问。数据传输过程应采用加密措施，防止隐私信息泄露。同时，系统应遵循教育领域的的数据伦理规范，保护未成年人的隐私权。

5、易用性需求

考虑到系统使用者包括教师、学生和家，界面设计应简洁明了，操作流程应直观易懂。系统应提供必要的使用指导和帮助文档，降低用户学习成本。

2.3 系统设计原则

在设计学生情感智能分析与监测系统时，我们需遵循一系列原则，以确保系统的可靠性、高效性、易用性、安全性以及可扩展性，同时满足用户对学生情感实时监测的需求。

以下是系统设计时的主要原则：

1、多模态融合与精准分析原则

多模态数据融合是情感智能分析系统的核心设计原则。系统应整合面部表情、语音语义和生理数据三种模态信息，通过动态权重融合算法消除单一模态的局限性，提升情感识别的鲁棒性和准确性^[25]。在算法设计上，应采取轻量化的决策级融合策略，既保证识别精度，又能适应教育场景下对实时性的要求。此外，系统应支持不同模态数据的时间维度对齐，处理模态缺失情况，确保多模态融合的有效性和连续性。

2、人本化与易用性原则

考虑到系统使用者包括教师、学生和家三类用户群体，系统界面设计应秉持人本化理念，将复杂的技术实现隐藏在直观的操作界面之下。对教师而言，系统应提供班级情绪分布的宏观视图和学生个体的微观分析；对学生而言，系统应强调情绪自我认知和情感支持；对家而言，系统应聚

焦孩子情绪变化与学业表现的关联。此外，系统应尽量简化硬件配置过程，采用即插即用的设计理念，降低用户使用门槛。

3、差异化与角色适应原则

针对不同用户角色的特定需求，系统应提供差异化的功能模块和数据访问权限。教师端应关注班级整体情绪态势和异常情况预警；学生端应注重个人情绪自我监测和情感疏导；家长端则应强调子女情绪与学业关联性的长期跟踪。各角色间的数据流转应严格遵循权限控制原则，确保学生隐私得到适当保护的同时，又能支持必要的信息共享与家校协同。

4、实时性与轻量化原则

学生情感状态的监测要求系统具备良好的实时性能。在硬件选择上，ESP32 和 ESP32S3 系列微控制器能够满足边缘计算的基本需求；在算法设计上，应优先采用计算复杂度较低的轻量化模型，如 DeepFace 面部情绪识别和基于大模型 API 的语音情感分析；在系统架构上，应采用边缘计算与云端服务协同的策略，将初步数据处理放在边缘端完成，复杂计算任务交由云端服务处理，从而在保证性能的同时降低硬件成本。

5、安全性与隐私保护原则

学生情感数据属于高度敏感的个人信息，系统设计必须高度重视数据安全与隐私保护。应采用多层次的安全防护措施，包括用户身份验证、数据传输加密、访问权限控制等机制。敏感数据的处理应尽可能在本地完成，减少不必要的网络传输。当需要调用云端服务如大模型 API 时，应确保数据匿名化处理，避免个人识别信息的泄露。同时，系统应遵循最小化原则，只收集必要的学生情感数据，并制定明确的数据保留策略。

6、可扩展性与可维护性原则

为适应教育场景需求的不断变化，系统应具备良好的可扩展性和可维护性。软件架构应采用模块化设计，各功能模块之间通过明确定义的接口进行交互，便于功能扩展和模块替换。数据库设计应考虑长期数据积累的需求，支持未来的数据挖掘与分析。同时，系统应提供完善的日志记录和监控机制，便于问题定位和系统维护。对于硬件部分，应预留足够的接口资源，支持未来传感器或执行器的扩展需求。

7、双向互动与反馈原则

有效的情感智能分析系统不应仅限于单向的数据收集和分析，还应提供丰富的反馈机制。系统设计应注重双向互动，通过灯光、声音、震动等多种感官方式向用户传递情绪状态反馈；通过大模型驱动的智能对话系统，为学生提供情感支持和心理疏导；通过可视化图表和数据分析，帮助教师和家长了解学生情绪变化规律，进行有针对性的干预。这种闭环的情感感知-分析-反馈机制，是系统发挥实际价值的关键。

2.4 系统总体方案设计

本系统选择 ESP32 和 ESP32S3 芯片作为学生情感智能分析与监测系统的微控制器核心，该系统由硬件组件和软件程序两部分组成。

硬件部分包括 ESP32 及 ESP32S3 微控制器、ESP32CAM 摄像头、INMP441 麦克风、MAX30102

心率传感器、WS2812 彩色灯带、震动马达、MAX98357A 音频解码器及 TFT 显示屏等。

软件部分包括 Node.js 服务端平台、前端 Web 应用以及大模型 API 接口调用模块。系统采用双主板分布式架构，ESP32 主控板负责情绪识别与反馈，ESP32S3 主控板专注于大模型对话功能。通过 WiFi 网络，硬件终端将采集的多模态数据传输至服务端进行处理分析，实现对学生情感状态的实时监测与评估。用户可通过 Web 应用界面查看情感分析结果，并获取相应的反馈与支持。Web 应用层包括了不同角色的三个客户端，分别是教师端、学生端和家长端，每个角色持有不同的功能权限，以保障学生的隐私。

按照物联网体系，本系统又可进一步划分为感知层、网络层和应用层。整体架构如图 2.5 所示。

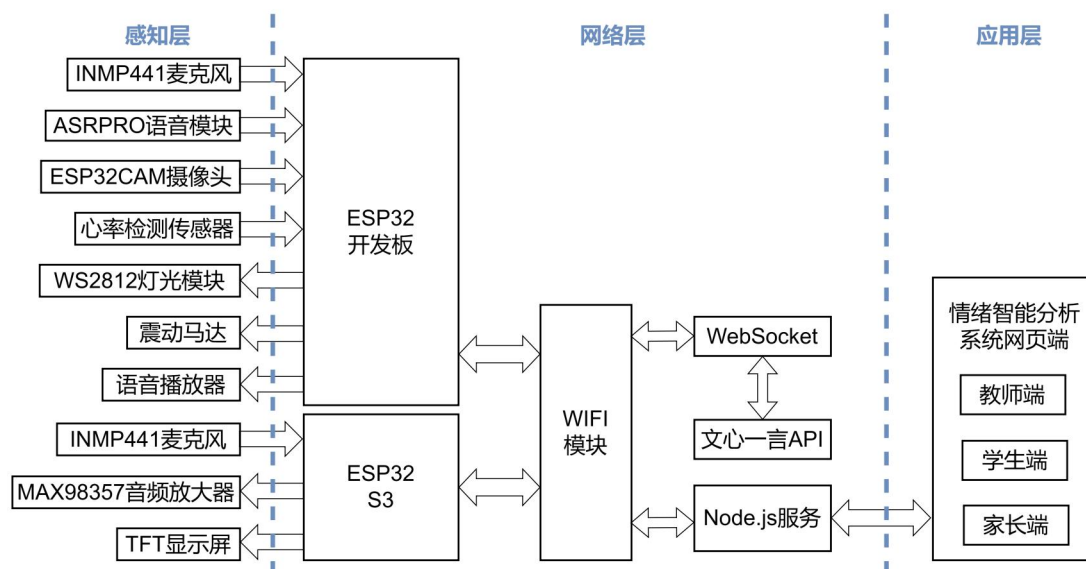


图 2.5 系统整体物联网体系架构

2.4.1 感知层

在本系统中，感知层集成了 ESP32CAM 摄像头模块、INMP441 麦克风、MAX30102 心率传感器以及 WS2812 彩色灯带等多种感知元件，它们的核心任务在于捕获学生的面部表情、语音信号及生理指标，并借助通信硬件将这些多模态数据传输至网络层，构建起与网络层及应用层之间的物理链路。

2.4.2 网络层

网络层以 Node.js 服务端平台为核心载体。通过 ESP32 和 ESP32S3 的 WiFi 功能，感知层所获取的多模态数据被无线上传至服务端。服务端承担着对感知层数据的多模态融合分析任务，并通过调用百度文心一言 API 实现智能对话功能。此外，经过服务端处理的情感分析结果能够顺畅流转至应用层，同时为情感反馈机制提供决策依据。

2.4.3 应用层

应用层基于 Web 技术开发了多角色用户界面，为教师、学生和家长提供差异化的功能模块。教师端可查看班级情绪分布、历史趋势及异常预警；学生端提供个人情绪分析和大模型对话功能；家长端则关注子女情绪变化。通过直观的数据可视化和便捷的操作界面，用户可实时掌握情绪状态并获取相应支持。

2.5 系统总体功能设计

由于功能模块化设计使系统整体架构清晰、扩展性增强，本系统采用模块化设计方法，面向教师、学生和家長三类用户群体构建差异化功能体系。

硬件系统基于 ESP32 双主板架构，完成多模态数据采集与实时反馈；服务端系统通过 Node.js 实现数据处理与分析，并开放 Web 应用接口；Web 前端系统针对不同角色提供定制化界面，支持情绪监测、智能交互与家校协同功能。教师端可通过班级情绪概览、学生管理、课程管理等功能实现学生情感状态的可视化分析与教学干预；学生端集成个人情绪仪表盘、历史记录查询及 AI 情感对话助手，支持情绪自我认知与疏导；家長端则专注于子女情绪监护、学业表现追踪及家校沟通，形成闭环心理健康管理机制。

系统中各功能模块通过数据接口实现互联，在保证系统鲁棒性的同时降低了新增功能的开发复杂度。系统总体功能模块如图 2.6 所示，其分层设计为教育场景下的情感计算提供了可扩展的技术框架。

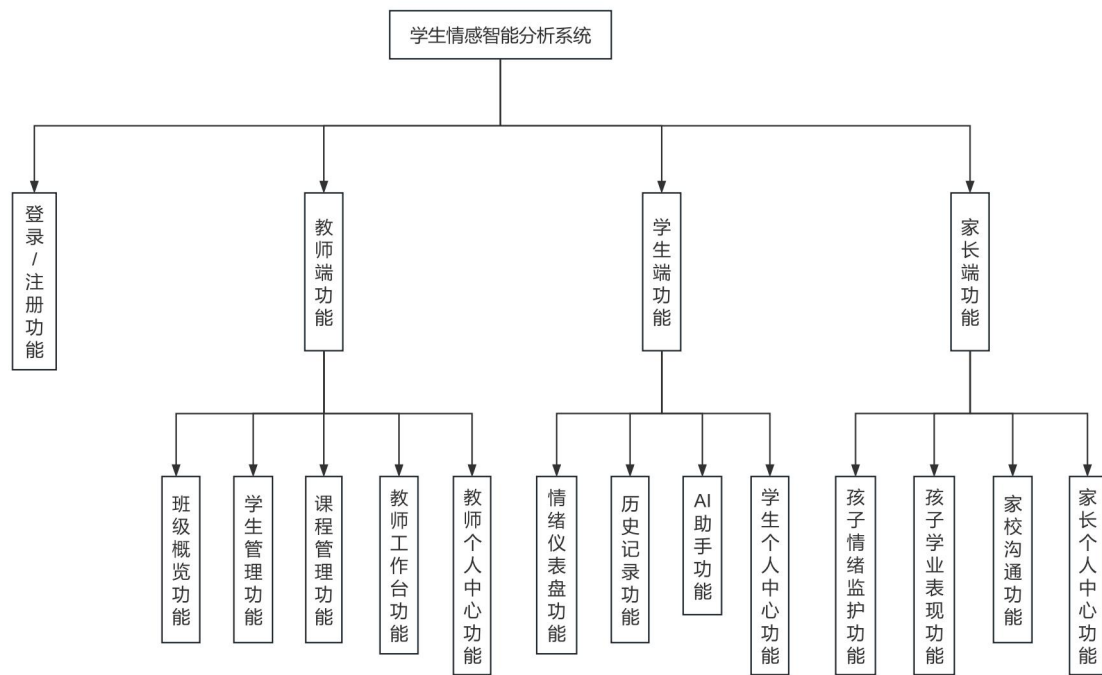


图 2.6 系统总体功能概览

2.6 本章小结

本章从可行性分析、需求分析和系统设计三个层面构建了学生情感智能分析与监测系统的总体方案。首先，从技术、经济及操作性三个维度论证了系统实施的可行性，确立了基于 ESP32 平台与多模态融合技术的轻量化架构。其次，通过详细的功能与非功能需求分析，明确了系统需实现多模态数据采集、实时情感识别、情绪可视化及家校协同等核心功能，并满足准确性、实时性与安全性等关键指标。最后，基于模块化与角色差异化原则，提出了感知层、网络层与应用层的三层次系统架构，为后续硬件设计与算法实现提供了框架支撑。本章工作为系统的具体开发与实现奠定了理论基础和技术路线。

3 系统硬件设计

本章是情绪识别和交互系统的核心环节，基于系统总体设计，本文将详细阐述所采用的硬件架构及各功能模块的设计方案。本系统硬件主要分为两大核心部分：情绪感知与交互主板 ESP32 和大模型对话主板 ESP32S3。

3.1 硬件架构设计

本系统主要实现三个核心功能：实时情绪识别、生理参数监测及大模型智能对话。系统采用双主板分布式设计架构，实现功能模块化与处理任务的合理分配。第一主板以 ESP32 为核心，搭配 ESP32CAM 模块构建视觉情绪识别系统；第二主板以 ESP32S3 为核心，专注于语音交互与大模型对话功能。这种架构设计不仅确保系统同时处理多种复杂任务，还使得各个功能模块能够并行工作，提高系统整体响应速度。系统与云端通过 Node.js 框架构建的服务进行数据交换，接入百度文心大模型 API 实现智能对话功能。硬件架构设计如图 3.1 所示。

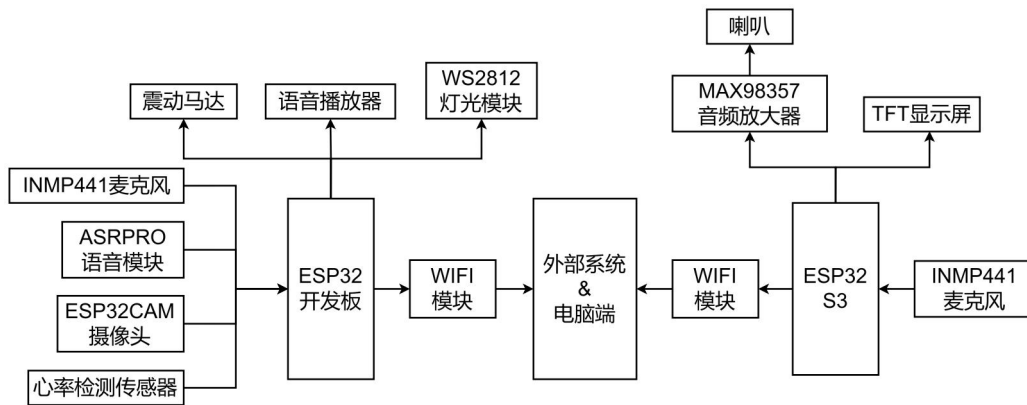


图 3.1 硬件架构设计图

3.2 ESP32 主控板模块

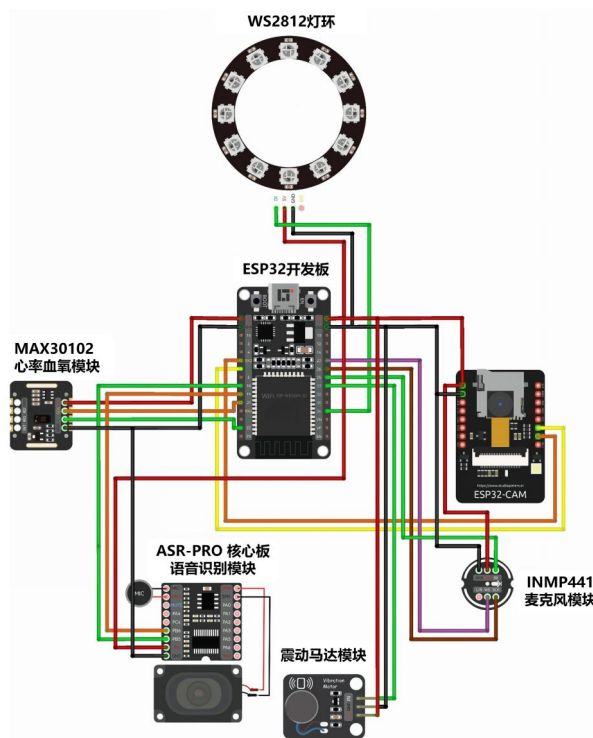


图 3.2 主控板 1 的串口连接图

考虑到情绪识别与反馈系统的功能需求和实时性要求，ESP32 主控板模块由主控芯片、摄像头模块、情绪反馈模块、语音识别模块、心率检测模块及数据传输模块组成。主控板 1 的串口连接如图 3.2 所示。

3.2.1 主控芯片简述

在情绪识别交互系统的设计中，主控芯片的选择对系统性能至关重要。作为系统的核心处理单元，主控芯片需要同时处理图像采集、情绪分析算法运算、传感器数据融合以及外设控制等任务。本系统选用 ESP32 作为第一主板的主控芯片，它是一款集成度高、功能强大的双核微控制器，具有丰富的外设接口与通信协议支持。ESP32 采用 Tensilica LX6 双核架构，工作频率高达 240MHz，内置 520KB SRAM 和 4MB 闪存，为复杂算法处理提供了充足资源^[26]。ESP32 主要功能特点如表 3.1 所示。

表 3.1 主要功能特点

参数描述	功能特点
处理器	双核 Tensilica LX6，主频高达 240MHz
内存	520KB SRAM
闪存	4MB
WiFi	IEEE 802.11 b/g/n Wi-Fi
蓝牙	蓝牙 4.2 和 BLE
外设接口	34 个可编程 GPIO，支持 UART、I2C、SPI、I2S、PWM 等
安全特性	硬件加密加速，安全启动
工作温度	-40°C~85°C
低功耗模式	深度睡眠电流低至 5 μ A
ADC	12 位 SAR ADC，最多 18 通道

ESP32 芯片凭借其强大的计算能力、丰富的外设接口以及集成的无线通信功能，非常适合作为情绪识别与反馈系统的核心处理单元。通过合理配置 GPIO 和外设资源，ESP32 能够同时处理多路传感器信号，实现图像采集与传输，并执行情绪分析算法，满足系统的实时性要求。

3.2.2 图像采集模块

ESP32CAM 模块是本系统情绪识别的视觉输入设备，基于 OV2640 图像传感器，能够提供高质量的实时图像流。ESP32CAM 集成了 ESP32-S 芯片、200 万像素 OV2640 摄像头以及 microSD 卡槽，为情绪识别提供了可靠的图像源。该模块采用 WiFi 无线传输方式，将采集的图像数据实时传送至计算机端进行处理分析，无需复杂的数据线缆，简化了系统的物理结构。

ESP32CAM 模块具备多种图像分辨率选项，可根据情绪识别算法的需求灵活调整。同时，该模块支持 JPEG 图像压缩，降低了数据传输负担^[27]。在硬件连接上，ESP32CAM 通过串口与主控 ESP32 通信，共享识别结果并执行相应的反馈指令。ESP32CAM 模块实物如图 3.3 所示。



图 3.3 ESP32CAM 模块实物图

3.2.3 语音识别与播放模块

本系统采用 ASRPRO 语音识别模块实现语音交互功能。ASRPRO 模块集成了高性能麦克风阵列和专用数字信号处理器，能够在嘈杂环境中实现高准确度的语音识别。该模块具有本地唤醒词识别功能，通过预设的唤醒词激活系统，进入主动录音状态，大幅降低系统能耗。

ASRPRO 模块与 ESP32 通过 UART 接口通信，采用简化的 AT 指令集进行配置和控制。在录音过程中，INMP441 I2S 数字麦克风负责采集高质量音频信号，其宽广的动态范围和低噪声特性确保了录音质量。INMP441 采用 I2S 数字音频接口直接与 ESP32 连接，避免了模拟信号传输过程中的干扰和失真问题^[28]。

当系统检测到用户情绪异常时，ASRPRO 模块还能播放预设的安抚语音，通过语言方式缓解用户的负面情绪。整个语音模块的设计确保了低延迟响应和自然流畅的语音交互体验。ASRPRO 语音模块和 INMP441 麦克风实物如图 3.4 所示。

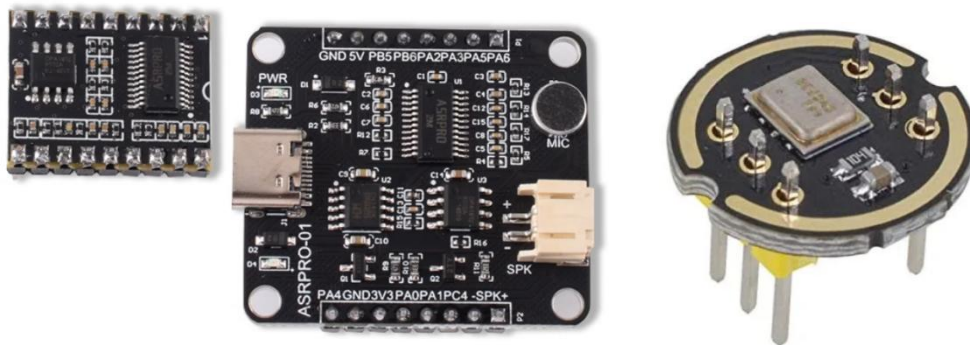


图 3.4 语音识别与麦克风模块实物图

3.2.4 心率检测模块

为全面评估用户情绪状态，本系统集成了基于 MAX30102 的心率检测模块，通过监测用户心率变化作为情绪判断的辅助依据。该模块采用先进的光容积脉搏波（PPG）技术，能够精确监测心率变化和情绪波动。工作流程如图 3.5 所示，实物图如 3.6 所示。

MAX30102 心率传感器的工作原理基于光容积法，通过利用人体组织在血管搏动时造成的透光率变化来测量脉搏和血氧饱和度。该传感器采用对动脉血中氧合血红蛋白和血红蛋白有选择性吸收特性的特定波长发光二极管作为光源。当血液在脉搏周期中流过血管时，由于动脉搏动充血容积的

变化导致光的透光率发生改变，此时由光电变换器接收经人体组织反射的光线，将其转变为电信号并放大输出，从而实现心率的准确测量^[29]。

心率传感器通过 I2C 总线与 ESP32 连接，实时传输心率数据。其中 SDA 引脚连接 ESP32 的 GPIO21，SCL 引脚连接 GPIO22，VIN 接 3.3V 电源，GND 接地。传感器支持 1.8V-5V 宽电压输入范围。ESP32 接收到传感器数据后，进行数字滤波和信号平滑处理，消除因手部运动等因素造成的干扰。通过设置 115200 的波特率，系统能够实现高精度心率数据的实时采集。

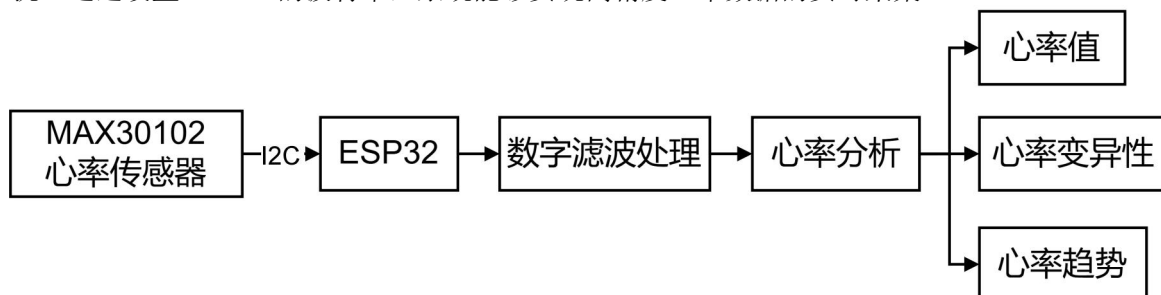


图 3.5 心率检测模块工作流程图

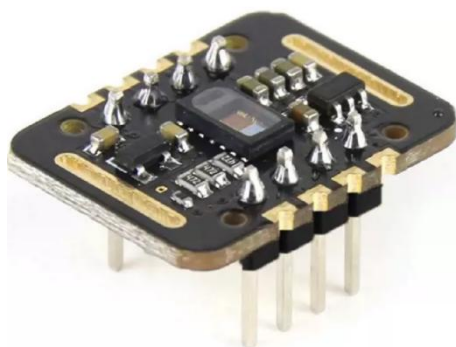


图 3.6 心率检测模块实物图

3.3 ESP32S3 大模型交互板模块

ESP32S3 大模型交互板模块作为系统的第二个处理单元，专注于提供流畅的语音交互体验和大模型智能对话功能。该模块由 ESP32S3 主控芯片、语音采集模块、音频播放模块和视觉反馈模块组成。

3.3.1 ESP32S3 主控芯片简述

表 3.2 主要功能特点

功能特点	参数描述
处理器	双核 Xtensa LX7，主频高达 240MHz
内存	512KB SRAM，8MB PSRAM
闪存	16MB
AI 加速	向量指令扩展，适合 AI 应用
WiFi	IEEE 802.11b/g/n
蓝牙	蓝牙 5.0 和 BLE
外设接口	45 个可编程 GPIO，支持 LCD、摄像头、I2S 等
安全功能	硬件加密加速，安全启动
工作温度	-40°C ~ 85°C

ESP32S3 作为第二代 ESP32 系列产品,针对人工智能和语音处理进行了专门优化。它采用 Xtensa LX7 双核处理器,主频高达 240MHz,集成了 512KB SRAM 和 8MB 闪存,并配备了专用的向量处理单元,可高效执行机器学习算法。ESP32S3 还内置了高速 PSRAM 接口,支持扩展至 16MB 外部 RAM,为大模型交互提供了充足的内存资源。

ESP32S3 的主要特点在于其强化的数字信号处理能力和 AI 加速功能,这使其特别适合语音识别和自然语言处理任务。同时,它保留了 ESP32 系列完善的外设接口和无线连接功能,包括 WiFi 802.11b/g/n 和蓝牙 5.0,确保与云端服务的稳定连接^[30]。ESP32S3 主要功能特点如表 3.2 所示。

3.3.2 语音采集与处理模块

大模型交互板采用高品质 INMP441 I2S 数字麦克风作为主要语音输入设备。INMP441 麦克风具有高达 61dBA 的信噪比和低至 -26dBFS±1dB 的灵敏度,能够准确捕捉用户的语音指令。其 20Hz-20kHz 频率响应范围确保了语音的高保真采集,为后续的语音识别提供了优质音频源。

语音采集过程中,INMP441 通过 I2S 总线将数字音频数据直接传输至 ESP32S3,这种无损数字传输方式避免了传统模拟麦克风中的信号衰减和干扰问题。ESP32S3 接收音频数据后,首先进行前置处理,包括降噪、回声消除和增益控制,然后再将处理后的音频数据打包发送至云端进行语音转文本和大模型处理。整个语音采集流程具有低延迟和高保真的特点,为用户提供自然流畅的语音交互体验。

3.3.3 音频播放模块

为了将大模型生成的回复内容转化为语音输出,本系统采用 MAX98357A I2S 音频解码放大器和配套喇叭组成音频播放模块。MAX98357A 是一款集成了 DAC、音频处理器和 D 类放大器的单芯片解决方案,支持 24 位/16bit 音频输入,输出功率最高可达 3.2W,为系统提供清晰响亮的语音输出。

MAX98357A 通过 I2S 总线与 ESP32S3 连接,接收从云端获取或本地生成的语音数据。ESP32S3 首先接收百度文心大模型的文本回复,然后通过云端 TTS 服务将文本转换为语音数据,再通过 I2S 接口传输给 MAX98357A 进行解码和放大,最终由喇叭播放出来。这一语音合成与播放流程实现了毫秒级的响应速度,使系统的语音交互体验更加自然流畅。音频解码放大器和喇叭模块实物如图 3.7 所示。

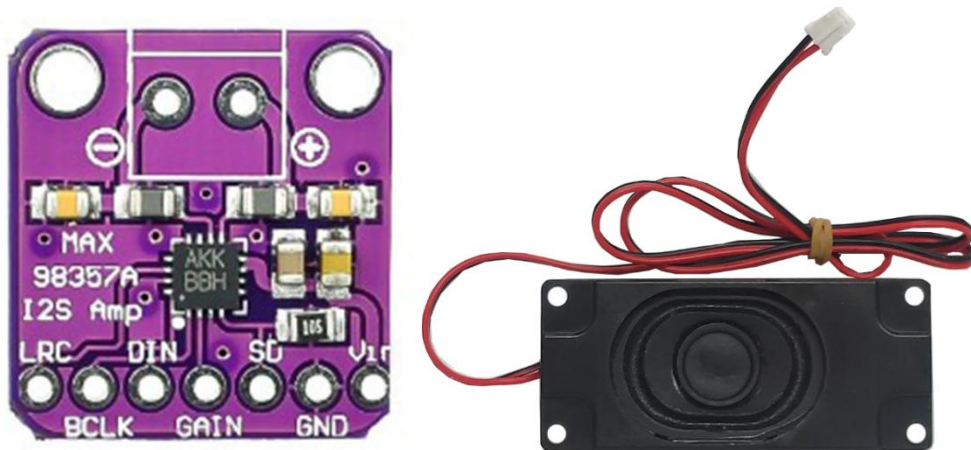


图 3.7 音频解码放大器和喇叭模块实物图

3.3.4 屏幕交互模块

为增强用户交互体验，ESP32S3 主板还集成了 1.8 寸 TFT 彩色显示屏，作为系统的视觉反馈界面。该显示屏基于 ST7735S 控制器，分辨率为 128×160 像素，支持 65K 色彩显示，通过 SPI 接口与 ESP32S3 连接，具有刷新速度快、功耗低的特点。

在交互过程中，显示屏主要用于展示与对话内容相匹配的表情符号，增强系统的情感表达能力。根据大模型分析的语义内容和情绪倾向，系统会选择合适的表情符号显示在屏幕上，与语音输出形成协同效应。此外，显示屏还能显示系统状态信息，如连接状态、录音状态等，提高系统使用的便利性和直观性。TFT 显示屏模块实物如图 3.8 所示。



图 3.8 TFT 显示屏模块实物图

3.4 系统通信与数据处理

为实现双主板之间的协同工作以及与云端服务的无缝连接，本系统设计了完整的通信架构和数据处理流程。系统通信主要包括板间通信、本地处理和云端交互三个层面。

在板间通信方面，ESP32 主控板与 ESP32CAM 通过串口通信实现图像数据传输和控制指令交换。ESP32 主控板与 ESP32S3 大模型交互板之间则采用 WiFi 网络进行数据共享，两块开发板连接到同一局域网中交换状态信息和控制指令。这种设计使得两个处理单元能够相对独立工作，同时又能在需要时共享关键信息。

在本地处理层面，ESP32CAM 摄像头采集帧流；而图像识别算法则由连接到同一网络的计算机 Python 程序执行，计算机通过 Python 程序接收 ESP32CAM 传来的图像，使用 OpenCV、Deepface 和 Matplotlib 库进行人脸检测和情绪识别分析，然后将结果通过 ESP32CAM 回传给 ESP32 主控板，以点亮灯带。此外，集成在 ESP32 主板的语音模块将识别到的语音文件以 WAV 的格式存放在本地计算机中，用于电脑端识别并进行语音情绪识别。

在云端交互层面，系统采用 Node.js 框架构建的服务作为数据接收，ESP32S3 通过 WebSocket 处理与百度文心大模型 API 的通信。ESP32S3 将语音结果发送至百度文心 API 获取回复内容，再将结果返回给 ESP32S3 进行语音合成和播放。整个通信架构设计如图 3.9 所示。

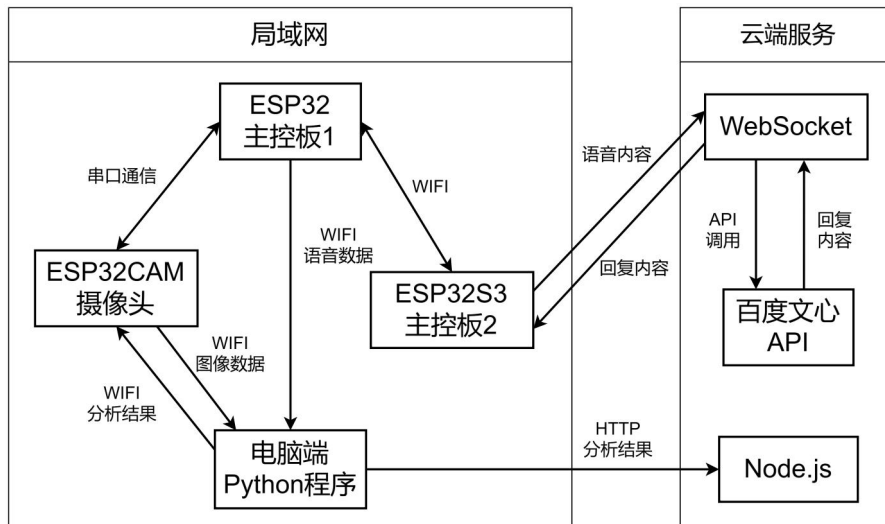


图 3.9 系统通信架构设计图

3.5 本章小结

本章详细阐述了情绪识别与交互系统的硬件设计方案，从整体架构到各功能模块的具体实现。系统采用双主板分布式设计，ESP32 主控板负责情绪识别和生理参数监测，ESP32S3 专注于大模型对话功能。各功能模块包括图像采集、情绪反馈、语音识别与播放、心率检测、语音采集与处理、音频播放以及视觉反馈等，通过合理的通信架构实现协同工作。这种模块化设计不仅提高了系统的可靠性和扩展性，还使系统能够同时处理多种复杂任务，为用户提供全面而自然的情感交互体验。

4 学生情感智能分析系统算法设计与实现

在完成系统硬件架构设计后，本章着重介绍学生情感智能分析系统的核心算法实现与传输机制。针对传统单模态情感识别的局限性，本章提出多模态情感指数融合算法，通过集成面部表情、语音语义及生理数据的三重感知，构建后期融合的多模态决策模型，以提升情绪判别的鲁棒性。同时，本章介绍了数据采集-计算-反馈的闭环机制，并结合自适应多模型预测框架与情绪阈值机制，为教育场景下的实时情感监测与干预提供了完整的算法支持。

4.1 多模态情感指数算法

在现代教育环境中，学生情感状态的准确识别对于提升教学质量和个性化教学具有重要意义。单一模态的情感识别方法往往存在局限性，难以全面捕捉复杂的情感表达。例如，面部表情可能因文化背景或个体差异而表现不一致，语音情感分析则可能受环境噪音干扰，而生理数据如心率虽然客观但难以直接对应具体情绪类别^[31]。因此，设计一种能够综合多种感知渠道、融合不同模态信息的情感指数算法，成为学生情感智能分析系统的核心需求。本节介绍的多模态情感指数算法通过整合面部表情、语音语义以及生理特征等多维度数据，能够更加全面、稳健地识别学生的情感状态，为后续的教学决策提供可靠依据。

4.1.1 基于 Deepface 的人脸情绪指数算法设计

随着人机交互和情感计算领域的快速发展，人脸情感识别技术已成为计算机视觉研究的重要分支。本节将详细介绍基于 OpenCV 和 DeepFace 框架的人脸情绪指数算法，该算法能够实现实时、高精度的面部情绪分析。

DeepFace 是一个轻量级开源框架，专注于人脸识别与属性分析，包括年龄、性别、情绪和种族等多维特征。该框架的核心优势在于整合了 VGG-Face、Facenet、ArcFace 等多种前沿深度学习模型，并简化了从人脸检测到特征分析的完整流程^[32]。在情绪识别任务中，用户只需调用 DeepFace.analyze 接口并指定 actions=['emotion'] 参数，即可通过预训练的卷积神经网络模型快速获取愤怒、厌恶、恐惧、快乐、悲伤、惊讶、中性七种基本情绪的预测结果。DeepFace 框架与 OpenCV 的结合提供了一个端到端的解决方案，用于实时视频流中的人脸情绪识别。OpenCV 主要负责视频捕获和初步的面部检测，而 DeepFace 则专注于高级情感分析任务。这种结合利用了 OpenCV 在图像处理方面的高效性和 DeepFace 在深度学习模型应用方面的优势。

1、Deepface 情绪识别算法实现原理

DeepFace 的情绪识别模块基于经典的 CNN 架构设计，输入图像经过人脸检测和对齐操作，确保面部区域标准化。对齐过程通过仿射变换调整面部关键点的位置，减少姿态变化对分类的影响。模型采用多层卷积-池化结构提取表情特征。模型以 FER2013 数据集训练模型为基准，其架构可分解为：

(1) 卷积层。在本层中，使用多个 3×3 卷积核堆叠，激活函数为 ReLU，公式为：

$$C_{out} = \text{ReLU}(W * X + b) \quad (4.1)$$

其中 W 为卷积核权重， X 为输入特征图， b 为偏置项。

- (2) 池化层。本层采用最大池化压缩特征维度，增强空间不变性。
 - (3) 全连接层。将展平后的特征映射到高维空间，最终通过 Softmax 输出七类情绪的概率分布。
- Deepface 的损失函数使用交叉熵损失函数来优化分类任务，公式为：

$$L = -\sum_{i=1}^7 y_i \log(p_i) \quad (4.2)$$

其中 y_i 为真实标签的 one-hot 编码， p_i 为模型预测的概率分布。优化器通常选择 Adam 或 SGD，结合学习率衰减策略提升收敛效率。

2、模型性能验证

DeepFace 的预训练模型在包含 28,709 张标注图像的 FER2013 数据集上进行了性能验证。该数据集涵盖了实验室控制环境与自然场景下的面部表情。根据文档记录，模型在 FER2013 数据集上的七分类准确率达到 57%，相较于该数据集上 34% 的基线模型准确率，表现相当优异^[32]。

为了直观展示多分类的性能表现，图 4.1 展示了 CNN 多分类模型的混淆矩阵。混淆矩阵是一种用于评估分类模型性能的重要工具，它通过将实际标签与模型预测标签进行对比，清晰地展示模型在各个类别上的分类效果。通过混淆矩阵，可以直观地观察到模型在不同类别上的准确率、召回率以及误分类情况。

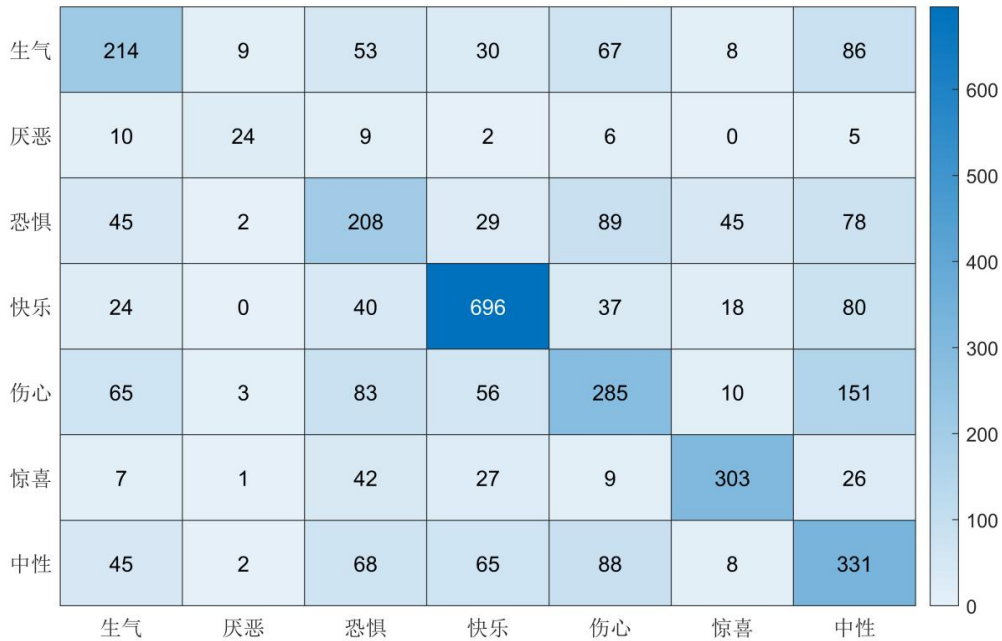


图 4.1 Deepface 在 FER2013 上的混淆矩阵热力图

最终，系统会接收七种情绪的概率分布，并组成学生的面部情绪指数作为单模态。

4.1.2 基于文心一言的语音情绪指数算法设计

本节提出了一种结合音频文件处理和大型语言模型的语音情绪识别算法，算法实现流程图如图 4.2 所示。

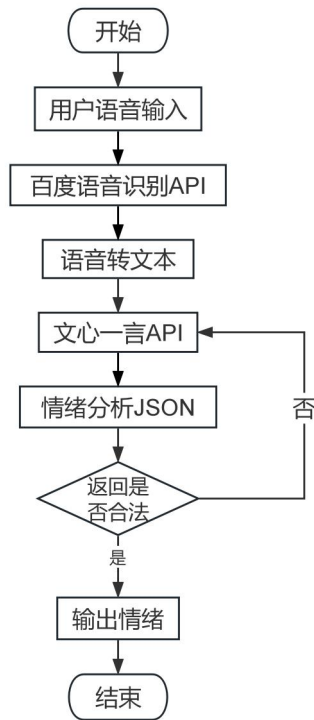


图 4.2 语音情绪指数算法流程图

该算法通过处理 WAV 音频文件将语音转换为文本，并利用百度文心一言模型进行情绪分析。本节详细阐述了音频文件处理、语音识别、文本情绪分析的实现原理及技术路线。算法能够有效识别七种基本情绪，并提供情绪强度的量化评估。综合评估认为，结合大模型的情绪分析准确率和运行效率综合优于传统方法。

1、音频文件处理

基于 WAV 文件的离线语音识别与实时语音采集系统存在显著差异，其处理流程主要包括以下几个关键步骤：系统首先读取并解析 WAV 文件头部信息，获取包括采样率、位深度和声道数等关键参数；随后根据识别 API 的技术要求进行必要的格式转换，统一采样率为 16kHz、转换为单声道等。针对较长的音频文件，系统会进行智能分段处理以避免超出 API 的处理容量限制，同时采用降噪算法提高音频信噪比，并通过音量归一化处理确保音频振幅处于最佳识别范围。由于 WAV 文件采用无损压缩格式完整保留原始音频信息，其处理过程相对简化，但仍需严格控制处理质量以满足 API 的输入要求，这是确保识别准确性的重要前提。

2、语音识别 API 选择

本系统评估了多个主流语音识别 API 的适用性，包括百度语音识别 API、科大讯飞语音听写 API、阿里云智能语音交互 API 和 Google Cloud Speech-to-Text API。经过综合对比分析，最终选择百度语音识别 API 作为核心识别引擎，主要基于其在中文语言环境下的出色表现。

百度语音识别 API 不仅全面支持 WAV、MP3 等常见音频格式输入，还内置了音频切分、静音检测等智能化预处理功能；其特色在于专门针对汉语语音特征进行了算法优化，支持批量文件处理，并在实际测试中展现出优异的中文识别准确率，完美契合本研究的核心需求。基于 WAV 的语音转文

本流程图如图 4.3 所示。

3、基于百度文心一言模型的情绪强度识别

本系统使用的百度文心一言 ERNIE-Bot-4 是基于 Transformer 架构的自回归生成式大型语言模型，通过海量数据训练实现高效的语言理解与生成。该模型采用知识增强预训练技术，通过融合结构化知识提升语义表征能力，并结合多阶段训练范式持续优化模型性能。为增强模型泛化能力，系统采用多任务联合优化策略，同步提升语义理解、逻辑推理等核心能力^[33]。

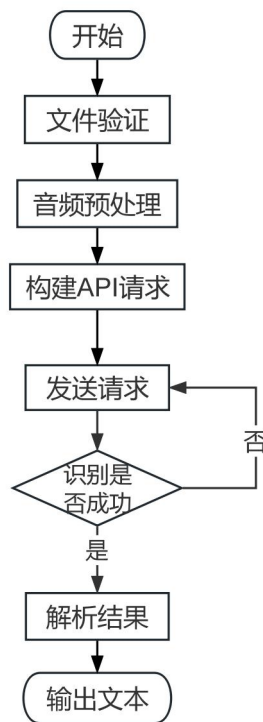


图 4.3 基于文件的语音识别流程图

模型具备长上下文建模特性，可有效处理超长文本的语义连贯性问题，并支持跨模态情感分析，能够准确识别文本中的显性情感表达和隐含情感倾向。在情绪分析任务中，模型通过综合分析词汇特征、句法结构和篇章级语境信息，实现高精度的情感倾向判别，尤其擅长处理反讽、隐喻等复杂情感表达。由于其技术文档较为成熟，且在中文任务中表现优异，因此选用为语音模态的情绪识别模型。

4、Prompt 工程

本研究采用 Prompt 工程技术构建了专门的情绪分析指令框架，旨在引导大型语言模型准确执行特定任务。该框架基于以下设计原则：首先明确规定情绪分析任务的具体要求，确保模型目标导向的一致性；其次限定输出格式为结构化 JSON 数据，包含预先定义的七种基本情绪类别及其对应概率分布，既降低输出不确定性又便于后续解析处理。

为优化模型表现，技术实现中特别设置了 0.1 的温度参数，通过抑制生成随机性来提升情绪判断的稳定性。此外，Prompt 中明确约束模型仅返回标准化 JSON 格式的结果，有效避免冗余解释造成的噪音干扰，从而保证输出数据的纯净度和可用性。具体提示词如表 4.1 所示。

表 4.1 本系统提示词内容

提示词内容
<p>请对下面这段文本进行情感分析，只需要分析出情感类别和强度。请严格按照以下七种情感类别分类："happy"(快乐), "sad"(悲伤), "angry"(愤怒), "disgusted"(厌恶), "fearful"(恐惧), "neutral"(平静), "surprised"(惊讶)。</p> <p>情感强度使用 0 到 1 之间的数值表示。</p> <p>只返回 JSON 格式的结果，不要有其他解释。格式如下：</p> <pre>{ "emotion": "情感类别(英文)", "confidence": 情感强度(0-1 之间的数值), "expressions": { "happy": 快乐的概率值, "sad": 悲伤的概率值, "angry": 愤怒的概率值, "disgusted": 厌恶的概率值, "fearful": 恐惧的概率值, "neutral": 平静的概率值, "surprised": 惊讶的概率值 } }</pre> <p>要分析的文本是: "\${text}"</p>

其中， $\text{\$}\{text\}$ 为上文语音识别出的文本。

5、模型性能验证

本系统选用大模型 API 接口实现语音情感分析主要基于以下考量：一是词向量匹配、词典规则等传统轻量化模型存在语义理解局限，难以处理反讽、隐喻等复杂情感表达，例如“你不会以为我真的开心吧”这类反语极易被误判为正面情绪；二是本地化部署的 BERT 等预训练模型虽具备较强的语义理解能力，但其推理速度较慢且计算资源消耗大，难以满足系统对实时性和轻量化的双重要求。

表 4.2 不同模型对于讽刺的性能对比

测试文本	词典法	词向量	BERT	大模型 API	人工标注
"系统崩溃 3 次，这技术真是太先进了"	快乐	快乐	愤怒	愤怒	愤怒
"您可真是个大聪明呢"	快乐	平静	厌恶	愤怒	愤怒
"这次项目太成功了，客户连夜删库跑路"	快乐	悲伤	愤怒	愤怒	愤怒
"嗯，你说的对，反正错的永远是别人"	平静	平静	厌恶	厌恶	厌恶
"这服务员态度真好（摔盘子）"	快乐	愤怒	愤怒	愤怒	愤怒
"真棒，又加班到凌晨"	快乐	悲伤	愤怒	愤怒	愤怒
"领导英明，改需求第 10 版了"	快乐	愤怒	愤怒	愤怒	愤怒
"我完全没生气（微笑）"	平静	平静	愤怒	愤怒	愤怒
"这方案满是漏洞，简直完美无瑕"	快乐	悲伤	愤怒	愤怒	愤怒

测试文本	词典法	词向量	BERT	大模型 API	人工标注
"太感谢了，帮我浪费了 3 小时"	快乐	愤怒	愤怒	愤怒	愤怒

通过表 4.2 的对比测试可以验证，本方案在保持准确率的同时将平均响应时间控制到最优，有效平衡了识别精度与系统性能的关系。

4.1.3 多模态融合决策算法

为了提升情感识别的准确性，本系统采用了多模态融合的情感决策机制。然而，在追求高精度的同时，我们也必须兼顾系统的实时性和轻量化需求。因此，系统选择了决策级融合策略。其中，面部情绪指数依据第 4.1.1 节所述方法计算得出，语音情绪指数则通过第 4.1.2 节所提方法计算完成。

此外，本系统还引入了第三个模态，即心率数据。然而，心率数据往往只能作为辅助决策的依据，因为某些情绪状态可能会导致心率的相似变化，从而难以仅凭心率进行准确的情绪判断。由于硬件的限制，面部表情和心率数据能够实现实时检测，而语音识别功能则需要用户主动激活语音模块后才会启动录音。这导致三种模态在时间维度上存在显著的异质性。

鉴于此，本节将详细介绍系统所采用的模态对齐方法以及动态权重决策算法的设计，以解决多模态数据在时间同步和决策权重分配上的挑战。

1、时间维度对齐

为了将不同模态计算出的指数进行时间对齐，本系统将连续的课程时间划分为若干等长的时间片段，并在每个片段内独立计算情绪指数。这种方法本质上通过“化整为零”的策略，将复杂的时间对齐问题分解为多个独立小时间段内的局部计算，从而简化了整体分析过程。

在具体实现中，基于学生的应用场景，总课程与课间时长为 50 分钟，将其分割为 25 个时间窗口，每个窗口长度为 120 秒。这一设计基于两个主要考虑：其一，2 分钟的时间跨度足以捕捉到显著的情绪变化，例如学生从专注状态转向分心状态，情绪会发生一定变化；其二，该时长能够确保每个窗口内有足够的数据点进行有效计算。窗口的数学表示为：

$$W = \{W_1, W_2, \dots, W_{25}\} \quad (4.3)$$

其中每个 W_k 表示第 k 个时间窗口，包含该 120 秒内的所有传感数据。

2、模态数据处理

在每个时间窗口内，系统会收集三种模态的数据，包括来自 ESP32CAM 的面部表情数据，每帧进行采样；来自大语言模型返回的文本情绪强度数据；以及来自心率传感器的心率生理数据。在预处理阶段，对于面部和心率这两种连续型数据，采用简单的算术平均方法来消除瞬时波动。

$$\bar{E}_{face}^{(k)} = \frac{1}{120} \sum_{t=(k-1) \times 120}^{k \times 120} E_{face}(t) \quad (4.4)$$

其中， $E_{face}(t)$ 表示时刻 t 的原始面部情绪值。这个公式表示把 120 秒内所有的面部情绪值相加后求平均。对于心率数据采用同样的处理方式。由于文本分析通常是间断进行的，我们采取“最近邻”原则，即使用该窗口内最后一次检测到的文本情绪值作为整个窗口的代表值。如果窗口内没有文本数据，则记为缺失状态。

3、心率数据辅助决策

心率作为重要的生理指标，能够反映学生的情绪唤醒水平。与面部表情和语音不同，心率数据本身并不能直接提供七种情绪的概率分布，但可以作为辅助决策的重要参考。由于心率传感器只能采集简单的心率时序数据，本系统采用一种简单有效的心率辅助决策方法，将心率时序数据转化为情绪判断的辅助依据。

基于心理生理学研究，不同情绪状态下心率变化具有一定规律：积极情绪通常伴随心率适度升高；消极情绪中，恐惧和愤怒往往导致心率显著升高，而悲伤可能导致心率降低；中性状态下心率保持在个体基线水平^[34]。

考虑到实现简单性和计算效率，本系统采用基于基线比较的心率情绪辅助方法。首先，系统取课堂开始前5分钟的平均心率建立学生的心率基线 HR_b ，对于窗口 k 内的心率数据，计算其平均值 $\overline{HR}^{(k)}$ ，计算方式如下：

$$\overline{HR}^{(k)} = \frac{1}{|W_k|} \sum_{t \in W_k} HR(t) \quad (4.5)$$

接下来，定义心率变化率计算方式如下：

$$\Delta HR^{(k)} = \frac{\overline{HR}^{(k)} - HR_b}{HR_b} \quad (4.6)$$

基于心率变化率，系统采用一个简单的三级调整机制：如果 $\Delta HR^{(k)} > 0.1$ ，即心率显著升高，则增加激活性情绪（愤怒、恐惧、快乐、惊讶）的权重，降低抑制性情绪（悲伤、中性、厌恶）的权重，如果 $\Delta HR^{(k)} < -0.05$ ，即心率降低，则增加悲伤情绪权重，降低快乐情绪权重。如果 $-0.05 \leq \Delta HR^{(k)} \leq 0.1$ ，即心率相对稳定增加中性情绪权重。在本系统中，这可以通过一个简单的调整向量实现：

$$C_{HR}^{(k)} = [c_a, c_d, c_{fe}, c_h, c_{sa}, c_{su}, c_n]^T \quad (4.7)$$

其中， $C_{HR}^{(k)}$ 为对应的调整向量，用于后续动态权重融合中的情绪辅助决策向量。下标(a)对应 anger，即愤怒，下标(d)对应 disgust，即厌恶，下标(fe)对应 fear，即恐惧，上标(h)对应 happiness，即快乐，下标(sa)对应 sadness，即悲伤，下标(su)对应 surprise，即惊讶，下标(n)对应 neutral，即中性/平静。

4、动态权重融合

在动态权重融合公式中，我们依据教育学领域的研究经验，预先设定了各模态的基准重要性。具体而言，将面部表情的权重设为 0.45，因为视觉信息最能反映即时的情绪状态；语音特征的权重为 0.45，因为语言内容直接反映认知状态；而生理指标的权重为 0.10，主要起到辅助决策的作用，即预设的权重向量为 $w_b = (0.45, 0.45, 0.10)$ 。

在实际情况中，常常会遇到模态数据缺失的问题，例如在课堂上学生可能不会发言，导致文本模态数据缺失。为了解决这一问题，本文采用了一种动态调整机制：当某个模态数据缺失时，系统会自动根据模态可用性指示函数重新分配权重。

$$\phi_m(W_k) = \begin{cases} 1, & \text{模态 } m \text{ 在窗口 } W_k \text{ 有效} \\ 0, & \text{无效} \end{cases} \quad (4.8)$$

实际计算时的权重调整为：

$$w_m^{(k)} = \frac{\phi_m(W_k) \cdot w_{m,b}}{\sum \phi_i(W_k) \cdot w_{i,b}} \quad (4.9)$$

其中， $w_m^{(k)}$ 表示窗口 k 中模态 m 的实际权重， $\phi_m(W_k)$ 表示模态 m 在窗口 k 的可用性， $w_{i,b}$ 表示模态 i 的基础权重。该公式确保在模态数据缺失的情况下，可用模态能够按比例“吸收”缺失模态的权重。例如，当文本模态数据缺失时，面部表情和心率模态的权重将自动调整为 0.82 和 0.18，同时保持原有的比例关系。

最终的学生多模态情绪指数计算公式如下：

$$E^{(k)} = w_f^{(k)} \cdot \bar{E}_f^{(k)} + w_t^{(k)} \cdot \bar{E}_t^{(k)} + w_{hr}^{(k)} \cdot \bar{E}_{hr}^{(k)} \quad (4.10)$$

若第 10 个时间窗口中，面部表情的均值为 0.72，表明情绪倾向积极；心率的均值为 0.55，表明处于中等兴奋水平；而文本数据缺失。依据动态权重调整机制，面部表情权重调整为 0.82，心率权重调整为 0.18。该窗口的情绪值计算为：

$$E^{(10)} = 0.82 \times 0.72 + 0.18 \times 0.55 \approx 0.68 \quad (4.11)$$

这一结果表明，情绪状态介于中性与积极之间，主要由面部表情的积极倾向所驱动，同时心率数据也反映了适度的兴奋程度。

上文计算出的仅为情绪指数，用于软件端可视化的一部分，为了精确分类情绪，而非简单地将情绪按指数划分为积极中性和消极，为更全面地捕捉情绪状态并支持多维度可视化，本系统引入七种基本情绪的概率分布融合算法。设面部和文本情绪模态的概率分布向量为：

$$P_f^{(k)} = [p_f^{(a)}, p_f^{(d)}, p_f^{(fe)}, p_f^{(h)}, p_f^{(sa)}, p_f^{(su)}, p_f^{(n)}]^T \quad (4.12)$$

$$P_t^{(k)} = [p_t^{(a)}, p_t^{(d)}, p_t^{(fe)}, p_t^{(h)}, p_t^{(sa)}, p_t^{(su)}, p_t^{(n)}]^T \quad (4.13)$$

其中， $P_f^{(k)}$ 为 Deepface 输出的面部情绪概率分布， $P_t^{(k)}$ 为文心大模型按照 Prompt 返回的情绪强度概率分布。上标(a)对应 anger，即愤怒，上标(d)对应 disgust，即厌恶，上标(fe)对应 fear，即恐惧，上标(h)对应 happiness，即快乐，上标(sa)对应 sadness，即悲伤，上标(su)对应 surprise，即惊讶，上标(n)对应 neutral，即中性。参考情绪指数，采用基于置信度的加权平均方法进行融合：

$$P_{fu}^{(k)} = \frac{w_f^{(k)} \cdot P_f^{(k)} + w_t^{(k)} \cdot I_t^{(k)} \cdot P_t^{(k)}}{w_f^{(k)} + w_t^{(k)} \cdot I_t^{(k)}} \quad (4.14)$$

融合后的概率分布向量中 $P_{fu}^{(k)}$ 的每个元素代表相应情绪类别的概率，总和为 1。最后，将心率调整向量添加到融合后的情绪分布向量，并重新归一化：

$$P_{imp}^{(k)} = P_{fu}^{(k)} + C_{HR}^{(k)} \quad (4.15)$$

$$P_{fin}^{(k)} = \frac{\max(P_{tmp}^{(k)}, 0)}{\sum_j \max(p_{tmp,j}^{(k)}, 0)} \quad (4.16)$$

其中 $\max(\cdot, 0)$ 操作确保所有概率值非负，归一化确保概率总和为 1。 $P_{fin}^{(k)}$ 表示最终的情绪概率分布向量。在代码中，接收到各指数后进行简单的设计加权逻辑和指示函数动态调整权重，这无需消耗大量计算资源，相较于特征级融合方法更加符合适合教育场景下的应用，满足实时性。整个多模态决策的程序设计流程图如图 4.4 所示。

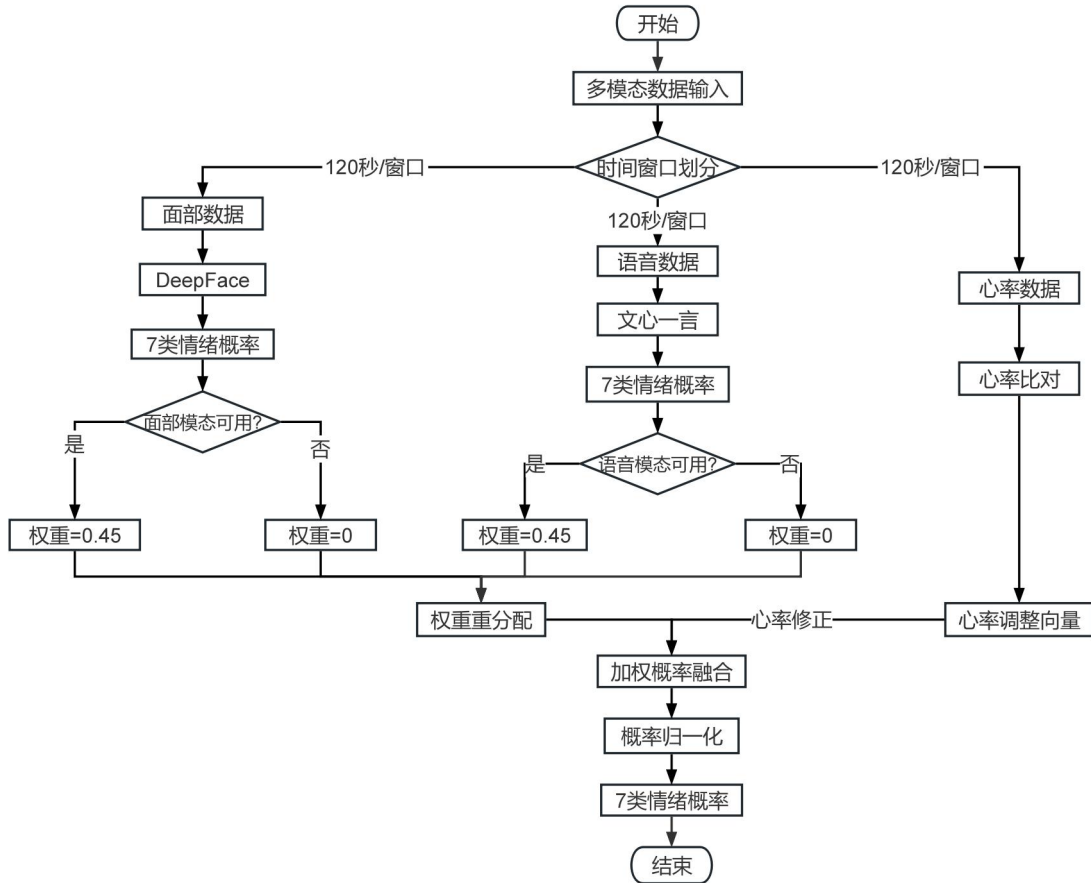


图 4.4 多模态决策程序设计流程图

4.2 多模型融合的情绪指数预测算法设计

学生情绪状态是影响学习效果和心理健康的重要因素。传统教育环境中，教师往往难以及时、准确地感知每位学生的情绪变化，尤其在班级规模较大的情况下。随着情感计算技术的发展，利用数据驱动方法预测学生情绪变化趋势成为可能，这有助于教师提前采取针对性干预措施，优化教学环境，提升学生学习体验。本研究提出了一种基于多模型融合的学生情绪预测方法，旨在为教育工作者提供更准确、可靠的情绪趋势预测工具。

在 4.1 节中，我们已经详细阐述了数据采集以及情绪指数计算的相关内容，因此此处不再重复介绍预处理等过程。在实际应用中，学生情绪指数的时序数据是动态更新的。鉴于此，本研究采用自适应多模型预测框架，能够根据数据特性动态选择最适合的预测模型。具体而言，模型选择框架

如图 4.5 所示。

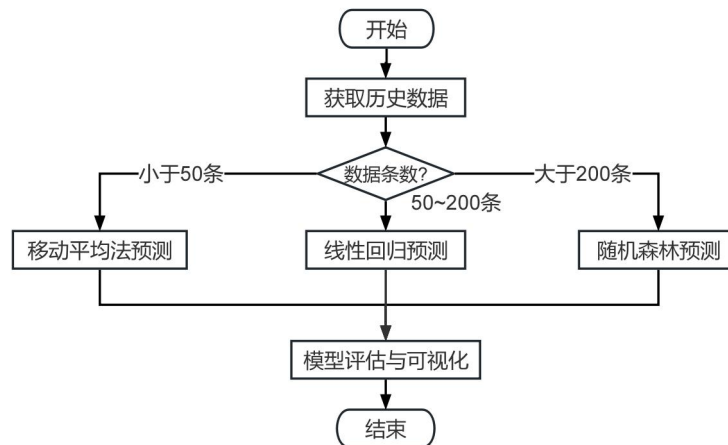


图 4.5 多模型情绪指数预测选择框架

4.2.1 移动平均法

当数据量太小，本系统定义为小于 50 条时，统计规律不明显，且线性回归、机器学习等复杂模型容易过拟合。移动平均计算简单，能平滑短期波动，提供稳定的趋势估计，适合数据稀疏的场景。移动平均法是一种简单而有效的时序预测方法，特别适合在数据量较少的情况下使用。

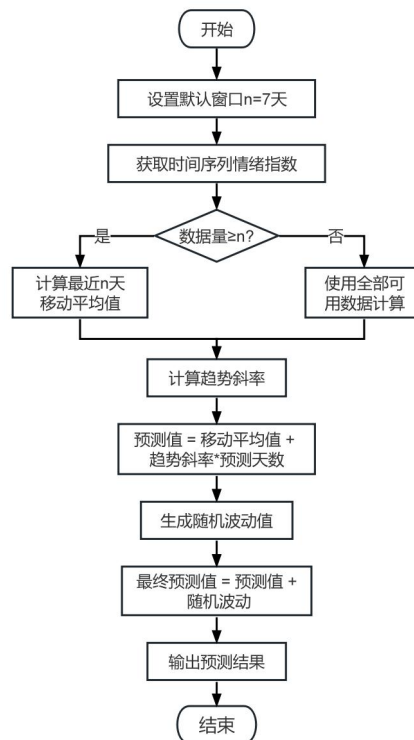


图 4.6 少量数据移动平均算法设计框架

其核心思想是通过计算最近 n 天的情绪指数的移动平均值来平滑数据波动，并结合趋势斜率来预测未来值^[35]。其实现步骤如下：

- (1) 使用最近 n 天的情绪指数计算移动平均值；
- (2) 计算最近数据的趋势斜率；

- (3) 基于移动平均值和趋势斜率预测未来值；
- (4) 添加随机波动以减少预测结果的平坦性。

在本系统设计的算法模块中，核心程序设计方法如图 4.6 所示。

4.2.2 线性回归模型

当数据记录数量在 50 到 200 条时，此时数据具有一定规模，可以拟合简单趋势。线性回归计算高效、解释性强，适合捕捉数据的整体变化方向，而不像移动平均仅能反映局部均值。然而，数据仍不够支撑更复杂的机器学习与深度学习模型，因此折中选择。线性回归是一种基础的统计模型，通过拟合特征与目标变量之间的线性关系进行预测^[36]。本系统的线性回归模型实现步骤如下：

- (1) 将时序作为特征 x ，情绪指数作为目标变量 y ；
- (2) 使用最小二乘法计算回归斜率和截距；
- (3) 添加周期性特征以捕捉周末和月度变化模式；
- (4) 添加随机噪声以模拟实际数据的波动性。

线性回归模型的公式表示为：

$$y = \beta_0 + \beta_1 x + \epsilon \quad (4.17)$$

其中， y 是预测的情绪指数， x 是时间戳， β_0 是截距， β_1 是斜率， ϵ 是误差项。在具体实现中，我们添加了周期性成分以捕捉时间特征：

$$y_{\text{pred}} = \beta_0 + \beta_1 x + \alpha \sin(2\pi x / 7) + \epsilon \quad (4.18)$$

其中 $\alpha \sin(2\pi x / 7)$ 代表每周的周期性变化。

4.2.3 随机森林模型

当数据量足够大时，随机森林能更好捕捉非线性关系、交互特征和复杂模式，且抗噪音能力优于线性模型。移动平均或线性回归可能丢失细节，而随机森林通过集成学习提升预测稳定性，适合大规模数据的精细化建模。随机森林模型作为各大时序预测任务的基线模型，在保证高准确率的同时，处理时间效率远远优于基于 CNN、LSTM 等深度学习网络模型。因此，在本应用场景时，选用随机森林来处理大规模数据集。

随机森林回归是一种集成学习方法，通过构建多个决策树并平均其预测结果来提高预测准确性和稳定性。当数据量足够大时，随机森林能捕捉到数据中的复杂非线性关系^[37]。本系统的随机森林模型实现步骤如下：

- (1) 构造训练特征集，包括天数，即线性特征、周周期性特征，定义为 $\sin(2\pi x/7)$ 和 $\cos(2\pi x/7)$ 、月周期性特征，定义为 $\sin(2\pi x/30)$ 和 $\cos(2\pi x/30)$ ；
- (2) 使用情绪指数作为训练标签；
- (3) 训练随机森林模型，在本系统中使用 25 棵决策树；
- (4) 使用训练好的模型预测未来情绪指数。

随机森林相比单一模型具有诸多优势。它能够处理非线性关系和特征交互，这使得它在复杂数据环境中表现出色；同时，随机森林不易过拟合，对异常值也不敏感，进一步增强了模型的稳定性

和鲁棒性；此外，它还可以自动评估特征重要性，为特征选择和模型解释提供了便利。随机森林模型的训练和预测过程可形式化为：

$$\hat{f}(x) = \frac{1}{25} \sum_{b=1}^{25} T_b(x) \quad (4.19)$$

其中， $\hat{f}(x)$ 是随机森林预测的结果，25 是决策树数量， $T_b(x)$ 是第 b 棵决策树对样本 x 的预测。

在前端展示部分，系统的核心功能之一是呈现历史情绪指数数据以及预测数据的折线图，具体展示页面将在第 5 章详细阐述。目前，由于本系统尚未正式落地应用，尚未积累足够规模的时序数据集，因此未能充分评估模型的性能，选择该算法体系是基于已有研究。接下来，为了验证模型在时序预测的有效性，本文基于 Kaggle 开源的时序数据集，进行模型的评估。

4.2.4 模型的评估

时序预测是机器学习和数据挖掘领域的重要研究方向，其方法论涵盖传统统计学模型、机器学习算法及深度学习架构^[35]。经典方法如线性回归（Linear Regression）和自回归积分滑动平均模型（ARIMA）因其参数可解释性和计算效率，至今仍在简单时序模式中表现良好；以决策树（Decision Tree）、随机森林（Random Forest）和梯度提升树（Gradient Boosting Trees）为代表的集成学习方法通过组合弱学习器提升预测精度，尤其擅长处理非线性关系；而长短期记忆网络（LSTM）和卷积神经网络（CNN）等深度学习模型则通过隐状态机制或局部感知野捕获时序依赖关系与空间特征。近年来，Transformer 凭借自注意力机制（Self-Attention）在序列建模中崭露头角，但其在时序预测中的有效性仍需验证。

为全面评估模型性能，本文采用均方根误差（RMSE）、平均绝对误差（MAE）、决定系数（R²）、平均绝对百分比误差（MAPE）及运行时间作为评价标准。RMSE 通过平方项强化大误差的惩罚力度，MAE 则提供误差绝对值的无偏估计；R² 衡量模型对目标方差的解释程度，MAPE 将误差归一化为百分比以提升可比性，具体结果如表 4.3 所示。

表 4.3 不同模型的预测性能对比

模型	RMSE	MAE	R ²	MAPE	运行时间
LSTM	9.02	5.76	0.95	16.17	33.93
CNN	9.43	5.86	0.95	15.14	35.91
CNN+LSTM	10.14	6.74	0.94	19.65	43.18
线性回归	8.84	5.16	0.96	12.84	0.01
决策树	9.5	5.36	0.95	12.61	0.07
随机森林	8.72	4.99	0.96	12.25	0.10
XGBoost	8.71	4.99	0.96	12.22	0.23
SVM	21.34	18.03	0.74	67.97	0.51
梯度提升树	8.72	5.18	0.96	13.03	1.98
LightGBM	8.81	5.04	0.96	12.4	0.2
CatBoost	8.97	5.32	0.95	13.44	0.89
Transformer	43.16	34.16	-0.06	105.41	130.54

XGboost 虽是多数表格时序数据的 benchmark，但因本系统情绪指数为日频数据，样本量有限易导致过拟合，故未采用。未来随着真实环境下数据集的逐步采集，我们将对预测模块进行重构，以确保其在实际应用中的准确性和可靠性。

4.3 硬件与服务端交互子系统设计

4.3.1 ESP32 主控制板与服务端通信子程序

实现 ESP32 主控制板与 Node.js 服务端的通信，需进行系统性的开发环境配置与网络连接建立。首先需在 Arduino IDE 中配置 ESP32 开发环境，安装 ESP32 开发板支持包、WiFiClient 和 HTTPClient 等相关库文件；其次配置 WiFi 连接参数，包括接入点 SSID 和密码，确保 ESP32 能够稳定接入局域网；然后配置 Node.js 服务器的 IP 地址和端口信息；最后通过 HTTP 协议建立与 Node.js 服务端的连接，确保数据传输的可靠性。

通信子程序的核心代码模块主要包含以下功能：WiFi 连接管理模块、HTTP 请求构造模块、JSON 数据解析模块以及错误处理与重试机制。WiFi 连接管理模块负责监测网络状态并在断连时自动重连；HTTP 请求构造模块负责组装请求头与请求体；JSON 数据解析模块利用 ArduinoJson 库处理服务器返回的结构化数据；错误处理与重试机制则确保在网络不稳定情况下系统依然能够可靠运行。

4.3.2 系统采集数据上传服务端

系统与 Node.js 服务端建立 HTTP 连接后，多模态数据的上传过程需遵循协议规范。首先，ESP32 主控制板从 ESP32CAM 接收图像情绪识别结果，同时采集 INMP441 传感器的语音数据和心率传感器数据，系统将处理后的数据按照 JSON 格式进行封装。

数据上传前，系统会执行网络状态检测，确认 WiFi 连接稳定且信号强度满足要求。当网络条件满足时，系统通过 HTTPClient 库发起 POST 请求，将 JSON 格式的多模态数据传输至 Node.js 服务端。

采集上传的多模态数据主要包括三部分：一是 ESP32CAM 采集并通过 OpenCV 和 Deepface 预处理后的面部情绪概率分布数据，包含七种基本情绪的概率值；二是通过 INMP441 麦克风采集并经文心大模型 API 分析后的语音情绪数据；三是通过心率传感器采集的生理数据。这些数据以统一的 JSON 格式封装，确保服务端能够正确解析。

4.4 服务端数据处理流程设计

ESP32 主控制板与 Node.js 服务端之间的通信采用 HTTP 协议进行数据传输，同时通过 WebSocket 实现控制指令的实时下发。当 ESP32 采集到人脸图像或语音数据后，通过 HTTP POST 请求将数据发送至服务端。服务端接收数据后，执行以下处理流程：首先对数据进行格式验证与内容规范化；其次将面部图像数据传递给本地 Python 脚本，调用 Deepface 进行情绪分析；然后将语音数据通过文心大模型 API 进行情绪识别；最后在 server.js 中执行第 4.1.3 节设计的多模态融合决策算法，计算出最终的情绪指数和情绪分布。

服务端处理流程中的关键技术包括：使用 child_process 模块调用 Python 脚本进行面部情绪分析，利用 axios 库与文心大模型 API 进行通信，以及采用 WebSocket 技术实现与 ESP32 的双向实时通信。当计算得到情绪指数后，服务端通过 WebSocket 将结果实时传回 ESP32，用于控制 WS2812 灯带的

颜色显示。

硬件与服务端的完整工作时序图如图 4.7 所示。

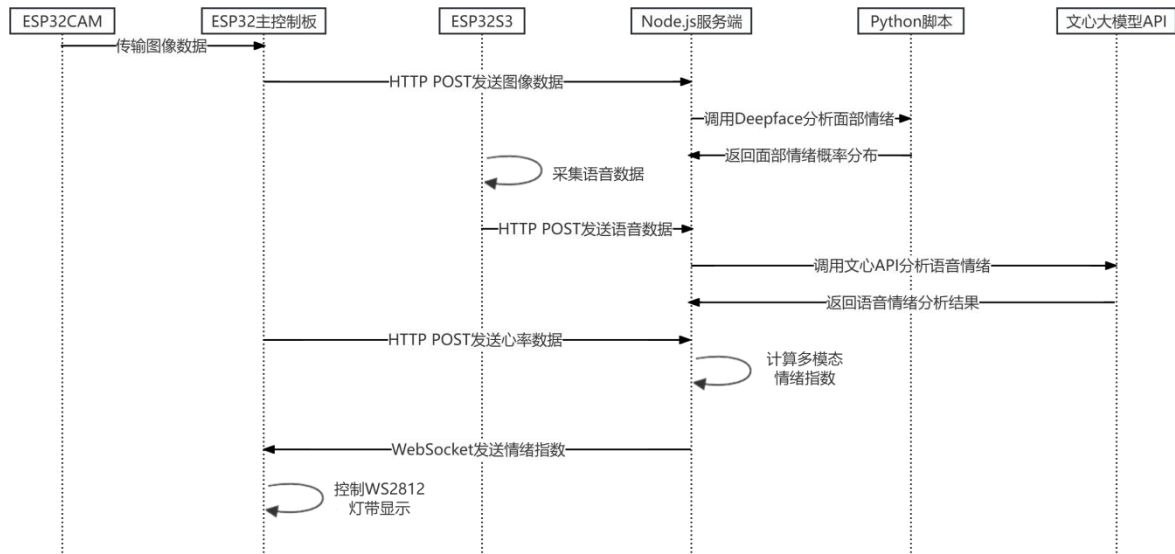


图 4.7 硬件与服务端通信与工作时序图

4.5 情绪阈值控制机制设计

基于 Node.js 服务端计算的多模态情绪指数，本系统设计了情绪阈值控制机制。该机制通过对来自不同传感器的数据进行综合分析，实现对使用者情绪状态的准确判断，并根据判断结果触发相应的反馈行为。

情绪阈值控制机制的核心是基于第 4.1.3 节描述的多模态融合算法计算得出的情绪指数。系统采用三级划分法，将情绪状态分为积极、中性和消极三个主要等级。具体划分标准为：情绪指数大于 0.6 时判定为积极情绪，ESP32 主控制板控制 WS2812 灯带显示绿色；情绪指数介于 0.4 到 0.6 之间时判定为中性情绪，灯带显示黄色；情绪指数低于 0.4 时判定为消极情绪，灯带显示红色。此外，系统还设置了极度消极的情绪阈值，当情绪指数低于 0.2 时，除显示红色灯带外，系统还会激活震动马达并通过 ESP32S3 的 ASRPRO 模块播放预先设定的安抚语音内容。

阈值控制机制的参数设计如表 4.4 所示，这些参数可根据实际应用场景进行调整，以满足不同用户群体的需求。

表 4.4 情绪阈值参数及触发条件列表

触发条件	参数说明	灯带颜色	附加动作
情绪指数 > 0.6	积极情绪判定阈值	绿色	无
$0.4 \leq \text{情绪指数} \leq 0.6$	中性情绪判定阈值	黄色	无
$0.2 < \text{情绪指数} < 0.4$	消极情绪判定阈值	红色	无
情绪指数 ≤ 0.2	极度消极阈值	红色	震动马达激活和安抚语音播放

情绪阈值控制逻辑在 Node.js 服务端和 ESP32 端均有实现。服务端计算出情绪指数后，将对应的控制指令通过 WebSocket 发送给 ESP32，指令中包含灯带颜色和是否触发附加动作的信息。ESP32 接收到指令后，根据指令内容控制 WS2812 灯带显示相应颜色，并在必要时激活震动马达和播放安抚语音。

4.6 本章小结

本章详细介绍了学生情感智能分析系统的算法架构及其实现方案。首先提出了融合面部、语音和生理数据的多模态情感指数算法，通过 DeepFace 框架实现实时面部情绪分析，结合文心一言大模型构建语音情绪识别模型，并设计动态权重融合策略解决多模态数据异质性问题。随后介绍了基于数据规模自适应的多模型情绪预测框架，涵盖移动平均、线性回归和随机森林三种预测模型。最后介绍了硬件端与服务端的交互系统设计，包括数据采集传输、服务端计算以及情绪阈值反馈机制。该体系在保证实时性的同时，能有效提升情感识别的全面性与准确性，为教育场景下的情感智能监测提供了可靠的技术支持。

5 学生情感智能分析软件端设计

本章主要介绍了学生情感智能分析软件端设计，包括开发环境，功能设计，数据库设计等，选择 5.1 节所列开发环境的主要理由在于其轻量高效的特性。相比于复杂的前端框架如 React 或 Vue，直接使用原生 Web 技术降低了系统复杂度；而 Node.js 生态系统则为 AI 服务集成提供了便利条件，尤其适合情感分析这类需要频繁与外部 AI 接口交互的应用场景^[38]。同时，基于文件的 SQLite 数据库避免了额外数据库服务的依赖，使系统能够以极低的资源占用实现部署，这对于教育场景下通常资源有限的服务器环境尤为适合^[39]。

5.1 开发环境

本情感智能分析系统采用前后端分离的轻量级 Web 应用架构，服务端基于 Node.js 环境构建，前端采用现代 Web 技术栈实现。

在服务端方面，系统核心采用 Express 框架搭建 Web 服务，通过其简洁的 API 结构和中间件系统实现了路由管理、请求处理等功能。数据存储层使用 SQLite 轻量级关系型数据库，其文件化存储特性使系统部署维护更为便捷。系统还集成了 axios、cors、multer 等模块分别用于 API 通信、跨域请求处理和文件上传管理，构成了完整的后端服务体系。

前端开发环境则采用原生 Web 技术栈，包括 HTML5、CSS3 和 JavaScript，并辅以现代前端工具增强开发效率。页面样式通过 TailwindCSS 实现响应式设计，图表展示功能依托 Chart.js 实现数据可视化，界面图标则由 FontAwesome 提供支持。前后端数据交互通过原生 Fetch API 实现，确保了系统的轻量化特性。

在开发工具链方面，主要依托 VSCode 作为集成开发环境，配合 Nodemon 实现服务端代码热重载，大幅提升了开发效率。数据库设计与调试借助 SQLite Browser 完成，前端调试与性能优化则通过 Chrome DevTools 实现。

5.2 功能模块设计

为实现对学生情感的智能分析与监测，本文设计了网页平台端的情感智能分析系统。

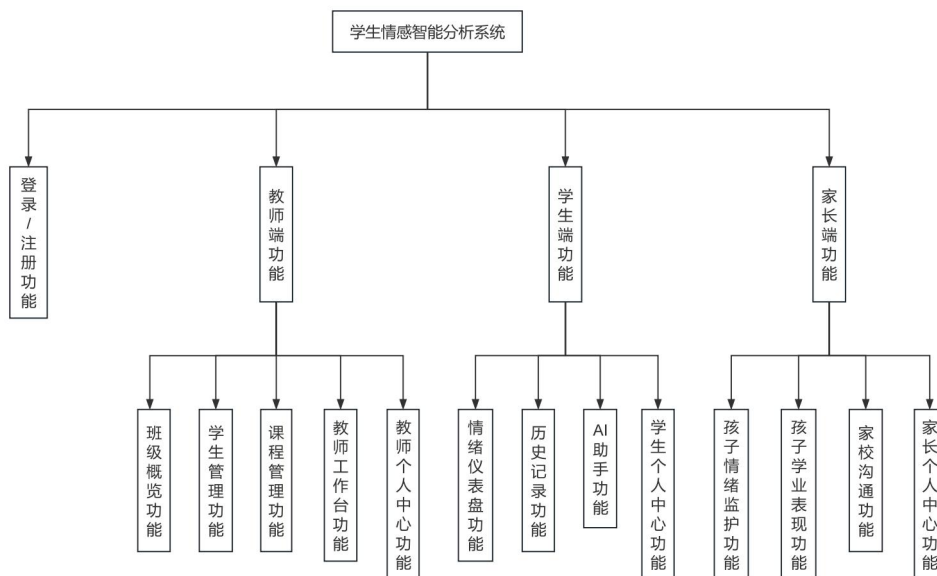


图 5.1 功能结构图

该系统的主要功能涵盖登录/注册、教师端功能、学生端功能以及家长端功能，其功能结构如图 5.1 所示。

5.2.1 登录功能设计

情感智能分析系统的登录功能基于多角色用户管理理念设计，支持学生、教师和家长三种不同角色的用户访问，各角色登录后进入对应的页面，获取与其权限相符的情感数据和操作界面。登录功能界面如图 5.2 所示。

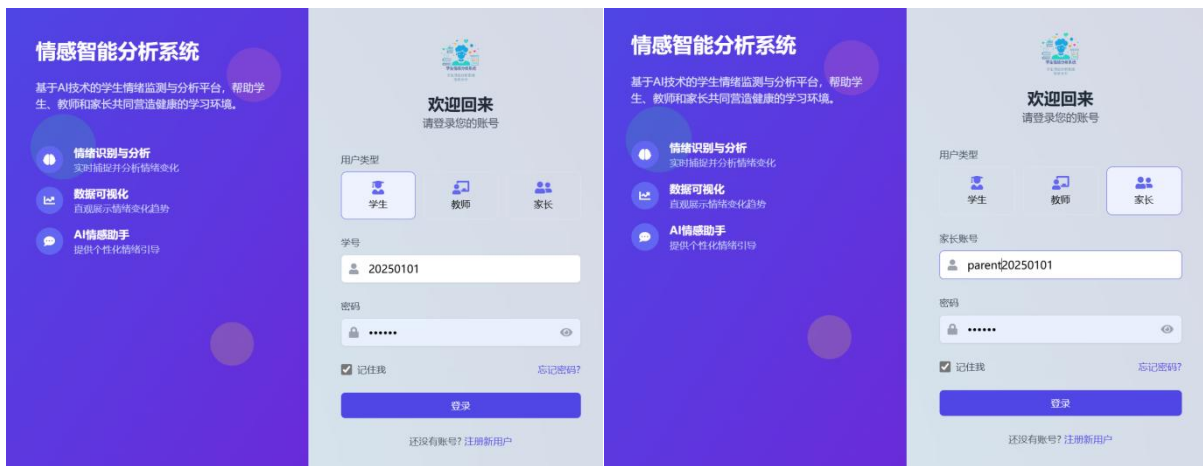


图 5.2 登录功能界面

登录功能与后端服务的集成主要通过 REST API 实现。客户端将用户输入的登录信息打包后发送 POST 请求至服务端验证接口，服务端返回验证结果和用户信息。根据返回的用户角色信息，系统自动导航至对应的功能页面，完成整个登录流程。

5.2.2 注册功能设计

情感智能分析系统的注册功能设计同样以角色分流为核心理念，为不同类型的用户提供定制化的注册流程，确保信息采集的精准性和系统数据的完整性。注册功能界面如图 5.3、图 5.4 所示，而家长端与学生端的页面一致。

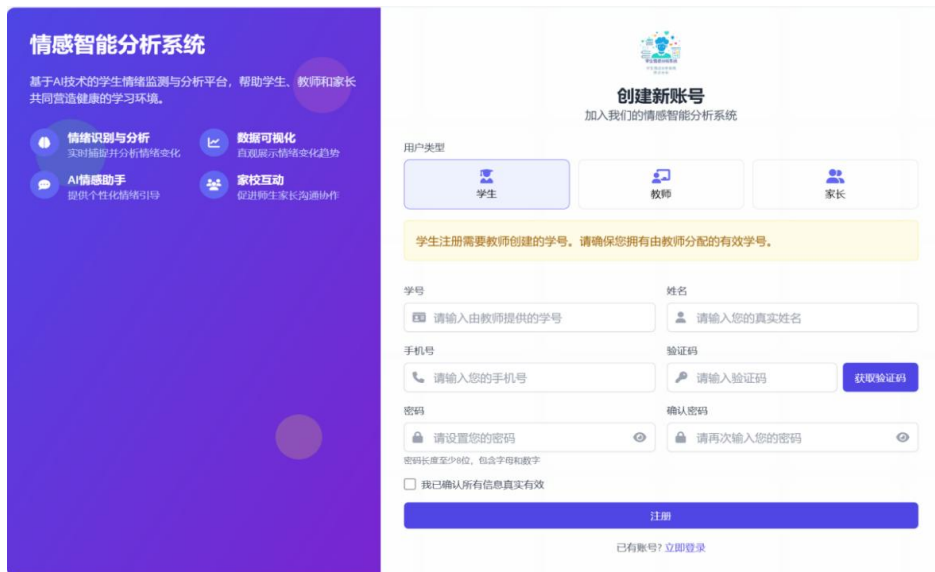


图 5.3 学生端注册功能界面

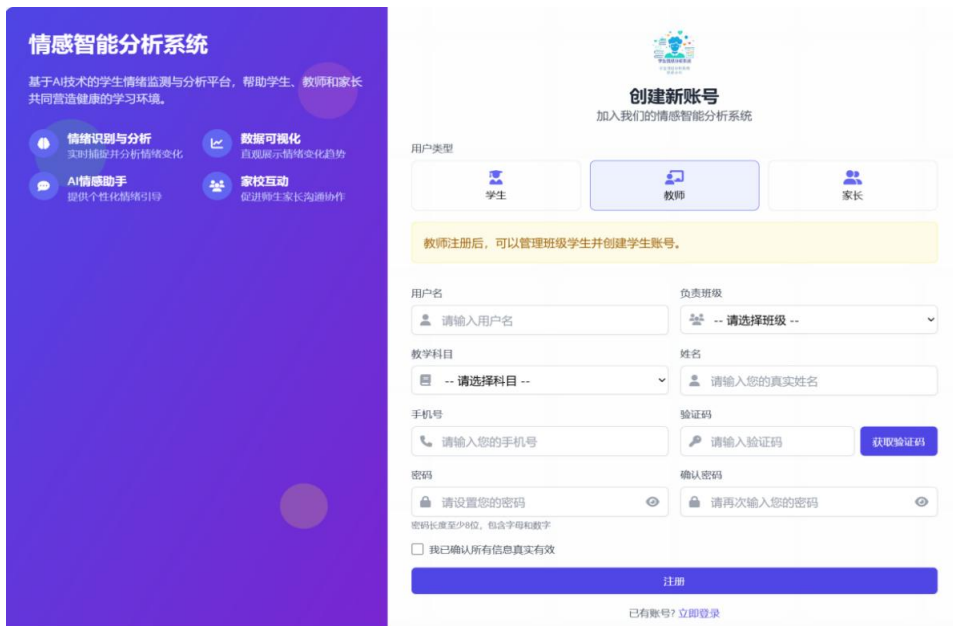


图 5.4 教师端注册功能界面

表单验证是注册功能的关键环节。系统实现了多层次的验证机制，包括学号存在性检查、手机号格式验证和密码强度评估。安全性设计贯穿注册功能的全过程。第一，通过手机验证码实现用户真实身份的确认，防止批量注册和恶意注册；第二，密码设置要求包含字母和数字且长度不少于 8 位，同时提供密码强度提示和双重输入确认机制；此外，系统还通过角色权限分离，确保只有教师可以创建学生账号，而学生和家長只能使用已存在的学号进行注册，这是基于数据库而设计的。

注册功能与后端服务的集成体现在学号验证通过 API 实时检查学号的有效性；注册信息提交后经服务端处理并存入数据库。

5.2.3 教师端功能设计

情感智能分析系统的教师模块设计以教学管理与情感分析协同为核心理念，为教师用户提供了全面的班级监控、学生情感分析和教学辅助功能，通过数据可视化与智能分析，助力教师实现情感导向的个性化教学。

1、班级概览功能设计

班级概览页面为教师登录后的首页，为教师提供班级情感动态的宏观视图，如图 5.5 所示。页面顶部展示班级整体情感雷达图，直观呈现当前班级的情感分布状况；中部区域则通过情感趋势图展示过去一段时间内班级情感变化曲线和课堂参与度排名，帮助教师识别可能的情感波动节点。班级情绪健康指数是由全班的學生由 4.1 节介绍的算法计算出的指数求平均值得出，情感指数预测是基于 4.2 节介绍的多模型算法，课程参与度是由教师工作台由教师手动输入。



图 5.5 教师端班级概览上半部分界面

下半部分的页面左侧显示“需要关注”的学生列表，如图 5.6 所示。点“查看详情”能够进入单个学生的情绪主页，出于隐私考虑，只能够查看学生较为基本的数据。点击“干预”能够进入到家校沟通页面，这个页面在下文介绍。基于情感异常检测算法自动筛选出情感状态发生显著变化或持续处于负面情绪的学生，提醒教师及时关注并干预。右侧支持导出 Excel 功能，点击后将可视化图表所用数据导出成表格。

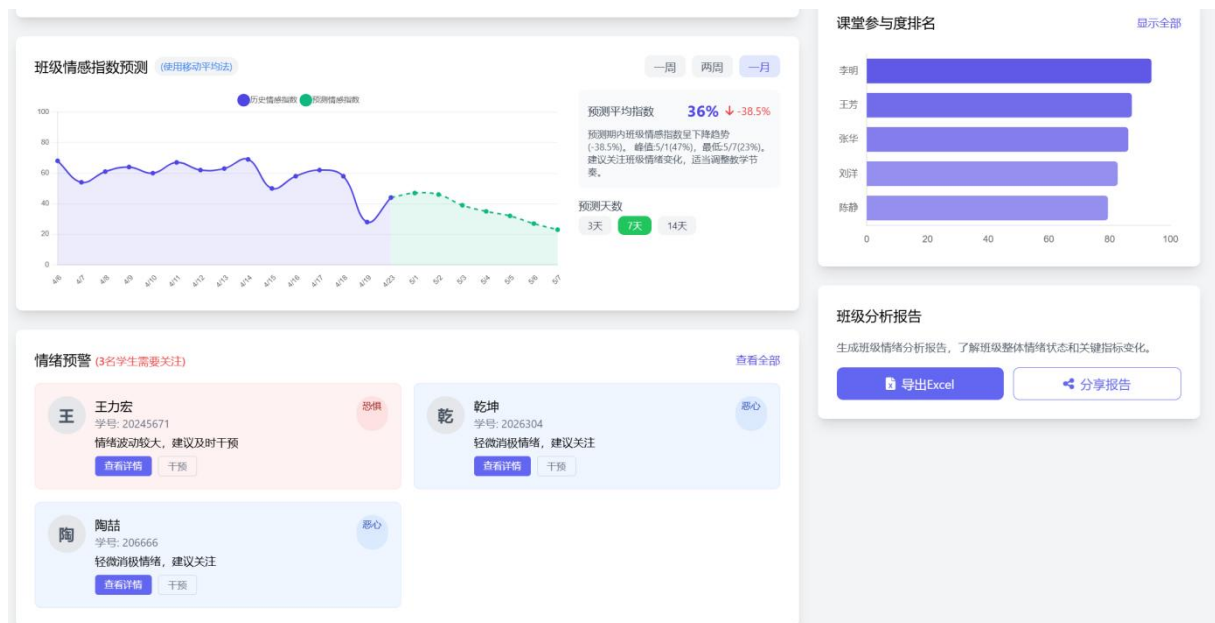


图 5.6 教师端班级概览下半部分界面

点击查看学生详情后，系统跳转到单个学生的情绪分析界面，上半部分为学生的个人基本信息、情绪指数、情绪曲线、情绪雷达图，如图 5.7 所示。

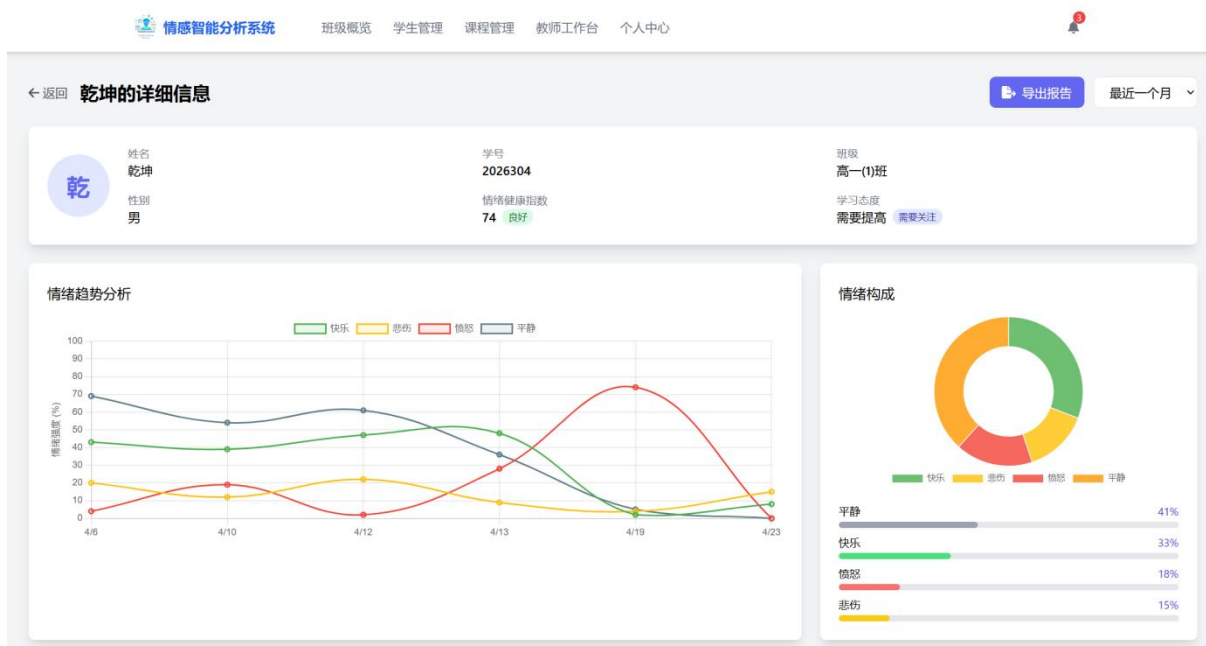


图 5.7 教师端查看学生详情上半界面

学生详情的下半部分为学业分析、课堂专注度、参与度、作业完成度的可视化分析，不同于来源于硬件的情绪数据，这些学业数据由教师工作台手工输入，目的是为了更好察觉情绪对学业的影响，如图 5.8 所示。

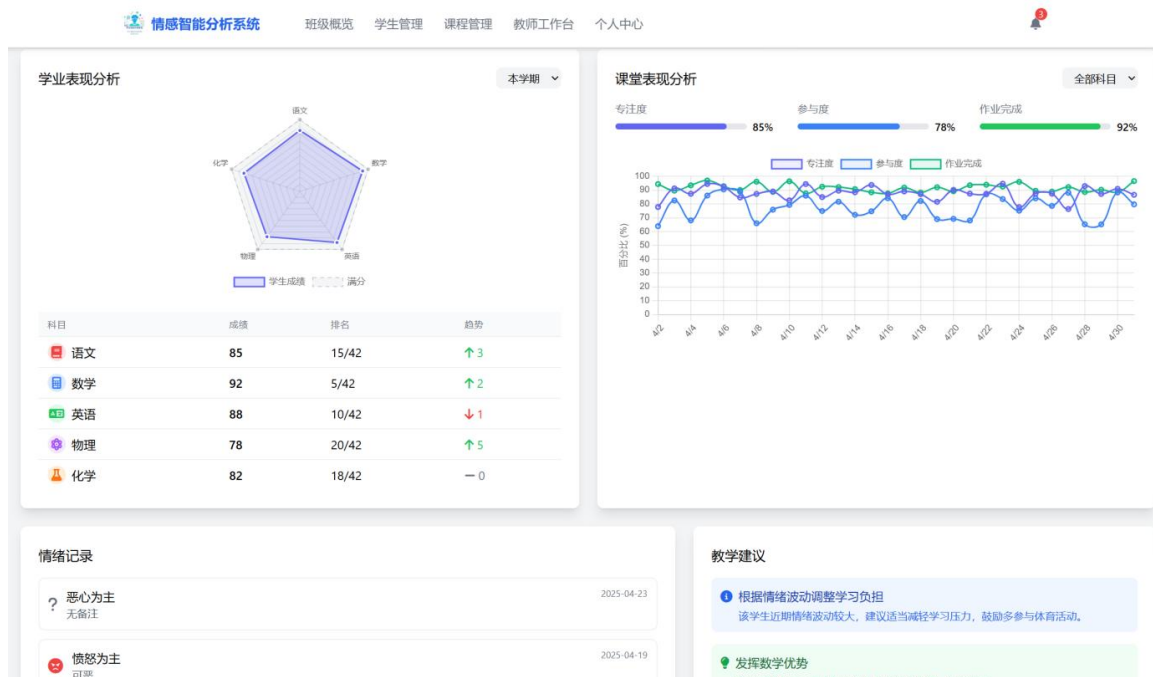


图 5.8 教师端查看学生详情下半界面

在班级概览页面点击“干预”按钮时，会弹出一个用于与家长沟通的界面。在此界面中，教师可以通过文字反馈的形式，直接向家长传达孩子在校的相关情况。为了更好地支持这一功能，我们专门设计了一个会话数据库，用于存储家校沟通的完整记录，方便教师和家长随时查阅过往的沟通内容，如图 5.9 所示。

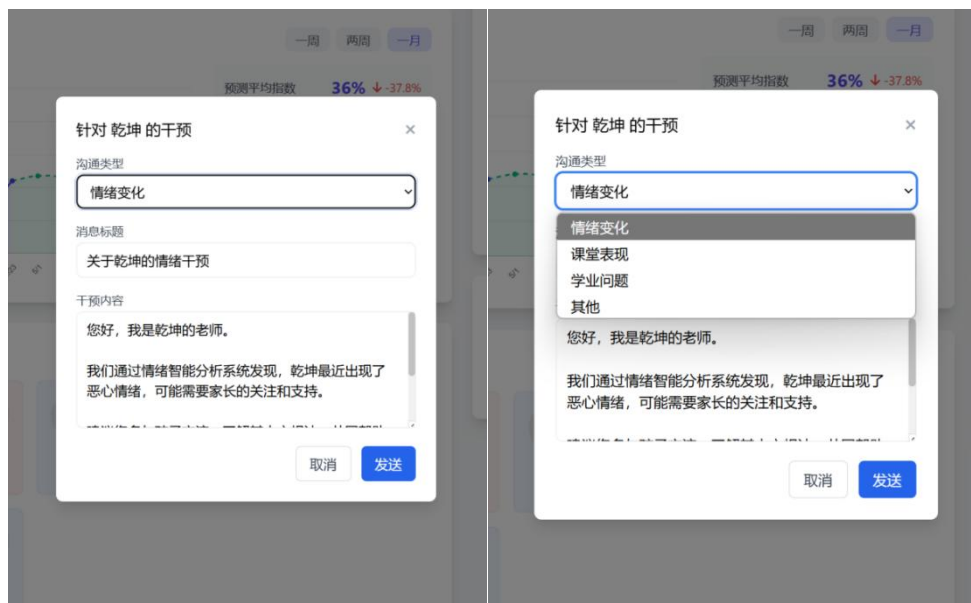


图 5.9 教师端班级概览“干预”界面

2、学生管理功能设计

学生管理功能为教师提供学生全面管理和个体情感分析能力。页面主体采用表格式布局，展示学生基本信息、最新情感状态、参与度和学习表现等指标，教师可通过表格顶部的搜索框和筛选器快速定位特定学生或具有相似特征的群体，如图 5.10 所示。

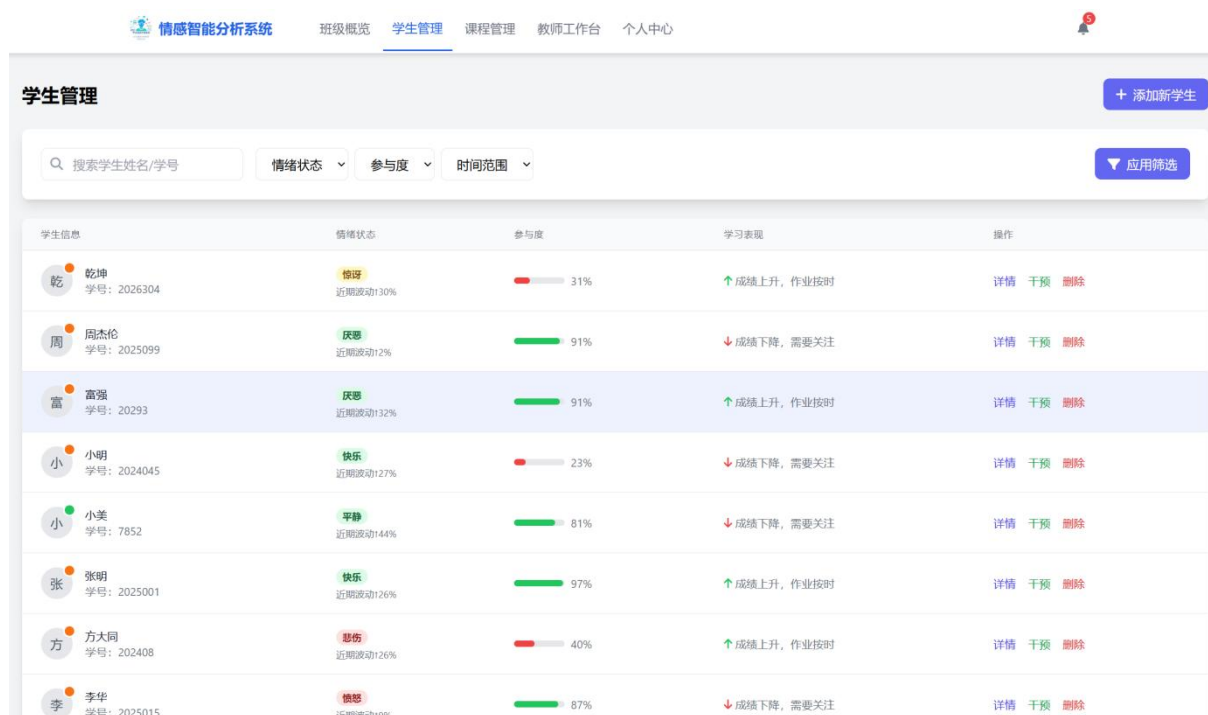


图 5.10 教师端学生管理界面

在这个页面中，详情和干预功能与班级概览页面保持一致。不过，该页面还具备一些独特的功能。教师可以在此删除学生的信息记录，也可以增加新的学生账号。关于学生信息记录和学生管理的功能介绍，将在教师工作台的功能设计小节详细介绍。

3、课程管理功能设计

课程管理为教师提供了课程规划、情感反馈采集等功能。页面主体分为"课程列表"和"课堂情绪分析"两大板块。由于课程表体系的数据库和后端数据设计较为复杂，因此针对此功能模块只设计了UI界面。

如图 5.11 所示，该功能的愿景是当某一节课上学生的情绪异常时，教师能够基于情绪分析调整教学计划，提高教学效率。



图 5.11 教师端课程管理界面

4、教师工作台功能设计

在教师工作台中，教师可以对学生信息进行增加、删除、查询和修改操作。只有当教师添加新学生并为其分配唯一的学号后，学生及其家长才能使用该学号注册并登录系统。这种设计不仅保证了硬件与软件之间的协同配合，避免数据异常的发生，还简化了数据库设计，提升了管理的便捷性和系统的稳定性。教师选择学生并填入信息后点击“保存数据”，数据将实时显示到学生历史数据的表格视图中，如图 5.12 所示。

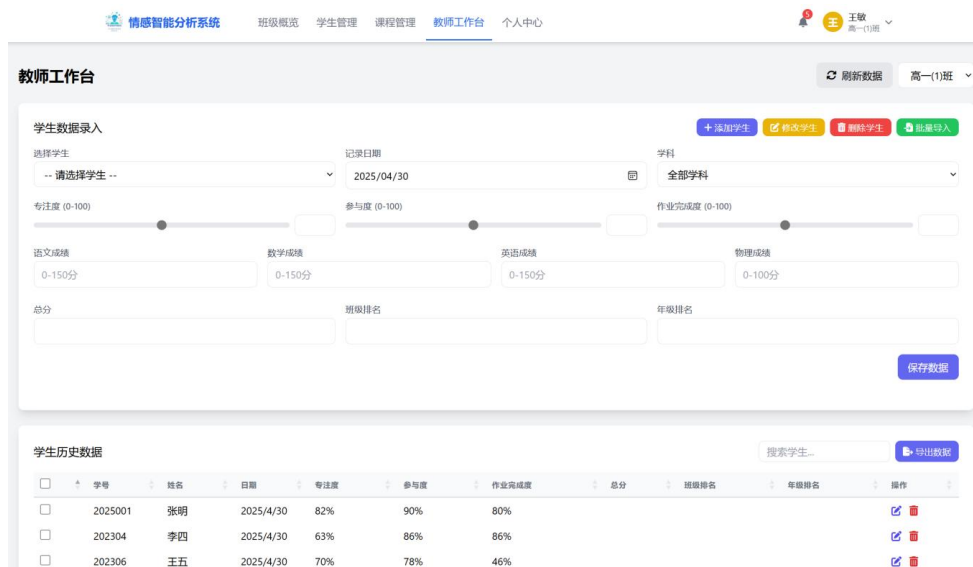


图 5.12 教师端工作台上部分界面

在选择了学生以后，工作台中间部分是对该学生的学业的可视化，包括课堂表现变化，成绩分析对比，这与学生详情页类似，目的是为了便于教师在填写时查看和检验，如图 5.13 所示。

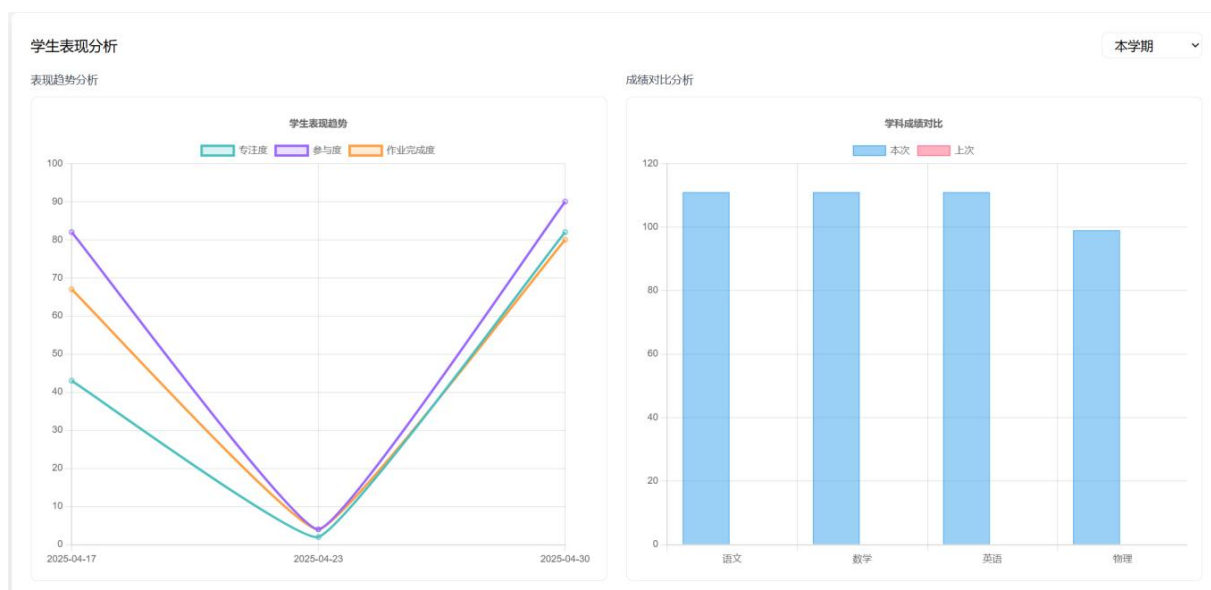


图 5.13 教师端工作台中间部分界面

在工作台的下半部分，是家校沟通的页面，在这里可以查看与家长和学生沟通的历史记录，并且能够发起对学生和家长单独的沟通，这些会话将在学生端和家长端单独呈现，确保了学生和家长的隐私，如图 5.14 所示。

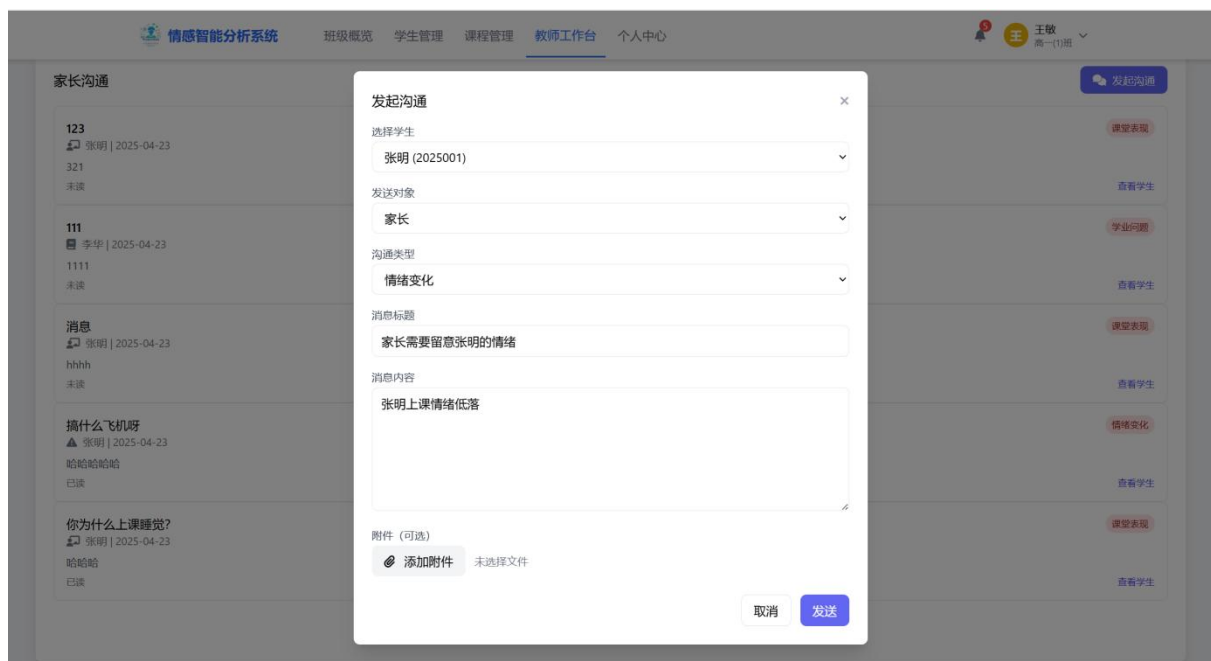


图 5.14 教师端工作台下半部分界面

5、教师个人中心功能设计

教师个人中心上侧用于可视化教师用户的基础信息，中间层用于统计一些班级信息和情绪监测，快捷工具和消息通知都是基于已有模块的再排布，方便教师查看和跳转，具体功能如图 5.15 所示。

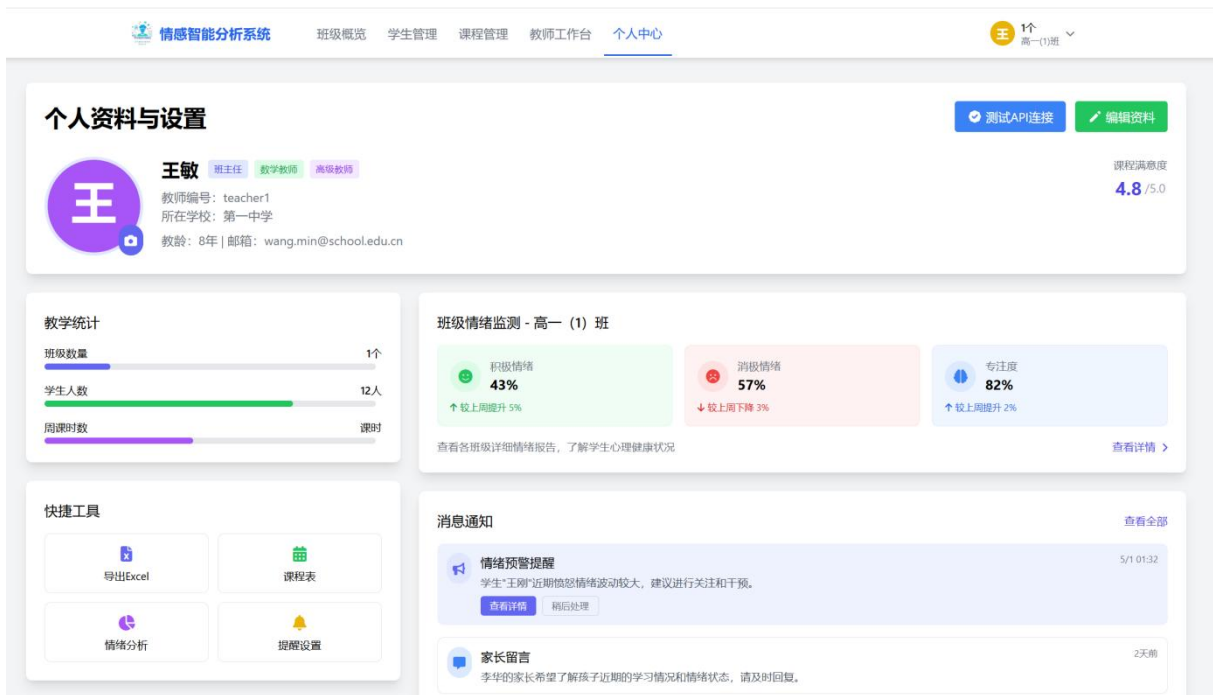


图 5.15 教师端个人中心界面

5.2.4 学生端功能设计

情感智能分析系统的学生模块设计以自我情绪认知与学习辅助为核心理念，为学生用户提供了实时情绪监测、历史数据追踪和个性化学习支持功能。通过直观的仪表盘展示情绪状态，结合 AI 互动助手提供心理疏导，同时整合课程安排与个人资料管理，形成了完整的情感与学习生态系统，帮助学生在认识自我情绪变化的基础上，优化学习状态，促进身心健康发展。

1、学生情绪仪表盘功能设计

学生仪表盘页面是学生端登录后的首页，为学生提供单模态情绪监测和多模态数据可视化功能。以多功能模块布局为设计特点，将复杂的情绪数据转化为直观易懂的图表和指标。

为了突出系统的功能，页面专门设计了各模态情绪识别的可视化模块，这部分内容仅对学生本人可见，以保护学生的个人隐私。

在仪表盘的上半部分，左上方是实时心率监测模块，数据由 ESP32 主板上的心率传感器提供；中间部分是表情情绪识别模块，由 ESP32CAM 提供支持；右侧是 AI 互动助手模块，点击即可跳转至 AI 助手功能。此外，左下角还设有实时情绪反馈模块，该模块通过 4.1 节提出的多模态融合算法计算出的结果进行可视化展示，为学生提供即时的情绪状态反馈，如图 5.16 所示。

学生仪表盘的下半部分左下角是“今日情绪指数曲线”模块，通过动态曲线图展示学生当天的情绪变化，帮助学生直观了解自身情绪的波动情况。中间区域则是语音和文本情绪识别模块。学生可以通过点击语音输入按钮，利用语音识别设备或直接用浏览器进行语音输入，并通过调用百度语音识别 API 将语音转化为文本。

此外，学生也可以直接输入文本内容，系统会调用百度文心 API 进行情感分析。经过第 4.1.2 节的程序设计，系统能够接收大模型以 JSON 格式回传的字段，并将其可视化为情感分析结果。

其中，详细情感分布展示了每种情感的强度，而右侧的情绪建议模块则提供了一些预设的文字

内容，以可视化形式为学生提供情绪管理建议，如图 5.17 所示。

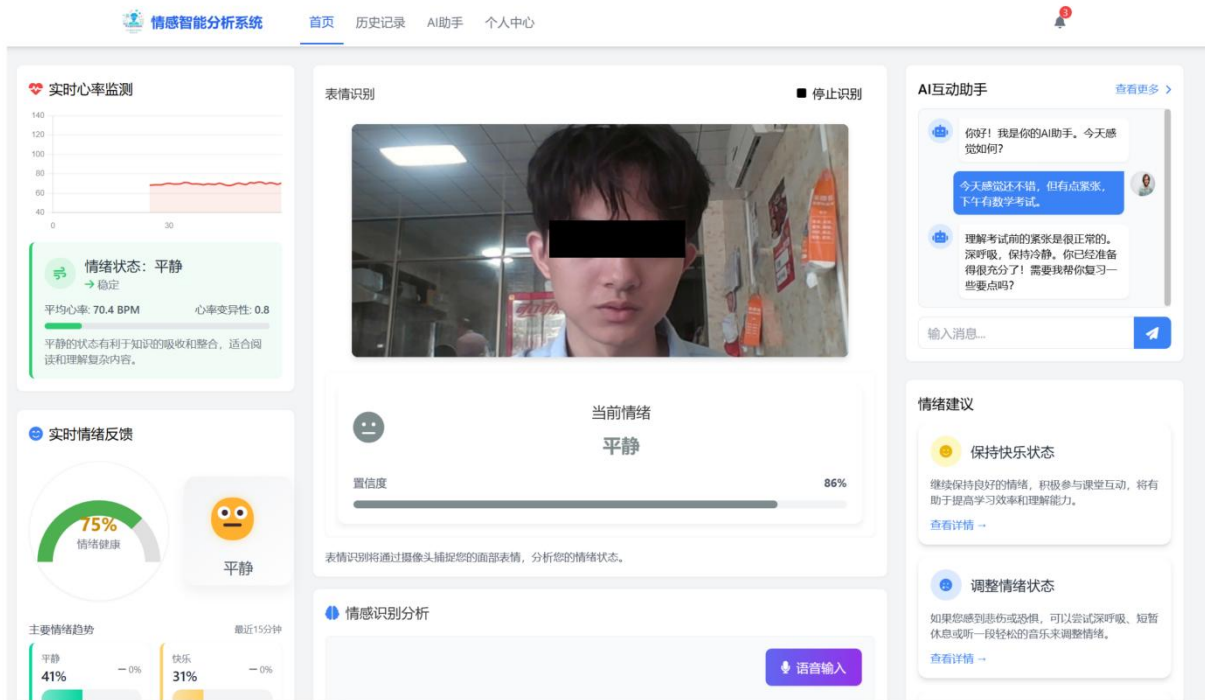


图 5.16 学生端首页仪表盘上半部分界面

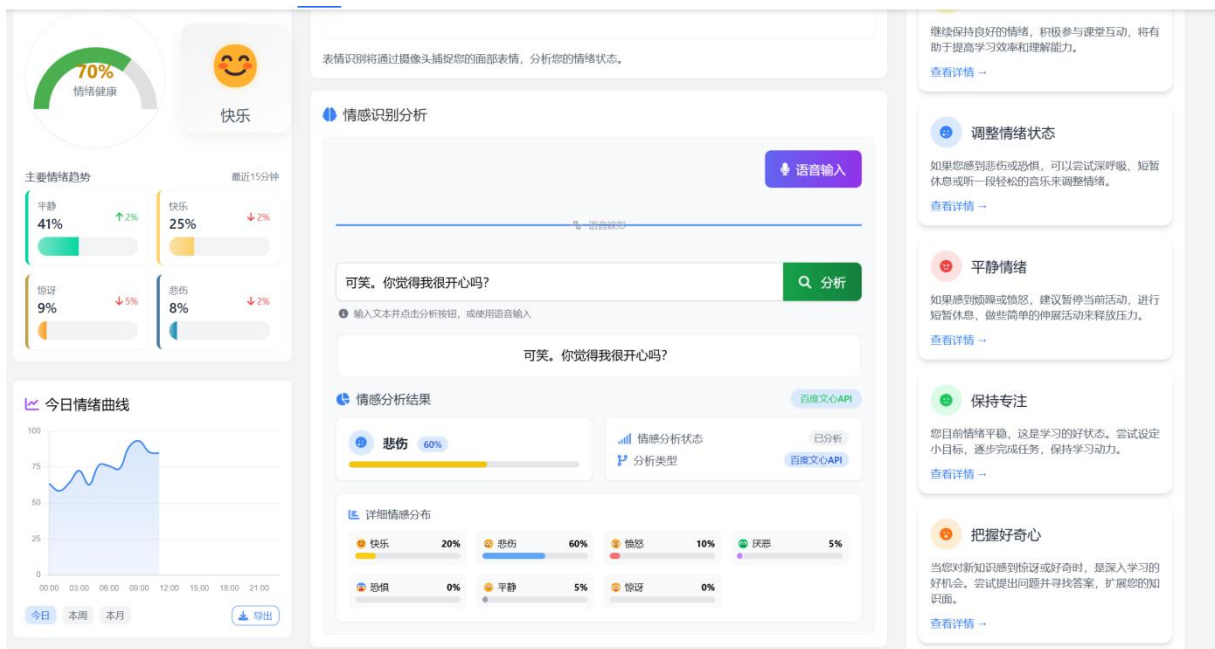


图 5.17 学生端首页仪表盘下半部分界面

2、历史记录功能设计

历史记录界面左上角是情绪时间线模块，该功能主要采集学生的语音文本内容，并记录学生激活语音模块后的每次记录，如图 5.18 所示。页面支持筛选功能，方便学生快速定位特定时间段的情绪记录，同时提供导出 Excel 功能，便于学生将数据进行本地化保存和进一步分析。

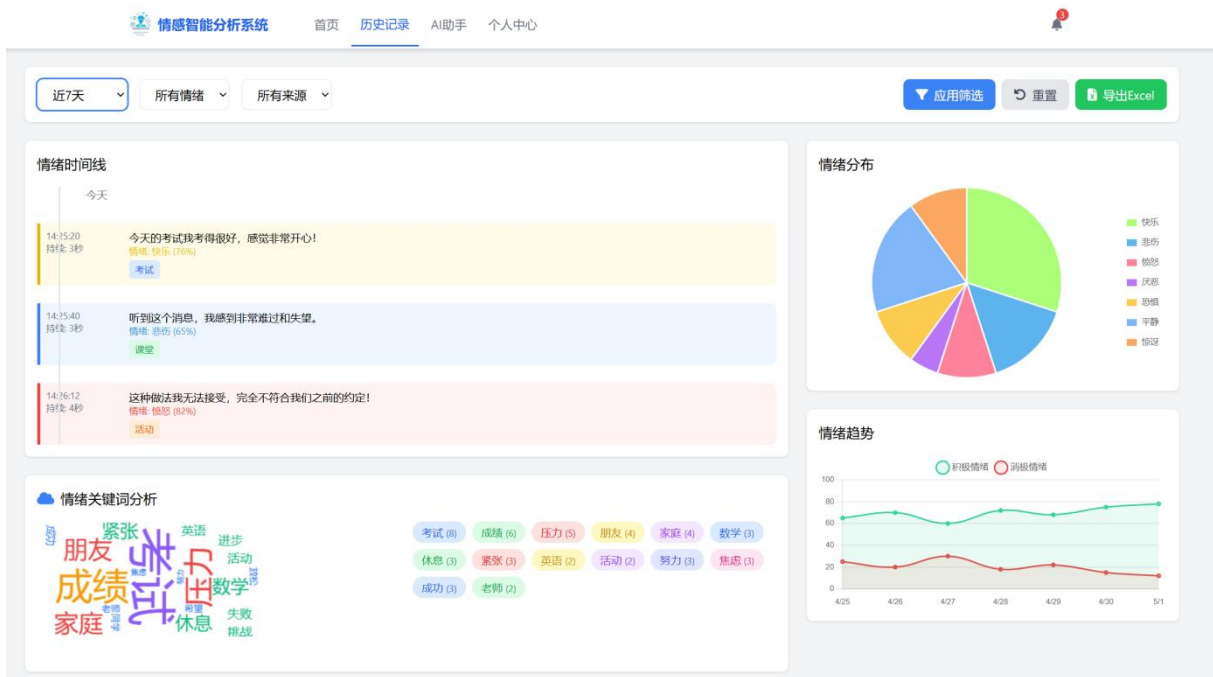


图 5.18 学生端历史记录部分界面

右上部分是情绪分布饼图，与雷达图类似，通过直观的图形展示学生情绪的多维度分布情况，帮助学生快速了解情绪的整体构成。左下方是情绪关键词分析模块，基于学生的所有记录，进行词频统计，并以词云图的形式进行可视化展示。右下角是情绪趋势模块，展示了积极情绪和消极情绪的占比，通过动态图表的形式，直观呈现学生情绪的正负倾向变化，帮助学生更好地理解自身情绪的整体走向。

3、AI 助手功能设计

学生智能助手页面整合了基于人工智能的对话系统，致力于为学生提供学业辅导、情感支持和生活建议，智能咨询平台。页面具备新建会话功能，通过基于情绪 Prompt 设计的百度文心大模型与学生进行互动聊天。该设计不仅能够帮助学生疏导不良情绪，还能解答学生在学习和生活中遇到的问题，满足大语言模型的基本功能需求。具体的 Prompt 设计如表 5.1 所示。

表 5.1 AI 聊天助手提示词

提示词
我是一名学生，请你扮演一个名为"心灵小助手"的情感智能 AI 助手，专注于帮助我处理学习和生活中的情绪问题。
当前我的主要情绪是: $\${emotionResult.primary}$
情绪强度: $\${emotionResult.intensity}$
情绪得分: $\${emotionResult.score}$
//上述三个变量用于后端接收数据并可视化
请根据我的情绪状态，提供适当的支持和建议。你的回复应该:
1. 表现出对我情绪的理解和共情
2. 给予针对性的建议和支持
3. 语气温和，避免过度正面或消极
4. 回复长度适中，不要太长
5. 不要直接提及你对我情绪的具体分析数据

提示词

如果我情绪消极(如悲伤、焦虑、愤怒)，提供安抚和缓解方法。

如果我情绪积极(如喜悦、平静、自信)，给予鼓励和保持这种状态的建议。

如果我表现出困惑或疲惫，提供清晰的指导和休息建议。

在图 5.19 所示的页面中，学生可以新建会话，并能够点击快捷话题与 AI 进行对话。右侧设有对话情绪分析模块，能够实时展示对话中的情绪分析结果。

未来将允许学生根据自身需求调整大语言模型的参数。为此，专门设计了大语言模型参数设计页面，学生可以在此调整常见的模型参数，如温度、最大长度等，以优化对话体验和情绪分析的准确性，如图 5.20 所示。



图 5.19 学生端 AI 助手部分界面

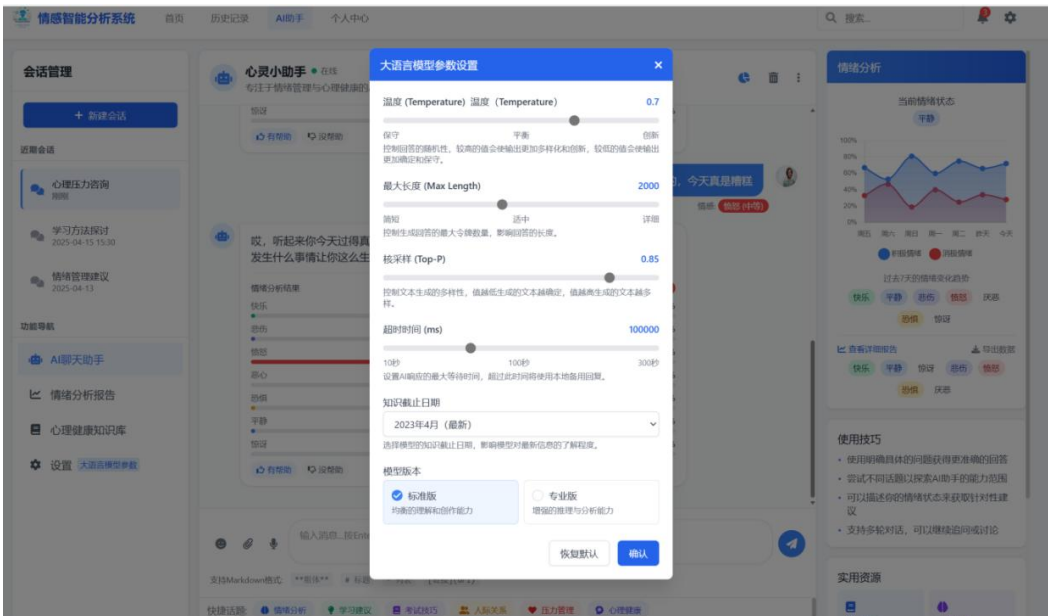


图 5.20 学生端 AI 助手参数设置部分界面

4、学生个人中心功能设计

与教师端功能相似，用户可以在此页面编辑学生的个人信息，同时提供情绪健康的简要概览，通过可视化图表呈现学生的情绪状态。此外，系统还支持数据导出功能，方便用户获取和分析数据。在界面右侧，设有教师留言区域，教师可以在教师工作台选择“学生”类型，向学生发送留言。值得注意的是，学生的学习表现需要在教师工作台手动输入，如图 5.21 所示。

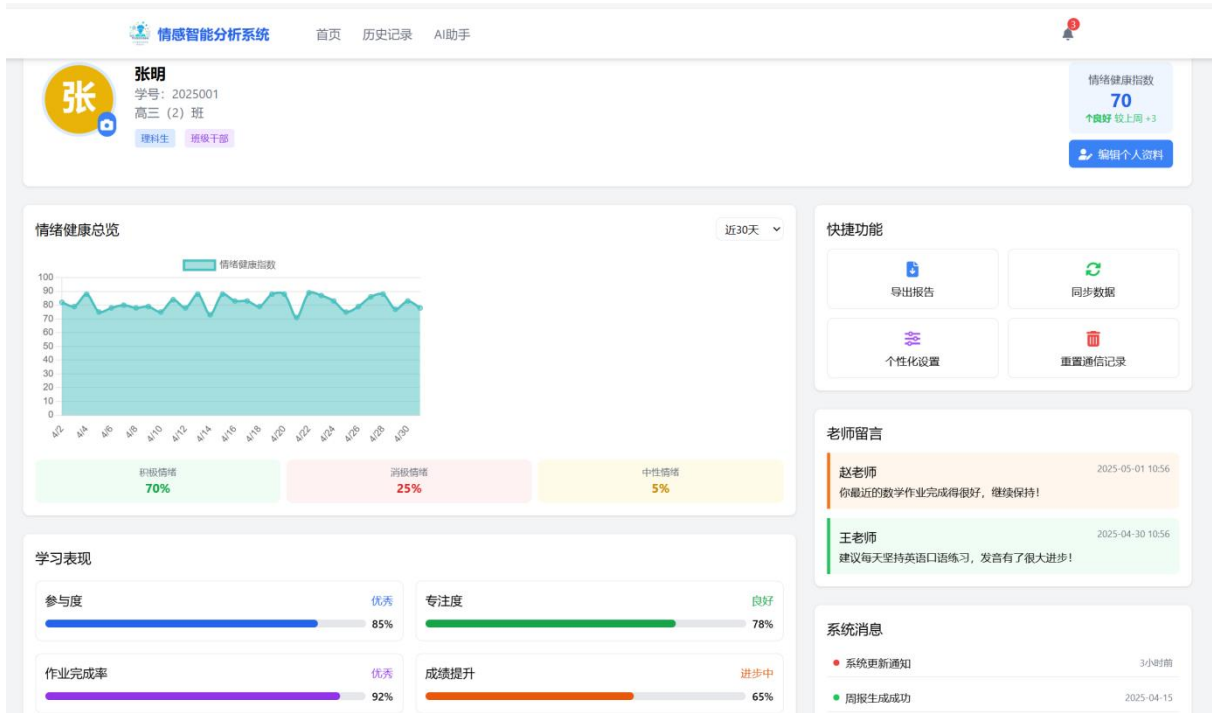


图 5.21 学生端个人中心界面

5.2.5 家长端功能设计

情感智能分析系统的家长功能设计以情绪监护与学业协同为核心理念，为家长用户提供全面的孩子情绪监测、学业表现追踪和家校沟通功能，通过数据可视化与智能分析，助力家长实时了解孩子的情绪状态和学习表现，促进家校协同育人。

1、孩子情绪监护功能设计

情绪监护页面是家长登录系统后的首页，它为家长提供了一个全面的孩子动态情感监测视图，如图 5.22 所示。该页面通过直观且多维度的可视化图表和数据展示，使家长能够清晰地了解孩子在不同时间段内的情绪变化趋势、情绪状态的分布情况以及潜在的情绪波动。它不仅呈现了孩子当前的情绪状态，还能够通过历史数据对比，帮助家长发现潜在的情绪问题，从而更好地关注孩子的心理健康。

页面顶部的今日情绪概览部分直观展示当前情绪分布。在页面下方，近一周学习情绪图表展示了包括专注度、参与度和情绪指数在内的情感与学习关联数据。页面右侧信息栏中的孩子情绪健康图表以饼图形式展现各类情绪的分布比例，情绪健康指数区域显示总体得分及趋势变化，情绪趋势图则直观呈现近 7 天内情绪的波动变化。



图 5.22 家长端情绪监护界面

2、孩子学业表现功能设计

学业表现页面能够提供孩子学业情况的全面分析报告。页面顶部展示了学生的基本信息，并配备了打印报告、导出 PDF 和导出 Excel 三个功能按钮，方便家长保存和分享孩子的学业报告。成绩趋势图采用折线图形式展示了孩子在语文、数学、英语和物理四个主要科目的成绩变化趋势，如图 5.23 所示。

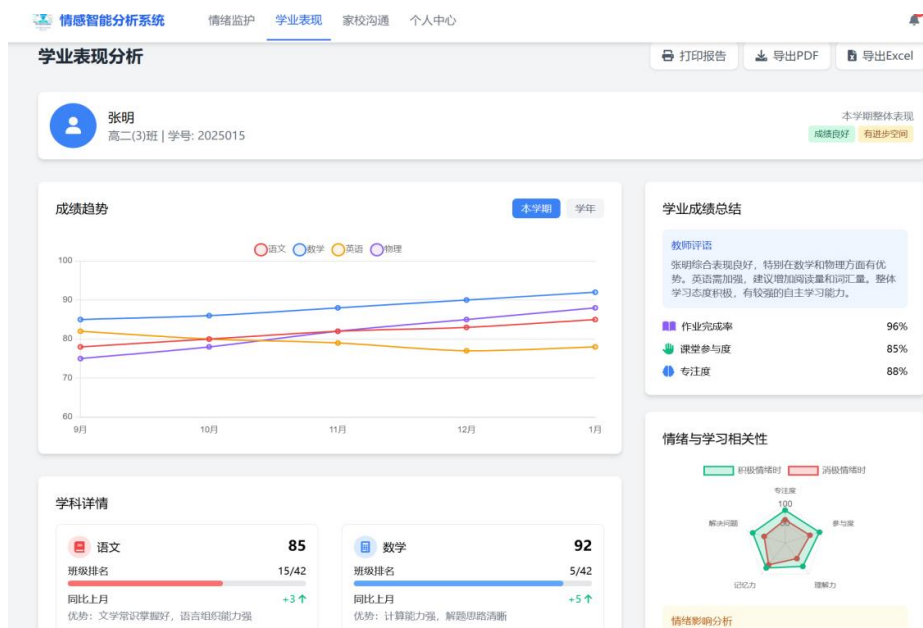


图 5.23 家长端学业表现界面

学科详情部分采用卡片式设计，分别展示各科目的具体表现，包括分数、班级排名、同比上月变化及优势分析。情绪与学习相关性部分通过雷达图直观展示积极情绪和消极情绪状态下学习能力的差异，强调了情绪管理对学习效果的重要影响。

3、家校沟通页面

页面主体部分是沟通记录列表，家长可以通过教师、沟通类型和时间范围三个维度进行筛选查询，系统会根据筛选条件动态加载相关的沟通记录。每条沟通记录清晰显示了消息标题、教师信息、时间、沟通类型、详细内容，并提供了回复功能，支持家长与教师进行持续的对话交流，这与教师

端和学习端类似，如图 5.24 所示。

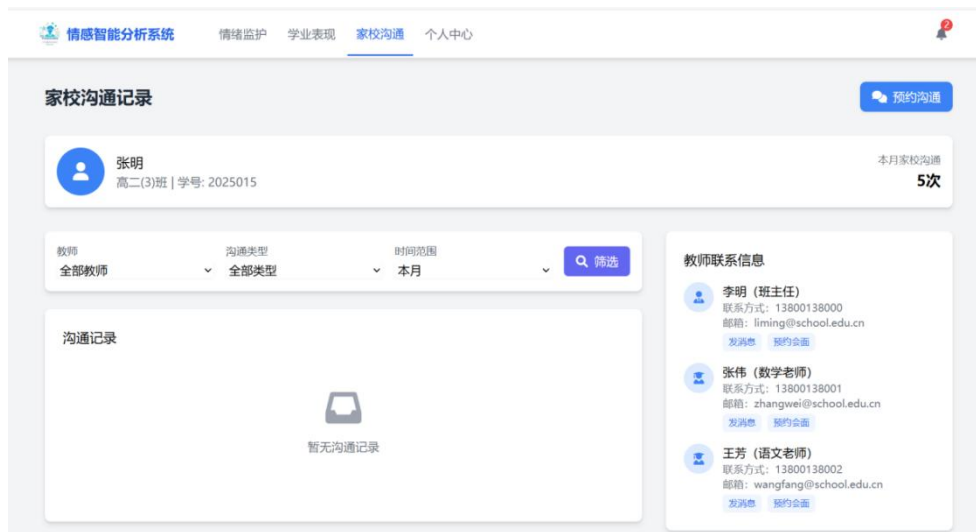


图 5.24 家长端家校沟通界面

4、家长个人中心功能设计

个人中心页面为家长提供了个人信息管理和系统便捷入口。页面顶部是个人信息卡片和孩子的信息概览，其他均是常规功能，与教师端和学生端类似，如图 5.25 所示。

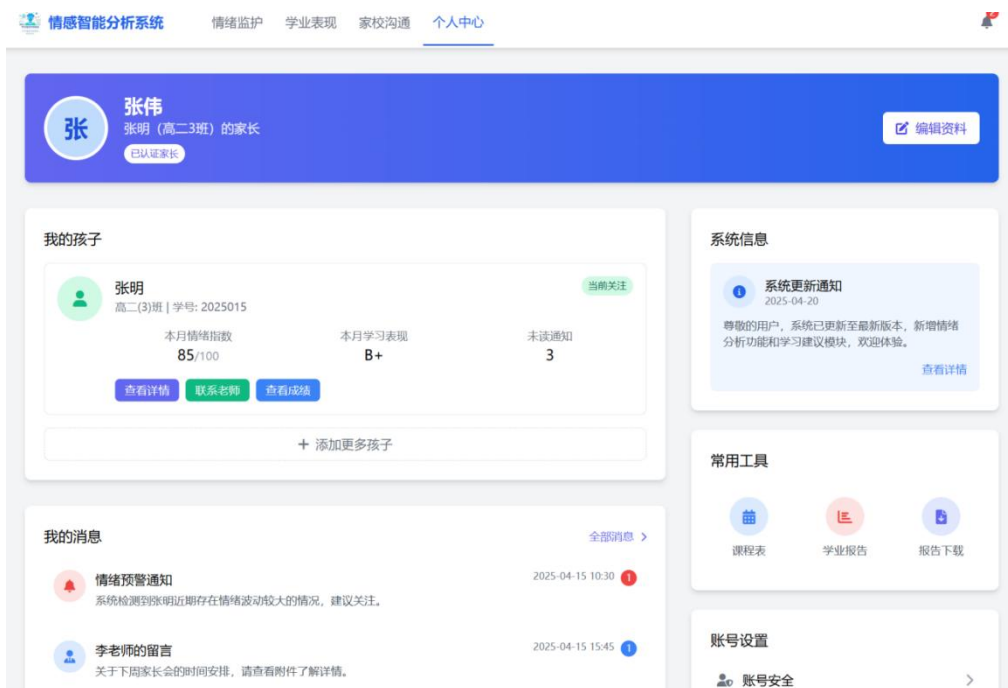


图 5.25 家长端个人中心界面

5.3 数据库设计

在学生情感智能分析系统平台的构建中，用户登录注册、单模态情绪指数存储以及多模态情绪指数存储等功能模块的存在，使得数据库的设计成为整个系统的核心环节。因此，本节将重点围绕数据库的概念设计、逻辑设计以及表结构设计三个层面展开详细阐述。

5.3.1 概念设计

在数据库的概念设计阶段，我们对情绪智能教育平台的数据需求进行了系统化分析，构建了反

映实体间本质关系的概念数据模型。该模型揭示了系统中情绪数据与学习表现之间的逻辑联系，为后续实现提供了理论框架^[40]。本节通过实体-联系模型（E-R 图）的方式展现系统各实体之间的关系。

由于数据表过多，绘制的 E-R 图过于混乱，因此分别绘制不同场景下的 E-R 图。

图 5.26 为教师端管理系统 E-R 图：

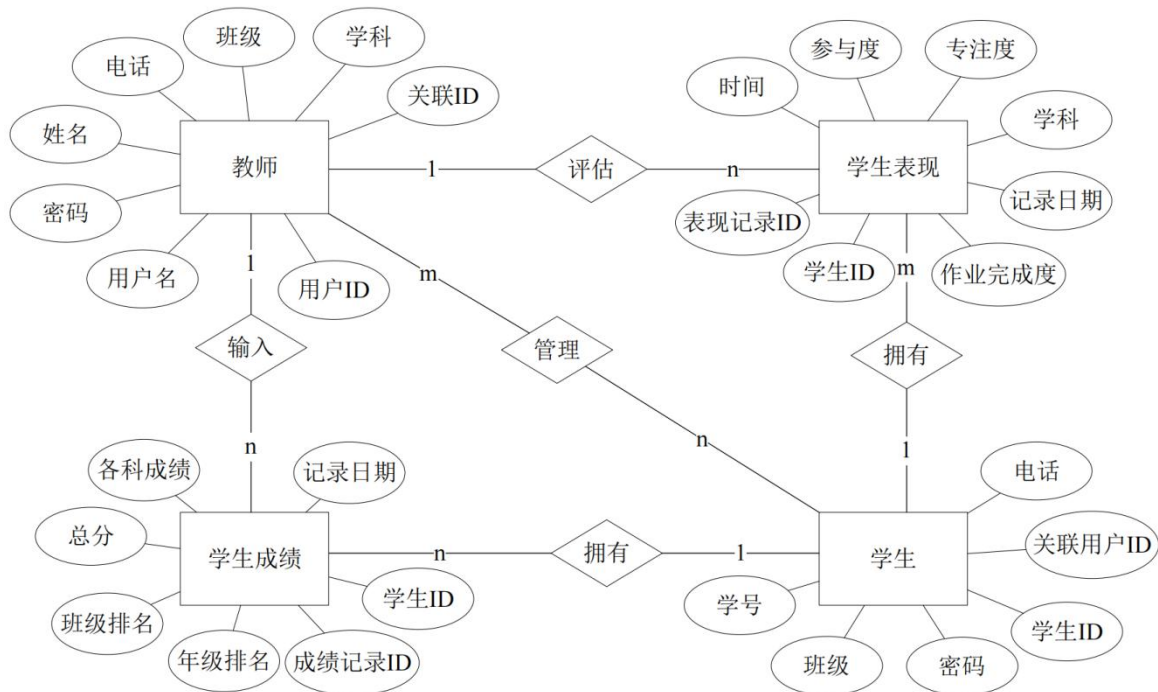


图 5.26 教师端管理系统 E-R 图

图 5.27 为情绪管理功能 E-R 图：

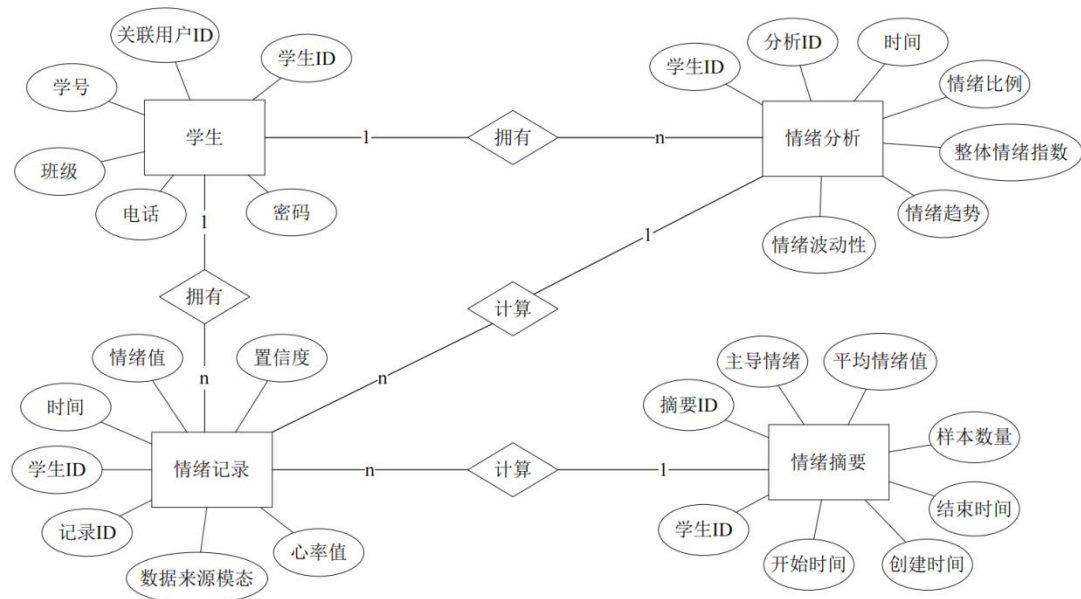


图 5.27 情绪管理系统 E-R 图

图 5.28 为家校沟通功能 E-R 图：

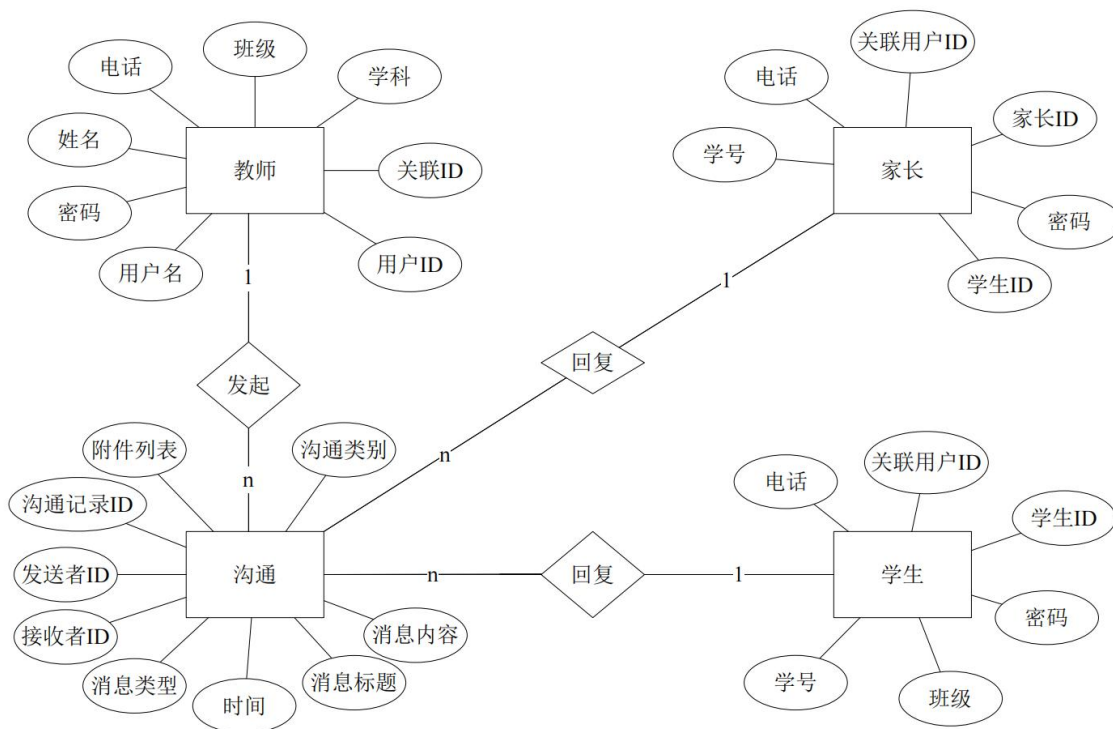


图 5.28 家校沟通 E-R 图

5.3.2 逻辑设计

数据库逻辑设计的核心任务在于将前期构建的概念模型有效地映射为数据库系统所接纳的数据模型逻辑架构。前文对数据库进行了概念模型的设计，明确了情绪智能教育系统实体的各个属性，将概念模型转换为关系模型，情绪智能教育系统数据库的关系模式如下所示：

用户表 { 用户 ID, 用户名, 密码, 姓名, 电话, 角色, 关联 ID, 班级, 学科 }

学生表 { 学生 ID, 关联用户 ID, 学号, 班级 }

情绪记录表 { 记录 ID, 学生 ID, 时间, 数据来源模态, 情绪值, 置信度, 心率值 }

情绪摘要表 { 摘要 ID, 学生 ID, 开始时间, 结束时间, 样本数量, 平均情绪值, 主导情绪, 情绪变化幅度, 时间 }

情绪分析表 { 分析 ID, 学生 ID, 整体情绪指数, 情绪波动性, 情绪趋势, 情绪比例, 时间 }

学生表现表 { 表现记录 ID, 学生 ID, 记录日期, 学科, 专注度, 参与度, 作业完成度, 时间 }

学生成绩表 { 成绩记录 ID, 学生 ID, 记录日期, 语文成绩, 数学成绩, 英语成绩, 物理成绩, 总分, 班级排名, 年级排名, 时间 }

沟通表 { 沟通记录 ID, 发送者 ID, 接收者 ID, 消息类型, 消息标题, 消息内容, 沟通类别, 附件列表, 时间 }

在设计过程中，情绪记录表的情绪值、摘要表的平均情绪值以及情绪分析表的情绪比例均涵盖七种情绪的具体数值，但在写作中统一以“情绪值”进行概括，以保持简洁性。

5.3.3 表设计

结合情绪智能教育系统功能特点和逻辑设计，将从用户表、学生表、情绪记录表、情绪摘要表、

情绪分析表、学生表现表、学生成绩表、沟通表八个方面进行表设计。八张系统数据库表如下表 5.2、表 5.3、表 5.4、表 5.5、表 5.6、表 5.7、表 5.8、表 5.9 所示：

表 5.2 用户表

字段名	数据类型	标识	约束
用户 ID	INT(11)	用户唯一标识	主键，自增
用户名	VARCHAR(50)	用户登录名	非空，唯一
密码	VARCHAR(255)	加密存储的密码	非空
姓名	VARCHAR(50)	用户真实姓名	非空
电话	VARCHAR(20)	联系电话	
角色	VARCHAR(20)	用户角色	非空
关联 ID	INT(11)	关联其他表 ID	
班级	VARCHAR(50)	所属班级	
学科	VARCHAR(50)	教师所教学科	

表 5.3 学生表

字段名	数据类型	标识	约束
学生 ID	INT(11)	学生唯一标识	主键，自增
关联用户 ID	INT(11)	关联用户表 ID	非空，外键
学号	VARCHAR(50)	学生学号	非空，唯一
班级	VARCHAR(50)	所在班级	非空
时间	DATETIME	创建或更新时间	非空

表 5.4 情绪记录表

字段名	数据类型	标识	约束
记录 ID	INT(11)	记录唯一标识	主键，自增
学生 ID	INT(11)	关联学生表 ID	非空，外键
时间	DATETIME	情绪记录时间	非空
数据来源模态	VARCHAR(50)	情绪数据来源	非空
情绪值	JSON	各种情绪值	非空
置信度	FLOAT	情绪识别置信度	非空
心率值	INT(11)	学生心率值	

表 5.5 情绪摘要表

字段名	数据类型	标识	约束
摘要 ID	INT(11)	摘要唯一标识	主键，自增
学生 ID	INT(11)	关联学生表 ID	非空，外键
开始时间	DATETIME	摘要开始时间	非空
结束时间	DATETIME	摘要结束时间	非空
样本数量	INT(11)	情绪记录样本数	非空
平均情绪值	JSON	各情绪平均值	非空
主导情绪	VARCHAR(50)	主要情绪类型	非空
情绪变化幅度	FLOAT(10)	情绪波动大小	非空
时间	DATETIME	摘要创建时间	非空

表 5.6 情绪分析表

字段名	数据类型	标识	约束
分析 ID	INT(11)	分析唯一标识	主键, 自增
学生 ID	INT(11)	关联学生表 ID	非空, 外键
整体情绪指数	FLOAT(10)	综合情绪指标	非空
情绪波动性	FLOAT(10)	情绪波动程度	非空
情绪趋势	VARCHAR(50)	情绪变化趋势	非空
情绪比例	JSON	不同情绪占比	非空
时间	DATETIME	分析创建时间	非空

表 5.7 学生表现表

字段名	数据类型	标识	约束
表现记录 ID	INT(11)	表现记录唯一标识	主键, 自增
学生 ID	INT(11)	关联学生表 ID	非空, 外键
记录日期	DATE	表现记录日期	非空
学科	VARCHAR(50)	学科名称	非空
专注度	INT(11)	课堂专注度	非空
参与度	INT(11)	课堂参与度	非空
作业完成度	INT(11)	作业完成情况	非空
时间	DATETIME	记录创建时间	非空

表 5.8 学生成绩表

字段名	数据类型	标识	约束
成绩记录 ID	INT(11)	成绩记录唯一标识	主键, 自增
学生 ID	INT(11)	关联学生表 ID	非空, 外键
记录日期	DATE	成绩记录日期	非空
语文成绩	FLOAT(5)	语文学科成绩	
数学成绩	FLOAT(5)	数学科目成绩	
英语成绩	FLOAT(5)	英语科目成绩	
物理成绩	FLOAT(5)	物理科目成绩	
总分	FLOAT(5)	总分成绩	
班级排名	INT(11)	班级内排名	
年级排名	INT(11)	年级内排名	
时间	DATETIME	记录创建时间	非空

表 5.9 沟通表

字段名	数据类型	标识	约束
沟通记录 ID	INT(11)	沟通记录唯一标识	主键, 自增
发送者 ID	INT(11)	关联发送用户 ID	非空, 外键
接收者 ID	INT(11)	关联接收用户 ID	非空, 外键
消息类型	VARCHAR(20)	消息类型	非空
消息标题	VARCHAR(100)	消息标题	
消息内容	TEXT(1000)	消息内容	非空
沟通类别	VARCHAR(20)	沟通目的	非空

字段名	数据类型	标识	约束
附件列表	JSON	附件链接	
时间	DATETIME	消息发送时间	非空

5.4 本章小结

本章详细阐述了学生情感智能分析软件端的设计方案，涵盖了开发环境选择、功能模块设计及数据库构建。系统采用轻量级 Web 架构，实现了教师、学生与家长三端用户的差异化功能体系，包括情绪监测、数据可视化、AI 交互与家校沟通等核心功能。通过设计的数据库结构，系统能够有效存储和管理多模态情绪数据，为学生情感智能分析与监测提供全面的技术支持。

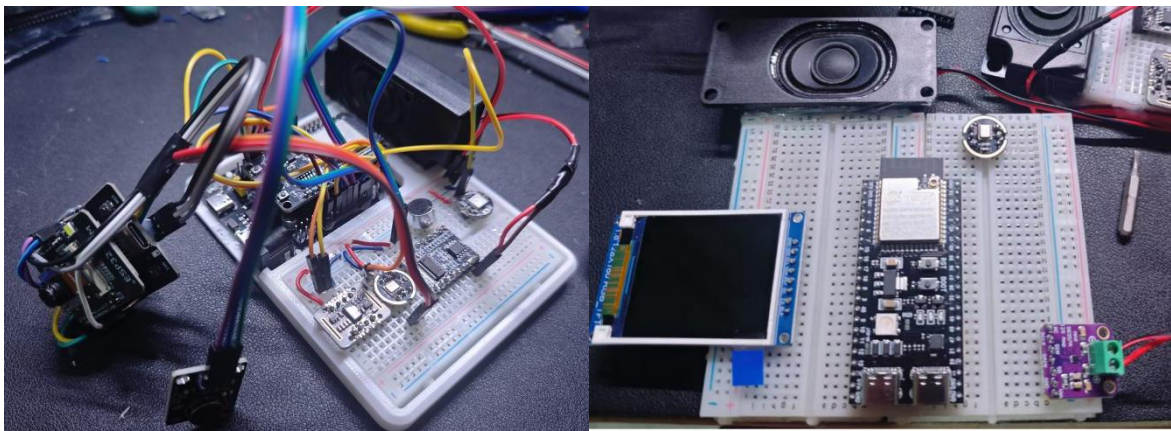
6 系统调试与分析

本章主要从 ESP32 单片机系统、Web 端云平台、前端用户界面三个系统出发进行系统调试。主要包括 ESP32 主控板及其外设模块的硬件安装和各项功能的调试；软件系统测试，包括白盒测试和黑盒测试，覆盖情感分析、多模态融合算法、预测算法等核心模块；前端用户界面各项功能的调试，包括登录注册、大模型情感分析、数据导出、AI 聊天助手等功能验证。

6.1 硬件系统安装和调试

6.1.1 系统安装

根据单片机系统硬件设计，将 ESP32 主控板、ESP32S3 主控板、ESP32CAM 摄像头模块、INMP441 麦克风、MAX30102 心率传感器、WS2812 彩色灯带、震动马达、MAX98357A 音频模块、TFT 显示屏等模块按照对应接口连接起来。硬件连接如图 6.1 所示。



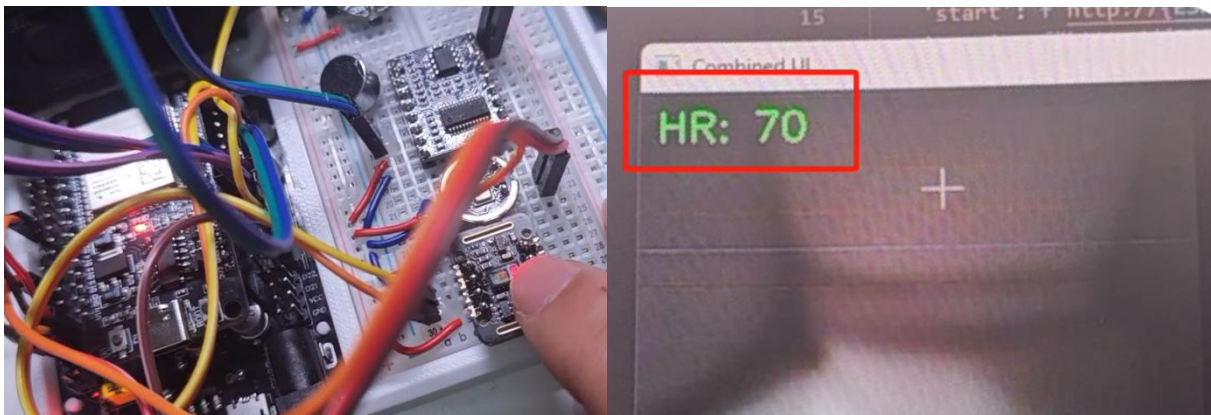
(a) ESP32 主板连接实物图

(b) ESP32S3 主板连接实物图

图 6.1 硬件安装连接图

6.1.2 心率传感器调试

当 ESP32 主板接入 Type-C 接口的电源后，用户可以通过按下按键来触发单片机进入开机模式。如图 6.2 所示，学生将手指平稳地按压在心率传感器的指定位置。此时，心率传感器将实时采集用户的心率数据，并通过 ESP32CAM 模块将这些数据传输至电脑端，以便进行进一步的分析处理。



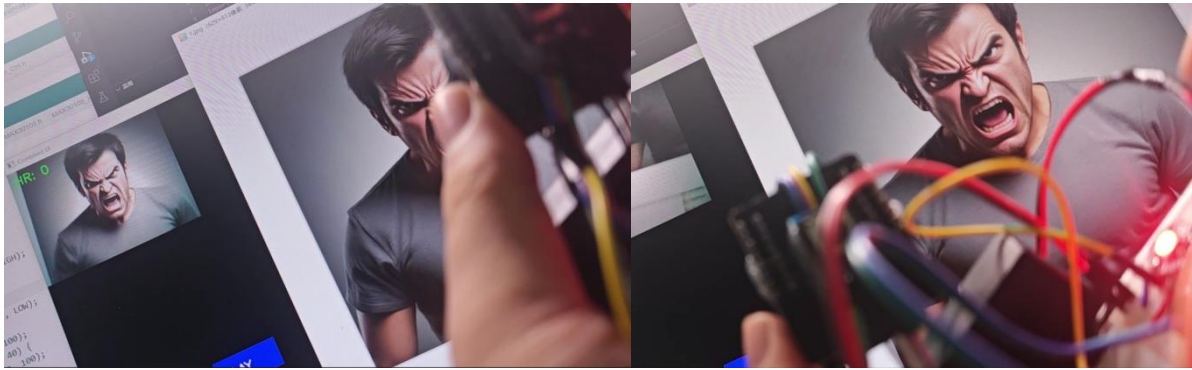
(a) 用户手指按压传感器视图

(b) 实时数据采集

图 6.2 ESP32 主板心率传感器调试图

6.1.3 ESP32CAM 摄像头调试

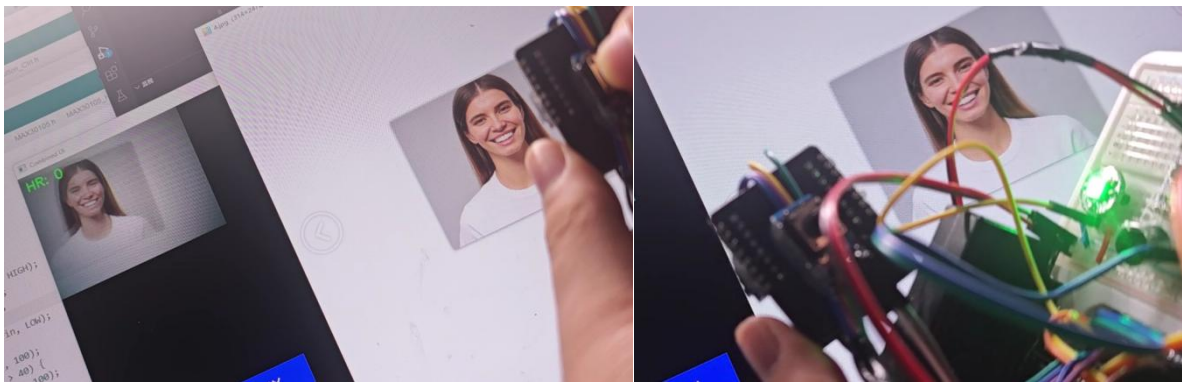
开机后，将硬件设备放置于桌面，系统即可自动启动人脸表情识别功能。为了便于调试，我们将摄像头对准预先准备好的标有“愤怒”情绪的图片。当系统检测到“愤怒”情绪的持续时间超过设定阈值时，如图 6.3 所示，设备会亮起红灯以示提醒。类似地，当系统检测到“快乐”的持续时间超过设定阈值时，如图 6.4 所示，设备会亮起绿灯。



(a) 摄像头对准“愤怒”图片

(b) 红色灯光反馈

图 6.3 消极情绪反馈调试



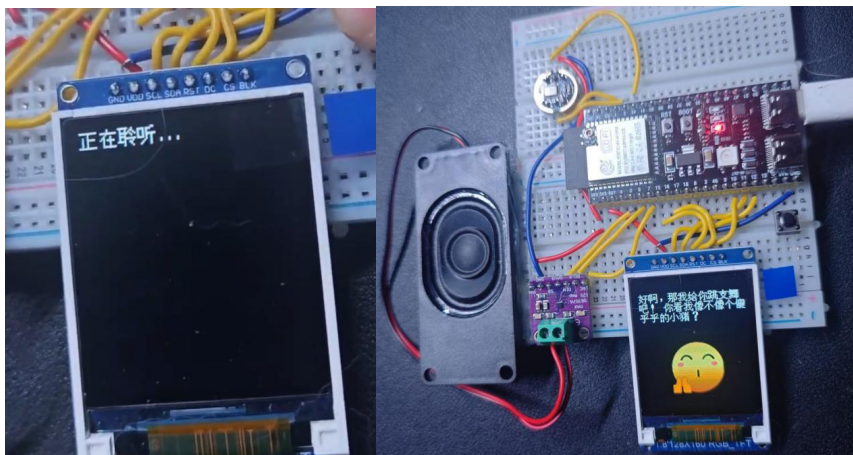
(a) 摄像头对准“快乐”图片

(b) 绿色灯光反馈

图 6.4 积极情绪反馈调试

6.1.4 ESP32S3 聊天对话调试

开机后，将硬件设备放置于桌面，系统可以通过按压按键模式打开语音聊天对话功能，也可以通过指定称呼来唤醒。



(a) 用户说话时视图

(b) AI 回复视图

图 6.5 ESP32S3 聊天对话调试图

为了便于反馈情绪，系统将通过 TFT 显示屏用于实时显示用户的对话内容，并且通过设计 emoji 表情来反馈用户的情绪状态，当用户正在说话时，屏幕会显示正在聆听，并且将用户说的话转成文本显示在屏幕上，通过调用 API 实现对话功能，如图 6.5 所示。

6.2 软件系统测试

6.2.1 白盒测试

白盒测试主要关注系统内部逻辑结构和代码实现，通过对系统内部设计和编码的理解来设计测试用例^[41]。针对本系统，白盒测试主要包括以下部分：

- (1) 进行代码覆盖率测试，确保测试用例能够覆盖系统中的所有代码路径，包括条件语句、循环语句和异常处理等；
- (2) 开展路径测试，检验所有可能的程序执行路径，特别是多模态融合决策算法中的条件分支；
- (3) 最后，实施数据流测试，分析变量在程序中的定义和使用情况，特别关注多模态数据的处理流程。

6.2.2 黑盒测试

黑盒测试不考虑系统内部结构和实现细节，而是通过系统的输入和输出来评估系统功能^[42]。针对本系统的黑盒测试主要包括以下部分：

- (1) 将输入数据划分为有效等价类和无效等价类，测试系统对不同类型输入的处理能力。
- (2) 测试系统对边界条件的处理能力，如情绪阈值边界、极端心率值等。
- (3) 基于经验预测系统可能出现的错误，设计针对性测试用例。

6.2.3 测试用例

情感分析测试用例如表 6.1 所示：

表 6.1 情感分析测试用例表

测试 ID	测试项目	测试输入	预期结果	实际结果
TA-001	面部识别准确性	不同情绪的图像数据集	正确识别 7 种基本情绪	与预期一致
TA-002	情绪识别实时性	持续视频流输入	在 1 秒内完成情绪识别	与预期一致
TA-003	文本情绪分析	不同情绪倾向的文本样本	返回正确的情绪类别	与预期一致
TA-004	情绪识别鲁棒性	低光照、侧脸等挑战场景	仍相对准确地判断情绪	与预期一致
TA-005	大模型情绪回复	表达负面情绪的用户输入	提供情感支持性回复	与预期一致

多模态融合算法测试用例如表 6.2 所示：

表 6.2 多模态融合算法测试用例

测试 ID	测试项目	测试输入	预期结果	实际结果
MA-001	权重动态调整	带缺失值的数据	动态调整权重，保持总权重为 1	与预期一致
MA-003	极端情绪检测	极度消极情绪的数据	触发极度消极情绪提醒	与预期一致
MA-005	情绪判断一致性	重复相同情绪场景数据	多次分析结果保持一致，浮动范围<10%	与预期一致

预测算法测试用例如表 6.3 所示：

表 6.3 预测算法测试用例表

测试 ID	测试项目	测试输入	预期结果	实际结果
PA-001	移动平均预测	30 条情绪数据	使用移动平均法进行情绪预测	与预期一致
PA-002	线性回归预测	150 条情绪数据	使用线性回归模型进行预测	与预期一致
PA-003	随机森林预测	300 条情绪数据	使用随机森林模型进行预测	与预期一致
PA-004	模型自动选择	从 30 增加到 300 条	根据数据量自动切换预测模型	与预期一致

系统功能测试用例如表 6.4 所示：

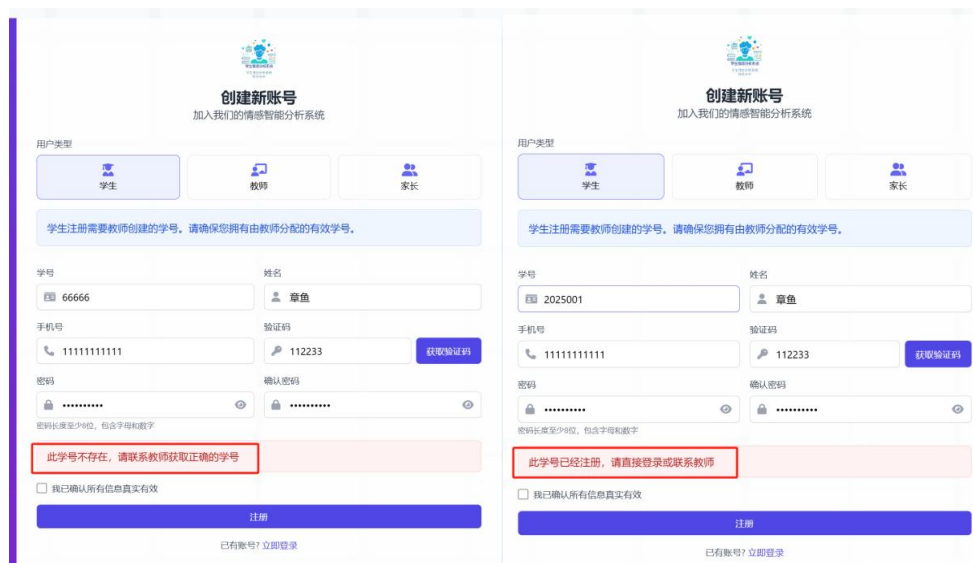
表 6.4 系统功能测试用例表

测试 ID	测试项目	测试步骤	预期结果	实际结果
SF-001	用户登录	输入不同角色账密	验证身份并跳转至对应页面	与预期一致
SF-002	情绪仪表盘	登录学生账号首页	显示情绪状态和历史趋势	与预期一致
SF-003	班级概览	登录教师账号首页	显示班级整体情绪分布等	与预期一致
SF-004	孩子报告	登录家长账号首页	显示子女情绪状态和学习表现	与预期一致
SF-005	AI 情绪助手	通过学生端访问 AI 助手， 输入问题	AI 根据情绪状态 提供个性化回复	与预期一致
SF-006	家校沟通	教师向家长发送消息	家长端能接收消息并回复	与预期一致
SF-007	数据导出	点击导出按钮	生成并下载指定的 Excel 文件	与预期一致
SF-008	情绪预警	模拟极低情绪指数	触发预警，发送通知给教师	与预期一致
SF-009	历史查询	设置日期范围 查询情绪记录	返回符合时间范围的情绪数据	与预期一致
SF-010	教师添加学生账号	教师工作台 添加新学生信息	创建学生账号并可被关联	与预期一致

6.2.4 核心功能测试

1、登录注册功能测试

在本系统的设计中，学生注册需凭借教师分配的学号进行操作。如图 6.6 所示，当输入的学号不存在时，系统将阻止学生完成注册；若学号已被其他学生注册，系统同样会弹出提示信息，提醒用户该学号已被占用。



(a) 学号不存在

(b) 学号已注册

图 6.6 登录注册功能测试

2、大模型情感分析功能测试

为了测试基于百度文心 API 的情感识别功能，我们在 server 文件中设计了终端打印功能，用于接收并展示识别结果。如图 6.7 所示，server 成功接收了文本数据，并将其情感识别的结果可视化展示在前端界面中。

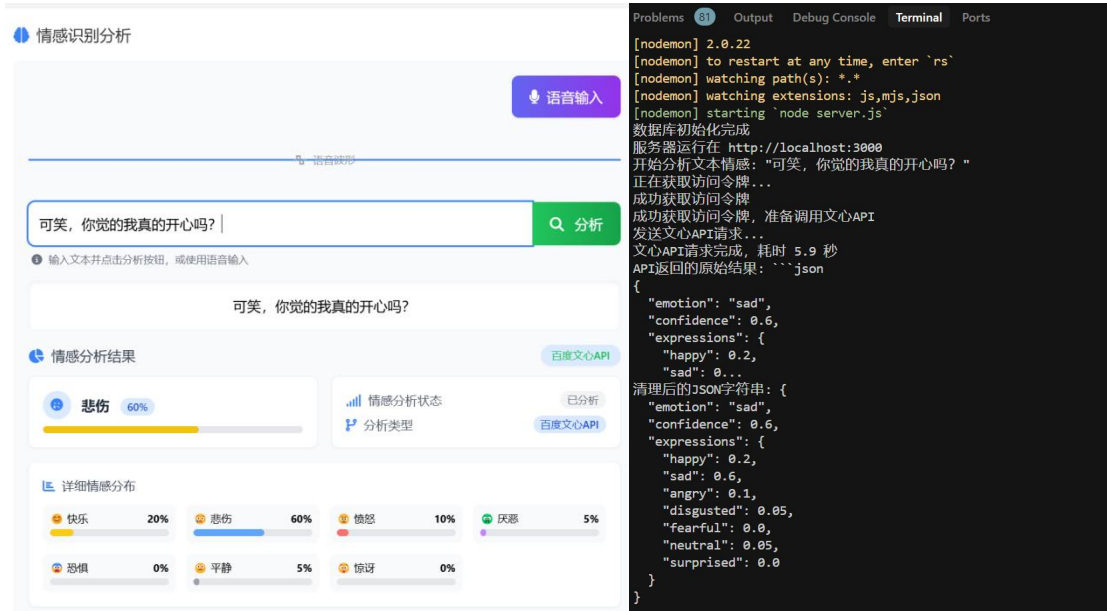


图 6.7 大模型情感分析功能测试

3、数据导出功能测试

在学生端的历史数据导出中，点击导出 Excel，系统能够准确地导出学生的情绪时间线数据和相应的对话内容数据，如图 6.8 所示：

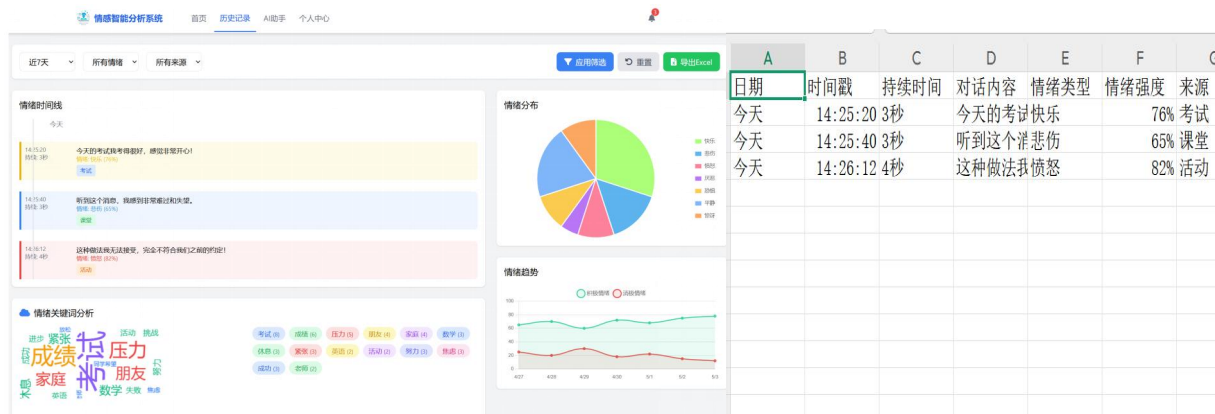


图 6.8 学生端数据导出功能测试

同样地，测试教师端的班级情绪分析报告如图 6.9 所示：



(a) 界面图

(b) 数据导出视图

图 6.9 教师端数据导出功能测试

4、AI 聊天助手功能测试

如图 6.10 所示，当输入情绪分析需求的对话时，大语言模型能够精准分析用户当前的情绪状态，并基于预先设计的 Prompt 进行回应。其回答不仅更加人性化，还带有浓厚的人情味，能够有效发挥疏导情绪的作用。



图 6.10 AI 聊天功能测试

6.3 本章小结

本章从 ESP32 单片机系统、Web 云平台、前端用户界面三个系统出发，进行相应的系统安装、各项功能测试和调试。ESP32 单片机硬件系统整体安装后，通过测试心率传感器、摄像头模块等外

设与主板的连接，检测到情绪数据和生理指标能够成功上传到云平台可视化界面和用户前端界面。接着测试系统是否能够根据识别到的情绪状态按照预设的逻辑控制灯带和其他反馈设备，最后检测了云平台 and 前端界面各个模块的设计与实现，包括登录注册、情感分析、数据导出、AI 聊天助手等核心功能。经测试表明，白盒测试和黑盒测试的结果与预期一致，各功能模块运行效果良好，系统能够有效进行多模态情绪识别与分析，符合该系统的设计要求。

7 总结和展望

7.1 全文总结

本论文设计并实现了一套基于多模态数据与大模型赋能的学生情感智能分析与监测系统，采用了硬件与服务端相结合的方式，其中硬件由 ESP32 和 ESP32S3 双主板硬件系统构成，服务端则由 Node.js 服务器平台和 Web 前端应用组成，这种架构设计有效降低了情感计算的复杂度，能够精准采集学生的面部表情、语音文本和心率生理数据并实现多模态融合分析，很好地达到了本系统设计的预期效果。尽管本文构建的情感监测原型在规模上与实际教育场景存在一定差异，但本文提出的多模态情感融合与大模型赋能的技术原理具有广泛的适用性，如需推广到实际教育环境中，只需扩展硬件部署规模和优化算法参数即可。

本论文主要完成的功能和研究成果如下：

(1) 针对硬件系统的整体架构进行了系统化设计，基于 ESP32 和 ESP32S3 双主板架构实现功能分离，其中 ESP32 主板负责情绪识别与反馈功能，集成了 ESP32CAM 摄像头模块、心率传感器、WS2812 彩色灯带和震动马达等，实现了面部表情识别和生理数据采集；ESP32S3 主板则专注于大模型对话功能，集成了 INMP441 麦克风、MAX98357A 音频解码器和 TFT 显示屏，提供了语音交互和视觉反馈能力。系统通过 WiFi 技术将采集的多模态数据传输至服务端，确保了数据处理的实时性和系统整体的协同工作能力。

(2) 设计并实现了多模态情感指数算法，包括基于 Deepface 的人脸情绪指数算法和基于文心一言的语音情绪指数算法，并通过动态权重融合机制解决了多模态数据的异质性问题。该算法能够有效处理模态缺失情况，并通过心率数据辅助决策，提升了情绪识别的鲁棒性。同时，研发了多模型融合的情绪指数预测算法，根据数据记录数量自适应选择移动平均法、线性回归模型或随机森林模型，为教学环境中的情绪干预提供前瞻性支持。

(3) 开发了面向教师、学生和家長三种角色的 Web 应用系统，实现了差异化的功能设计。教师端提供班级情绪概览、学生管理、课程安排及工作台等功能，帮助教师及时发现学生情绪异常并进行干预；学生端集成了个人情绪监测、历史记录查询、AI 助手对话等功能，支持学生情绪自我认知和情感疏导；家長端则提供子女情绪监护、学业表现查看和家校沟通功能，促进了家校协同育人。系统采用 Node.js 构建后端服务，SQLite 实现数据存储，确保了轻量化部署和高效运行。

(4) 在完成系统硬件与软件部分的设计后，对系统功能进行了全面测试，包括 ESP32 主控板与外设的功能验证、多模态情感融合算法的白盒测试、以及 Web 应用各模块的黑盒测试。测试结果表明，系统能够准确识别七种基本情绪，实现多模态数据的有效融合，并通过直观的可视化界面呈现情绪分析结果。同时，系统的 AI 对话功能能够根据学生情绪状态提供个性化的情感支持，提升了系统的人机交互体验。

7.2 工作展望

尽管本文已竭力在有限的时间和资源条件下完成系统设计与实现，但依然存在诸多可提升之处，以期在未来研究中取得更深层次的完善与优化。具体而言，以下方面有待后续探索：

1、多模态算法的性能优化

当前系统采用的多模态融合算法为决策级融合，虽然计算效率高，但可能无法充分挖掘模态间的互补信息。未来研究可探索特征级融合方法，如基于注意力机制的多模态深度学习模型，进一步提升情感识别的准确性。同时，应研发更适合教育场景的轻量化模型，在保证性能的同时降低计算资源需求，使系统能够在资源有限的教育环境中大规模部署。

2、大模型情感干预策略的优化

目前系统的情感对话功能基于预设的 Prompt 模板，虽能提供基本的情感支持，但缺乏个性化和针对性。未来工作可引入知识图谱和个性化学习模型，结合学生的学科知识背景、学习风格和历史情绪数据，构建更加智能化的情感干预策略。同时，可探索引入模型微调技术，根据教育心理学理论构建专门面向学生情感支持的本地化大模型，提升系统的适应性和专业性。

3、系统隐私保护机制的增强

考虑到学生情感数据的敏感性，未来系统应进一步强化数据隐私保护措施。可探索联邦学习技术在情感分析中的应用，实现模型训练与数据本地化处理的分离，减少原始敏感数据的传输与存储。同时，引入差分隐私等技术，在数据分析过程中添加适量噪声，确保个体隐私不被推断，同时又能保证群体数据分析的有效性。

4、教育价值评估体系的构建

后续研究应关注系统在实际教育环境中的应用效果评估，通过构建科学的评估指标体系，量化系统对学生学习效果、心理健康和师生关系的影响。可开展纵向追踪研究，分析系统长期使用对学生情绪调节能力和学习表现的影响，并根据评估结果不断优化系统功能和算法，使其更好地服务于教育教学实践。

参考文献

- [1]教育部等十七部门关于印发《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)》的通知[J].中华人民共和国教育部公报,2023,(05):4-9.
- [2]郑欣丽. 基于注意力机制的多模态情感分析方法[D]. 湖北:湖北工业大学,2024.
- [3]陈飞宇. 多模态情感分析算法研究[D]. 四川:电子科技大学,2022.
- [4]韩宇. 基于深度学习的基础教育课堂语音情感识别与分析[D]. 北京:北京交通大学,2023.
- [5]黄玲,孙之光,赵迎升. 浅谈智慧校园设计[J]. 长江信息通信,2024,37(2):204-206.
- [6]杜加永. 基于布鲁姆认知的多模态教育情感计算研究[D]. 山东:山东师范大学,2024.
- [7]詹勇鑫. 面向在线课堂评价的多模态情感分析方法研究[D]. 江西:江西师范大学,2024.
- [8]薛凯鹏,徐涛,廖春节. 融合自监督和多层交叉注意力的多模态情感分析网络[J]. 计算机应用,2024,44(8):2387-2392.
- [9]赖河菴,王丽丽,李伶俐,等. 基于记忆迭代网络的多模态情感分析[J]. 智能计算机与应用,2024,14(11):200-205.
- [10]杨梦,林泽珊. 在线学习中多模态情感分析研究综述[C]//2020 第十九届教育技术国际论坛论文集.

2020:46-50.

- [11]Healey J, Picard R. Digital processing of affective signals[C]//Proceedings of the 1998 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP'98 (Cat. No. 98CH36181). IEEE, 1998, 6: 3749-3752.
- [12]Woo S, Zubair M, Lim S, et al. Deep multimodal emotion recognition using modality-aware attention and proxy-based multimodal loss[J]. Internet of Things, 2025, 31: 101562.
- [13]Hu T, Collier N. iNews: A Multimodal Dataset for Modeling Personalized Affective Responses to News[J]. arXiv preprint arXiv:2503.03335, 2025.
- [14]Gan W, Dao M S, Zettsu K, et al. Iot-based multimodal analysis for smart education: Current status, challenges and opportunities[C]//Proceedings of the 3rd ACM Workshop on Intelligent Cross-Data Analysis and Retrieval. 2022: 32-40.
- [15]Lee J P, Jang H, Jang Y, et al. Encoding of multi-modal emotional information via personalized skin-integrated wireless facial interface[J]. Nature Communications, 2024, 15(1): 530.
- [16]Gahlan N, Sethia D. AFLEMP: Attention-based federated learning for emotion recognition using multi-modal physiological data[J]. Biomedical Signal Processing and Control, 2024, 94: 106353.
- [17]蒋艳双,崔璨,逯行,等. 双师课堂中的多模态学习情感分析:关键问题、逻辑理路与实施路线[J]. 现代教育技术,2022,32(4):13-20.
- [18]武法提,高姝睿,赖松. 同步课堂中基于多模态数据的专注度分析[J]. 现代远程教育研究,2025,37(1):93-101.
- [19]程二丽. 情感化智能教室中表情识别算法的改进研究与应用[D]. 湖北:华中师范大学,2016.
- [20]张新新,王芳,杨广学. 机器人技术在自闭症谱系障碍儿童教育中的应用研究进展[J]. 中国特殊教育,2018(11):24-32.
- [21]Taigman Y, Yang M, Ranzato M A, et al. Deepface: Closing the gap to human-level performance in face verification[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2014: 1701-1708.
- [22]俞彤. 文心一言 3.5 与讯飞星火 4.0 大语言模型文言文翻译准确性的分析[J]. 张江科技评论,2024(8):54-56.
- [23]刘月,杨伟樱,李晔. 基于大数据的高职学生心理健康监测系统[J]. 中国新通信,2023,25(24):113-115,15.
- [24]张虎成,李雷孝,刘东江. 多模态数据融合研究综述[J]. 计算机科学与探索,2024,18(10):2501-2520.
- [25]山东三千物联网科技有限公司. 一种集成电机驱动芯片 IO 全拓展的 ESP32 开发板:CN202222889055.5[P]. 2023-03-21.
- [26]韦仕艳,王小波,蔺佳兰,等. 基于 ESP32 CAM 的智能安全门设计概述[J]. 科学与信息化,2021(14):61,66.
- [27]刘宇,李仁想,杨梦凡,等. 语音控制垃圾桶的设计与实现[J]. 无线互联科技,2024,21(17):24-26,30.

- [28] 阚子杨,姜绍君,白煜民,等. 基于 MAX30102 的血氧饱和度监测[J]. 仪器仪表与分析监测,2024(1):1-3.
- [29] 谭伟伟. 基于 ESP32S3 的智能家居控制面板系统设计[J]. 电子产品世界,2023,30(8):79-81.
- [30] 任泽裕,王振超,柯尊旺,等. 多模态数据融合综述[J]. 计算机工程与应用,2021,57(18):49-64.
- [31] Serengil S, Özpınar A. A benchmark of facial recognition pipelines and co-usability performances of modules[J]. Bilişim Teknolojileri Dergisi, 2024, 17(2): 95-107.
- [32] Peng Q, Yuan H. Study on Generating Language Experiment Reading Materials through ERNIE Bot[C]//Proceedings of the 2024 International Symposium on Artificial Intelligence for Education. 2024: 535-539.
- [33] 邓光华. 基于心率变异性的情绪识别研究[D]. 天津:天津工业大学,2021.
- [34] 段利忠,刘思峰,卢奇. 移动平均法中移动步长的确定[J]. 北京工业大学学报,2004,30(3):378-381.
- [35] 方婷. 基于线性回归分析与 BP 神经网络的枣树需水量研究[J]. 南方农机,2025,56(5):44-46,50.
- [36] 方匡南,吴见彬,朱建平,等. 随机森林方法研究综述[J]. 统计与信息论坛,2011,26(3):32-38.
- [37] 朱嘉涵,何治斌,陆昊,等. 基于 NodeJS+VueJS 的船舶机务管理系统的设计与实现[J]. 计算机应用与软件,2024,41(11):39-45,77.
- [38] 黎孟雄,郭鹏飞. 基于 SQLite 的数据库原理自主实验平台设计[J]. 实验技术与管理,2016,33(11):157-160.
- [39] 陈强,王松旺,李言飞,等. 登革热监测系统概念数据模型构建[J]. 医学信息学杂志,2018,39(6):43-46,60.
- [40] 苟英. 白盒测试[J]. 科技资讯,2010(10):14-14.
- [41] 宋书华. 黑盒测试方法[J]. 中国科技纵横,2020(6):84-85

Design of an Intelligent Student Emotion Analysis and Monitoring System Empowered by Multimodal Data and Large Models

Abstract: With the in-depth development of artificial intelligence and deep learning technologies, the application potential of multimodal sentiment analysis in the educational field is increasingly evident. Traditional unimodal emotion recognition methods have limitations in capturing students' complex emotional states, while multimodal analysis significantly improves the accuracy of emotion recognition by integrating facial expressions, speech information, and physiological signals. Under the current background of educational informatization, there is an urgent demand for student mental health monitoring. However, existing methods face challenges such as poor timeliness, strong subjectivity, and difficulties in scaling, which limit their widespread adoption in campus environments.

To address these challenges, this paper proposes and implements a student emotion intelligent analysis and monitoring system based on ESP32 and ESP32S3 hardware platforms,

combined with lightweight multimodal fusion algorithms and large language models. The system aims to utilize low-cost, highly integrated embedded technology to fuse multi-source data including facial, speech, and heart rate information, providing educators, parents, and students with real-time, accurate, and convenient emotion monitoring and support tools.

At the hardware level, a distributed dual-mainboard architecture using ESP32 and ESP32S3 is adopted. The ESP32 mainboard integrates ESP32CAM and heart rate sensors to implement facial expression recognition, physiological data collection, and basic feedback. The ESP32S3 mainboard integrates a digital microphone, audio amplifier, display screen, etc., to achieve intelligent dialogue functions based on Baidu's ERNIE Bot API. At the software level, a Node.js-based server is constructed, SQLite is used for data storage, and multi-role web application interfaces for teachers, students, and parents are developed. On the algorithmic level, the system implements facial emotion recognition based on Deepface, speech emotion analysis using the ERNIE Bot API, designs a dynamic weight decision-level multimodal fusion algorithm, and introduces a data-volume-based multi-model emotion trend prediction method. Additionally, prompt optimization is employed to enhance the performance of large language models in emotional support dialogue tasks.

Finally, the system hardware platform was successfully built and debugged, with comprehensive functional testing and verification conducted on the software system, including white-box testing and black-box testing. The test results demonstrate stable operation of all system modules, compliance with design requirements, and effective integration of multimodal data for student emotion state analysis and monitoring, validating the feasibility and effectiveness of the design.

Key Words: Multimodal sentiment analysis; Student emotion monitoring; ESP32; Large language models; Data fusion