



广东技术师范大学
Guangdong Polytechnic Normal University



计算机科学学院

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

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

答辩人：黄月胜

指导老师：黎嘉文

专业：物联网工程

日期：2025.5.22



目录

01

研究背景及现状

02

研究总体系统设计

03

关键技术与创新点

04

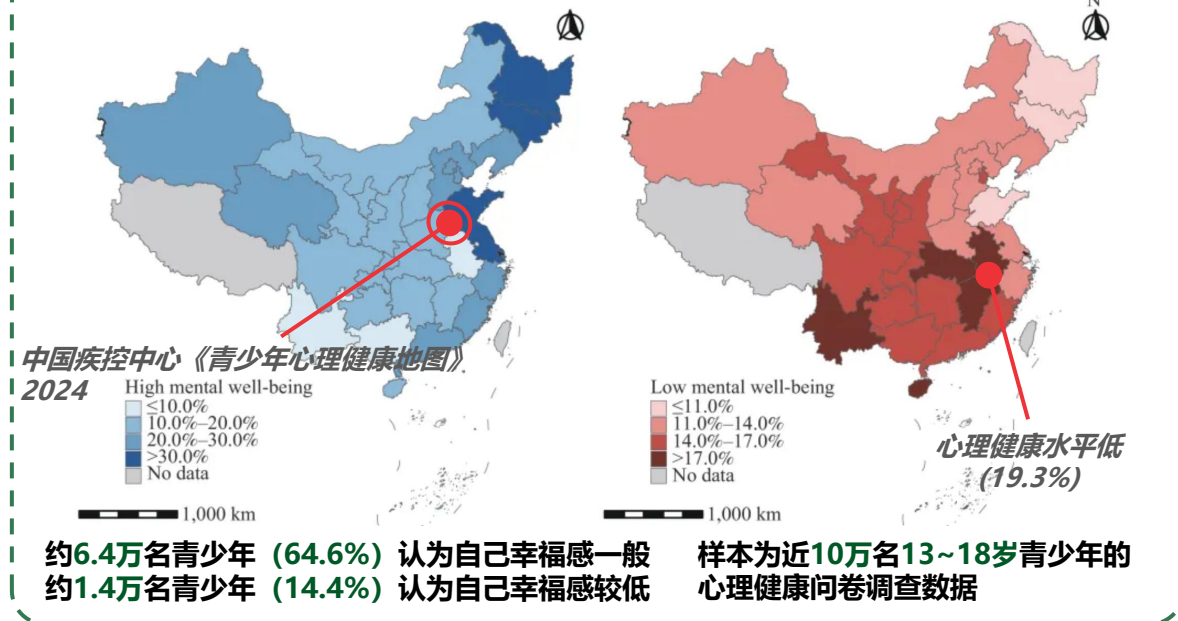
系统测试与评估

05

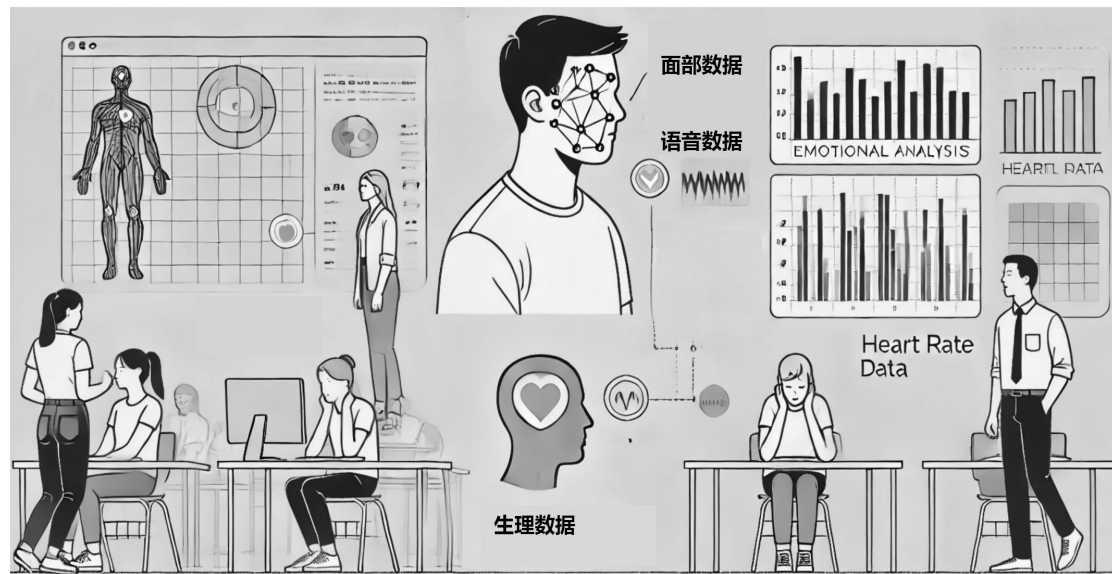
总结与未来展望

- 青少年心理健康形势严峻，教育部明确要求利用AI赋能教育心理
- 传统问卷方式时效性差、主观性强
- 情绪监测能够生成智能情感报告，助力家校协同优化学生情绪管理

青少年心理-亟待关注



多模态视角下的课堂情绪



需求

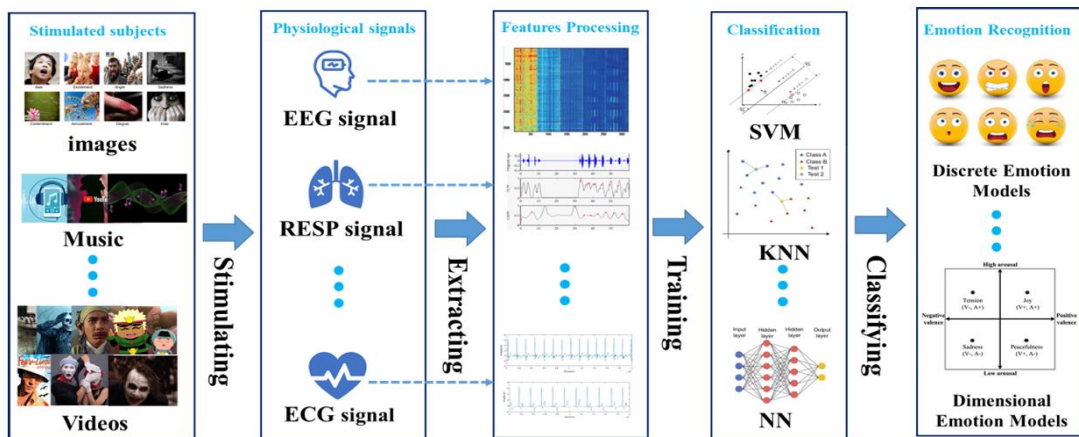
迫切需要构建学生智能情感监测与分析系统，优化课堂情绪管理策略

教育场景的多模态情感分析是数据融合技术与轻量化需求共同驱动的创新方向

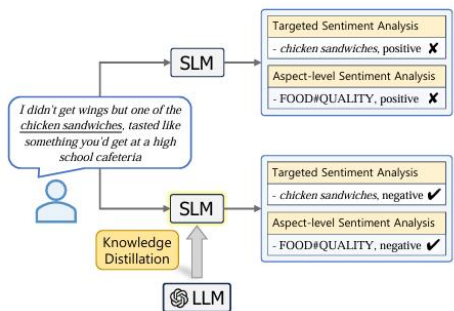
国外研究现状

多模态融合技术成熟

Wang et al, Information Fusion, 2022

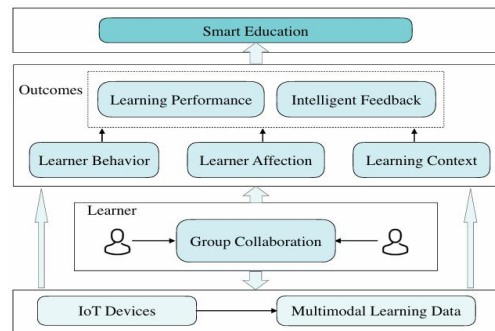


大模型驱动细粒度情感分析



Shen et al, Arxiv, 2025, 美国

物联网轻量化适配教育场景

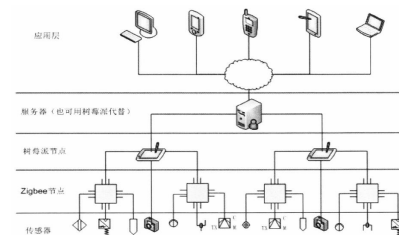


GAN et al, ICDAR, 2022, 日本

国内研究现状

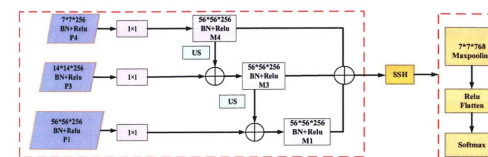
起步晚、但发展迅速

2016年 情感化智能教室 (面部表情)



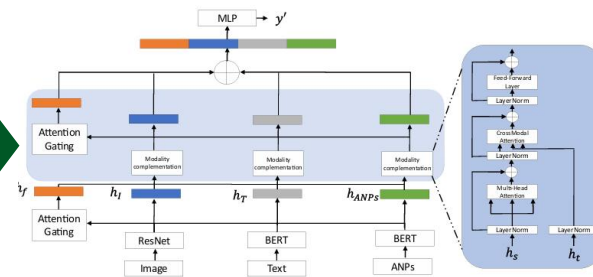
程二丽, 华中师范大学硕士毕业论文, 2016

2018年 智能教育机器人 (面部表情)



张新新等人, 中国特殊教育, 2018

2024年 在线课堂多模态情感分析 (文本+语音+视觉)



詹勇鑫, 江西师范大学硕士毕业论文, 2024

教育场景的多模态情感分析需兼顾轻量化、实时性与低成本

现状与不足

国外研究侧重高性能算法与跨学科应用，国内则聚焦教育场景融合与嵌入式部署探索

国外多模态情感识别技术成熟但受限于硬件成本，国内研究近年来快速发展但仍面临轻量化与实用化挑战

当前系统存在硬件成本高、算法实时性不足、多模态数据融合效率低等问题难以满足基础教育场景的普及需求

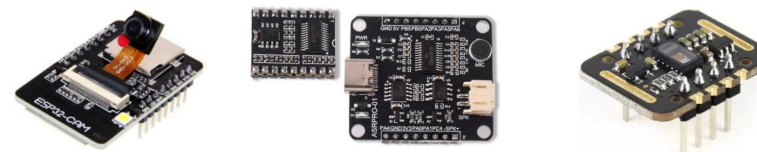
?



研究问题提出

1. 如何通过轻量化多模态融合算法提升嵌入式设备的实时情感分析能力？

2. 如何平衡系统成本与精度，实现教育场景的大规模部署？



3. 如何将情感识别结果有效转化为教学干预与心理健康支持？

班级情绪雷达图



班级情绪健康指数



研究目标

设计轻量化多模态情感分析硬件终端，实现教育场景下面部表情、语音及生理信号的实时同步采集，构建低成本、高能效的情感分析系统，为教师、家长提供可视化心理健康预警与干预支持

研究内容一

提升硬件设备的实时情感分析能力

- ✓ 采用动态权重决策级融合算法优化表情识别与语音情感分析的实时性
- ✓ 在本地完成轻量化计算，减少云端依赖，提升响应速度

研究内容二

面向教育场景的规模化部署方案

- ✓ 选用ESP32/ESP32S3低功耗硬件，结合模块化设计，降低硬件成本
- ✓ 通过异步数据传输，减少网络负载与服务器压力

研究内容三

情感识别结果的可视化与智能反馈

- ✓ 构建多角色Web平台，提供情绪趋势、异常预警等可视化分析功能
- ✓ 集成文心一言API，通过优化Prompt设计实现情感化对话为学生提供即时心理支持

解决问题

1. 如何通过轻量化多模态融合算法提升嵌入式设备的实时情感分析能力？

2. 如何平衡系统成本与精度，实现教育场景的大规模部署？

3. 如何将情感识别结果有效转化为教学干预与心理健康支持？

系统需求分析

总体需求分析

高效实时监测

智能预警机制

功能性需求

多模态数据采集

可视化与干预

非功能性需求

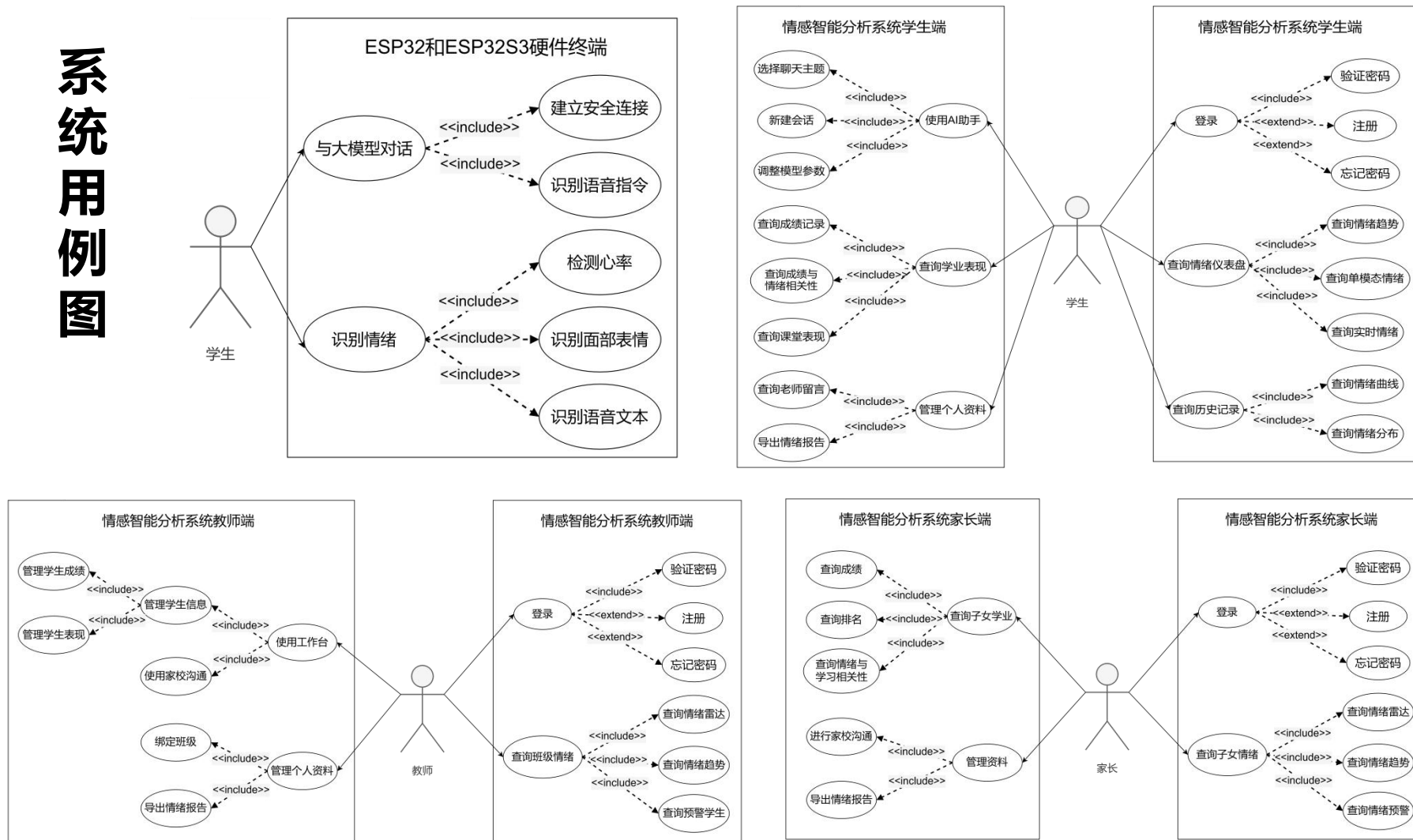
准确性

实时性

安全性

易用性

系统用例图



系统设计原则

人本化与易用性

差异化与角色适应

多模
态融
合与
精准
分析

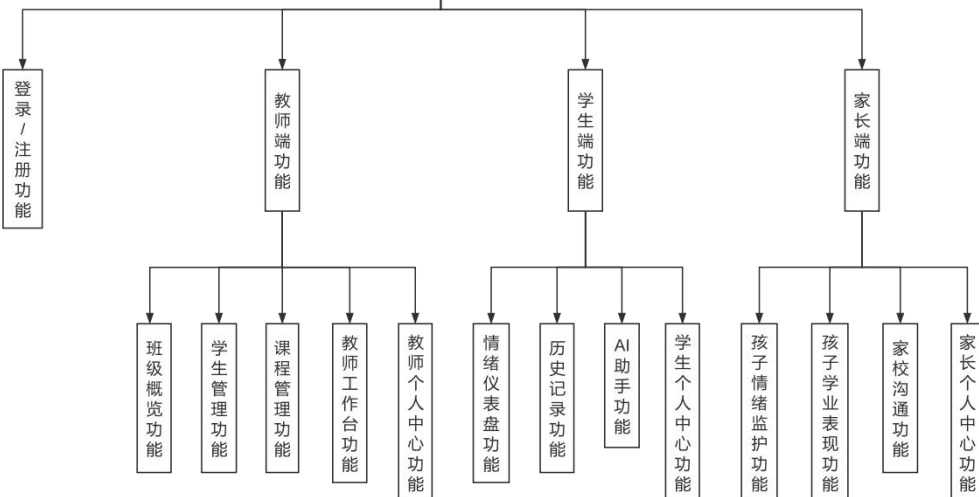
实时性与轻量化

安全性与隐私保护

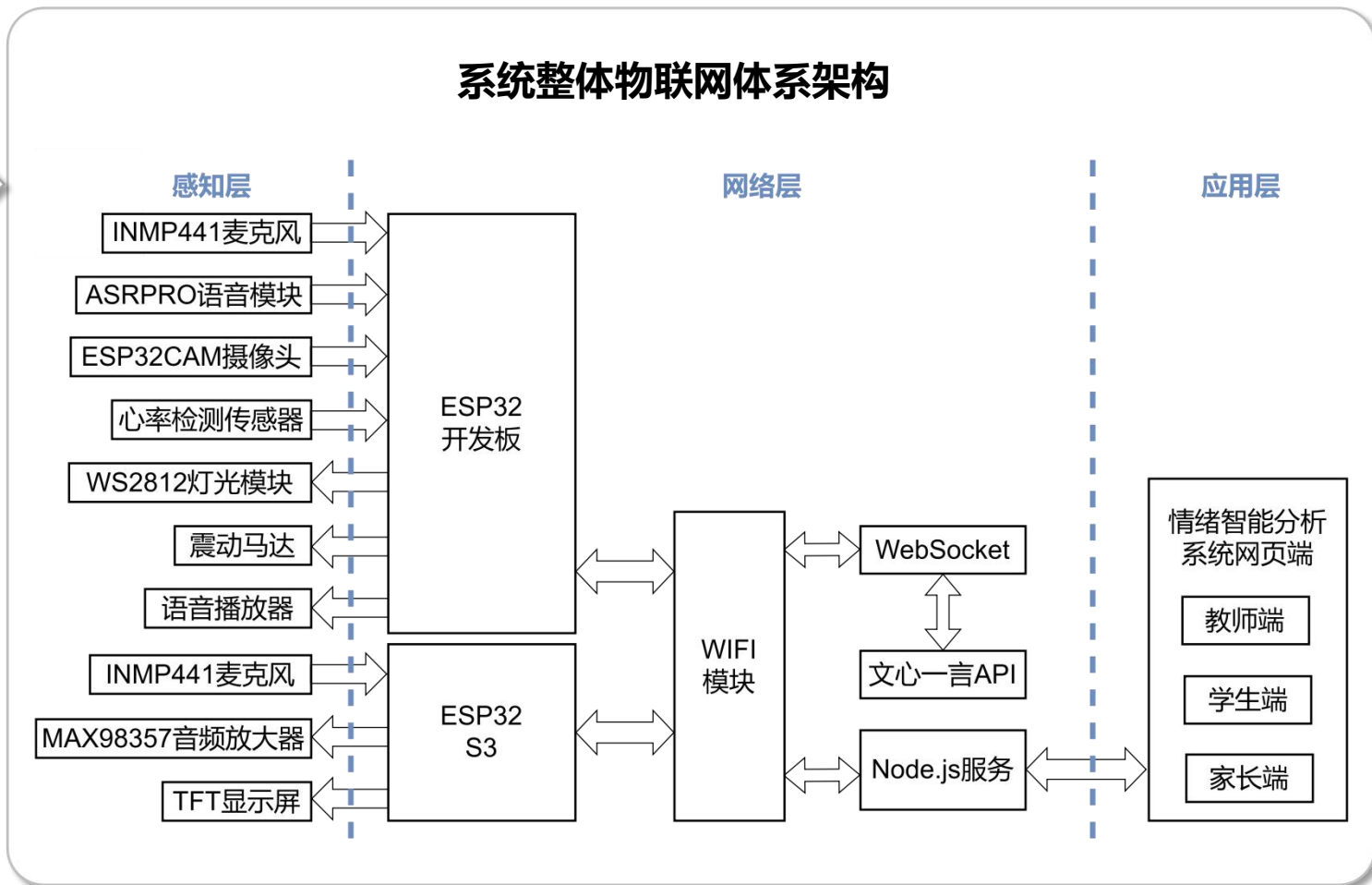
可扩展性与维护性

双向互动与反馈

学生情感智能分析系统



系统整体物联网体系架构

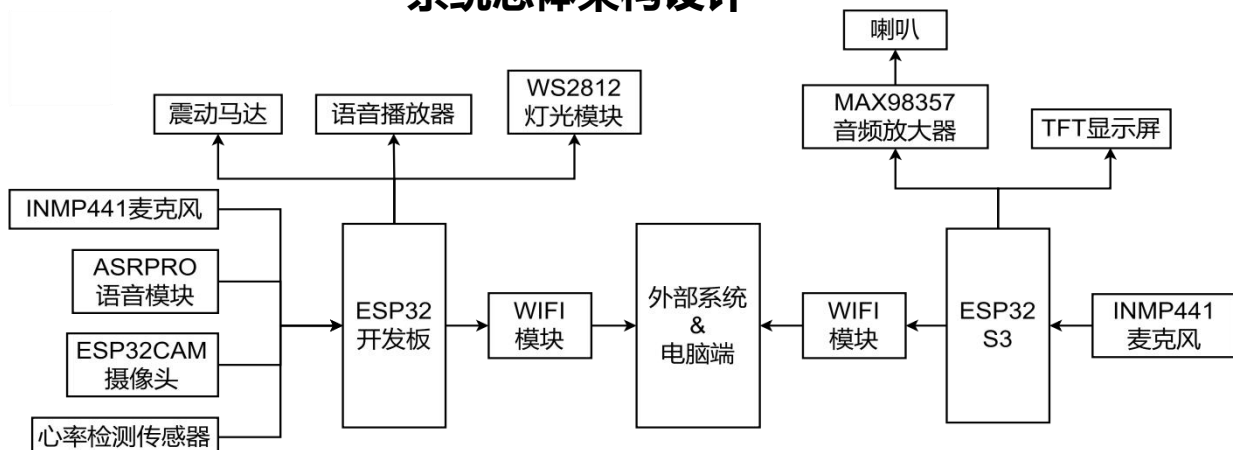


**系统中各功能模块通过数据接口实现互联
在保证系统鲁棒性的同时降低了新增功能的开发复杂度**

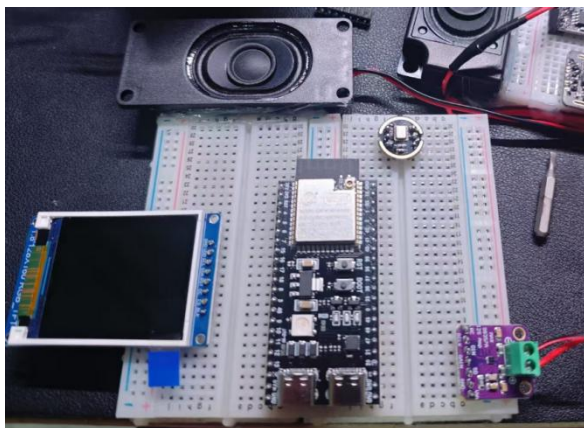
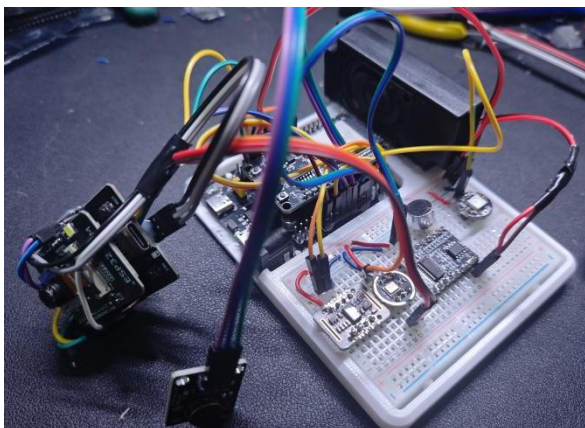
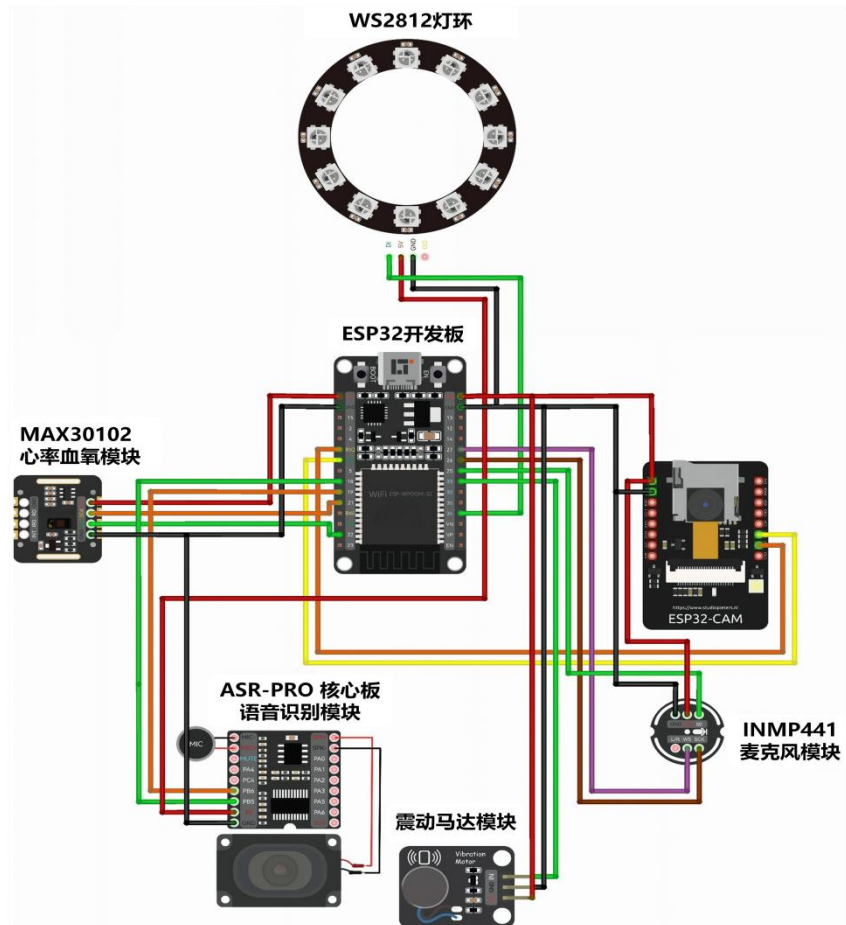
■ 系统硬件设计

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

系统总体架构设计



主板一串口连接图



系统核心算法一

多模态决策级融合算法, 精准高效、边缘端实时计算

七种基本情绪

基于美国心理学家Paul Ekman的情绪理论划分



视觉情绪分类器

Algorithm 1 DeepFace 情绪识别算法

- 1: 输入: 图像 I
- 2: 输出: 情绪概率 P
- 3: 初始化 DeepFace 模型, 加载预训练权重
- 4: 使用 OpenCV 检测人脸, 提取人脸区域 F
- 5: 对齐人脸 F , 调整关键点位置
- 6: 将对齐后的人脸 F' 输入模型
- 7: **for** 每层卷积层 **do**
- 8: 执行卷积操作, 激活函数 ReLU
- 9:

$$f(X) = \max(0, W * X + b)$$

- 10: **end for**
- 11: **for** 每层池化层 **do**
- 12: 执行最大池化操作
- 13: **end for**
- 14: 展平特征, 输入全连接层
- 15: 使用 Softmax 输出情绪概率 P

$$P_i = \frac{e^{z_i}}{\sum_{j=1}^7 e^{z_j}}$$

- 17: 使用交叉熵损失函数优化

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

- 19: 选择优化器 (Adam 或 SGD), 学习率衰减
- 20: 返回情绪概率 P

语音文本情绪分类器

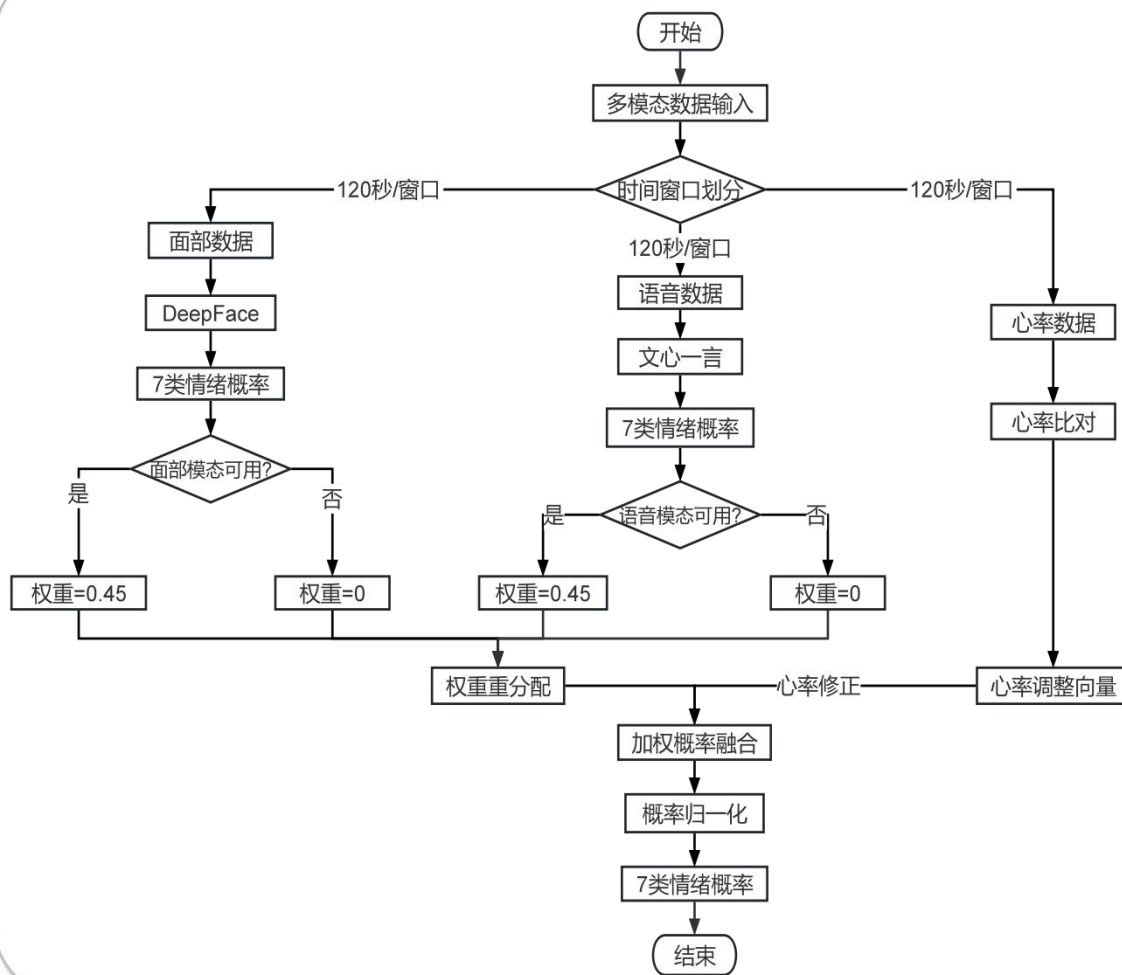
提示词内容

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)。
情感强度使用 0 到 1 之间的数值表示。
只返回 JSON 格式的结果, 不要有其他解释。格式如下:

```
{
  "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"
  }
}
```

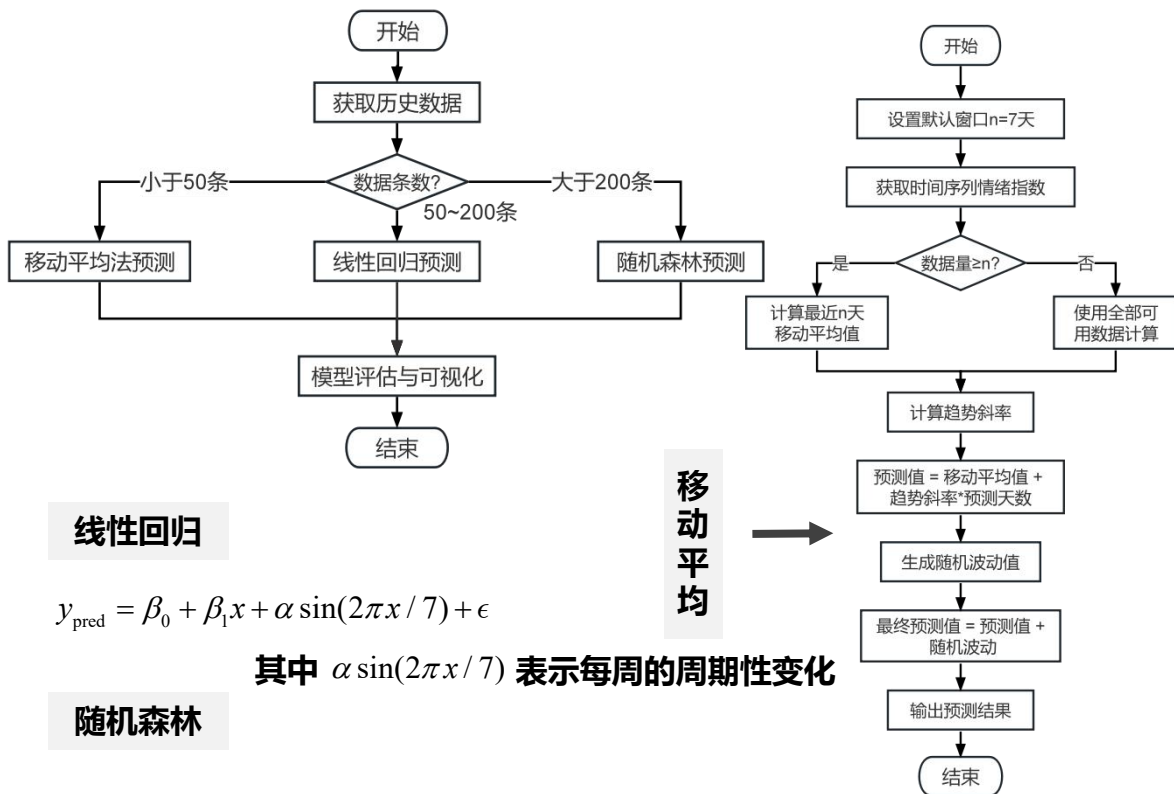
要分析的文本是: "\${text}"

动态权重决策级融合算法



■ 系统核心算法二

多模型融合的情绪指数预测算法



线性回归

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

其中 $\alpha \sin(2\pi x / 7)$ 表示每周的周期性变化

随机森林

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

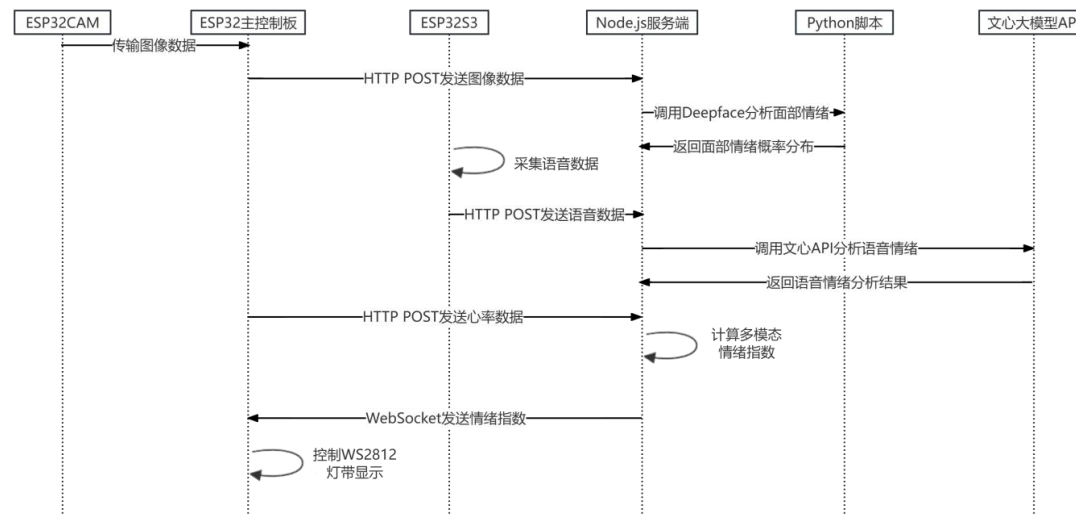
通过智能化的数据量感知门控设计，使模型在不同数据规模下都能保持优异的预测表现

移动平均

■ 系统核心算法三

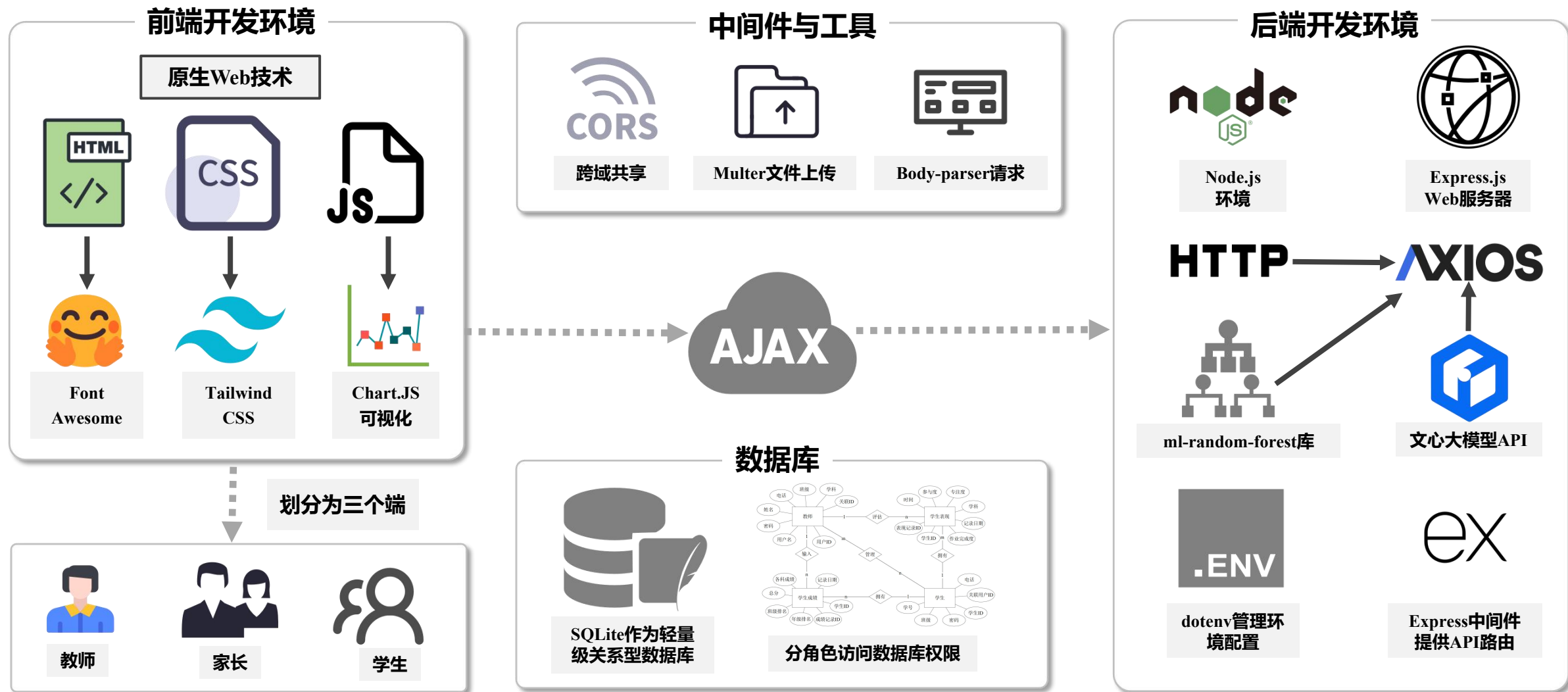
硬件与服务端通信算法

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



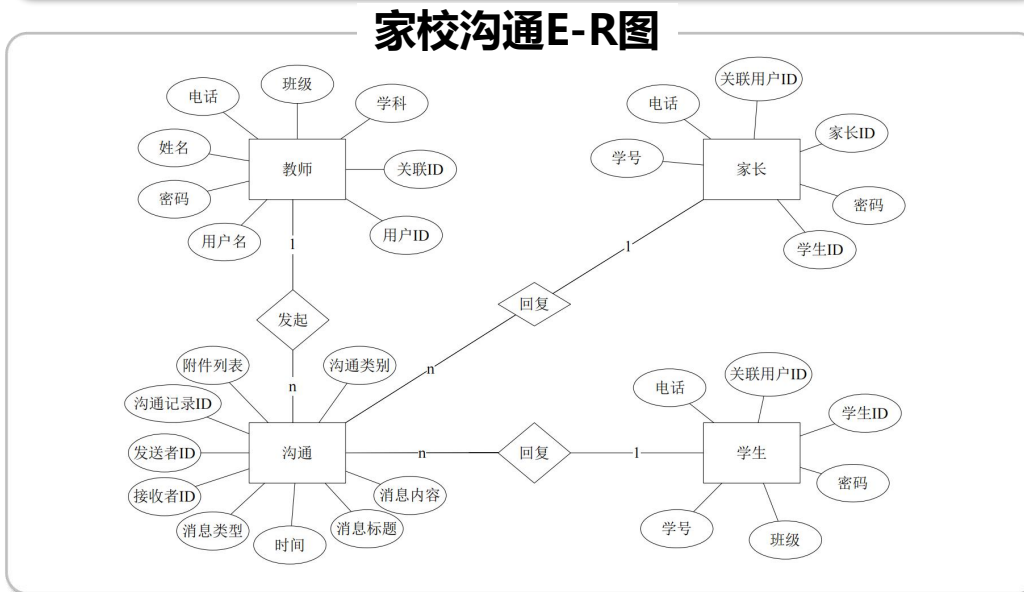
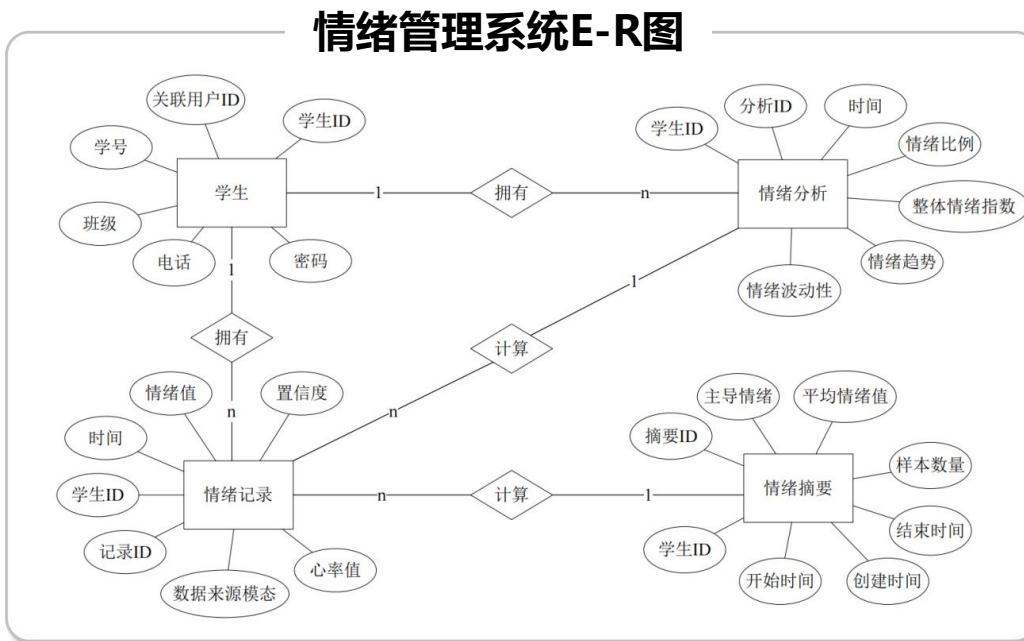
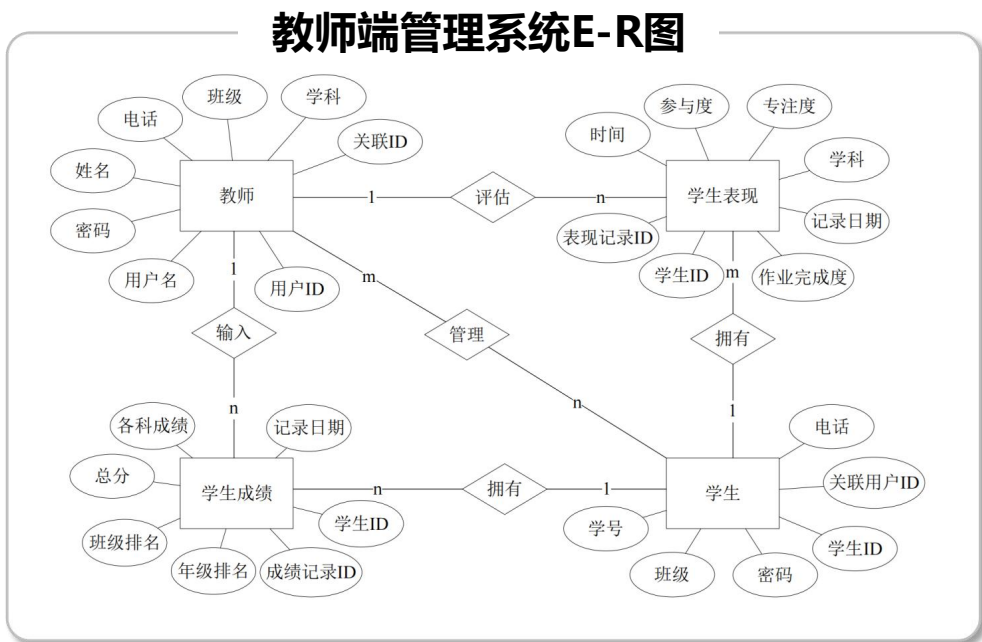
使用child_process模块调用Python脚本进行面部情绪分析，利用axios库与文心大模型API进行通信，以及采用WebSocket技术实现与ESP32的双向实时通信。当计算得到情绪指数后，服务端通过WebSocket将结果实时传回ESP32，用于控制WS2812灯带的颜色显示

系统软件总体设计



系统数据库设计

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



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

系统数据库设计



数据库表设计

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
数据来源模式	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 5.7 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 5.2 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 5.5 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 5.8 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 5.3 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 5.6 情绪分析表

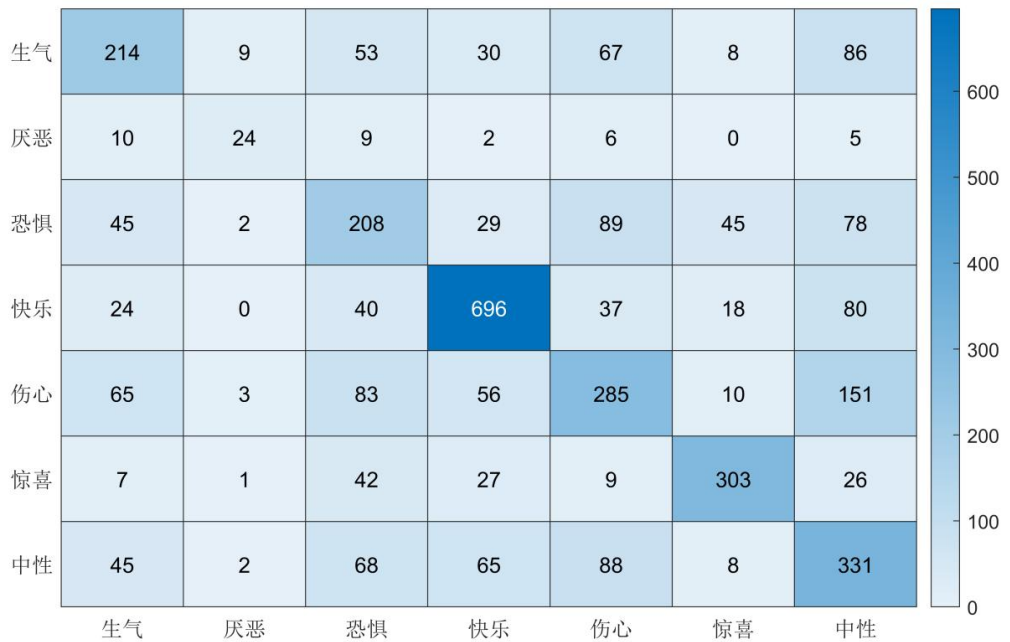
字段名	数据类型	标识	约束
分析 ID	INT(11)	分析唯一标识	主键, 自增
学生 ID	INT(11)	关联学生表 ID	非空, 外键
整体情绪指数	FLOAT(10)	综合情绪指标	非空
情绪波动性	FLOAT(10)	情绪波动程度	非空
情绪趋势	VARCHAR(50)	情绪变化趋势	非空
情绪比例	JSON	不同情绪占比	非空
时间	DATETIME	分析创建时间	非空

表 5.9 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
附件列表	JSON	附件链接	
时间	DATETIME	消息发送时间	非空

■ 由于缺乏满足同时具有三种模态的数据集，因此首先在单模态数据集上验证各模态七分类性能

Deepface在FER2013上的混淆矩阵热力图



Baseline Accuracy	34%
Our work(Deepface) Accuracy	57%

多模型文本分类对比 (识别标签/响应时间)

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
"太感谢了，帮我浪费了 3 小时"	快乐	愤怒	愤怒	愤怒	愤怒

Approach	Accuracy	量级
BERT	58%	中等
Our work(LLM)	86%	轻量

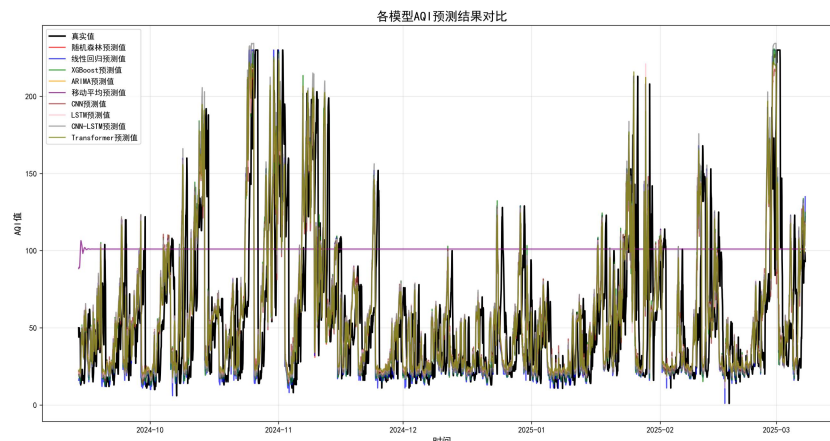
在公开多模态数据集MOSEI (文本+图像) 中验证多模态融合算法的优越性

模型	文本准确率	图像准确率	融合准确率	计算资源需求
GloVe-LSTM(文本)	70.93%	-	-	低
DeepFace(图像)	-	63.75%	-	低
特征级融合 (GloVe+Deep)	-	-	72.34%	中
决策级融合 (本文)	-	-	71.79%	低
ConCluGen (Baseline)	-	-	66.48%	中
MARLIN(特征级, SOTA)	-	-	80.63	高

部分结果数据来源——Paper with Code

- 由于MOSEI数据集的数据集分类粒度与本文不一致, 因此评价指标为三分类结果 (积极、中性、消极)
- 由于计算资源限制, 只对比了部分简单的融合模型, 特征级融合结果来源于已有文献
- 由于缺乏真实环境的数据, 心率数据的辅助决策需要后续实验采集数据后再进行补充

在Kaggle公开时序预测数据集上验证了关键算法二的优越性



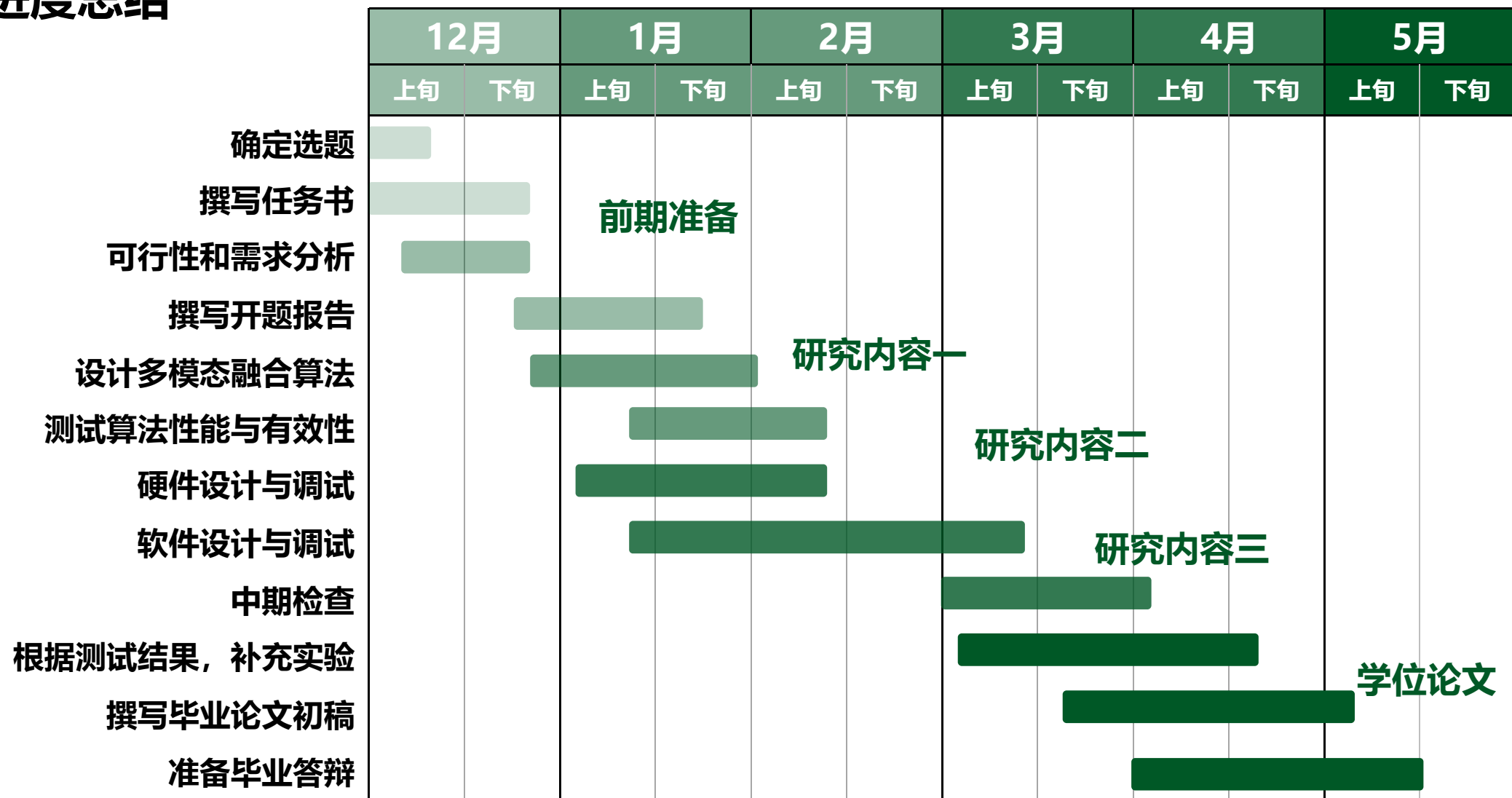
模型	RMSE	MAE	R ²	MAPE	运行时间 (秒)
LSTM	9.02	5.76	0.95	16.17	33.93
CNN	9.43	5.86	0.95	15.14	35.91
CNN+LSTM	10.14	6.74	0.94	19.65	43.18
线性回归	8.84	5.16	0.96	12.84	0.01
决策树	9.50	5.36	0.95	12.61	0.07
随机森林	8.72	4.99	0.96	12.25	0.10
XGBoost	8.71	4.99	0.96	12.22	0.23
SVM	21.34	18.03	0.74	67.97	0.51
梯度提升树	8.72	5.18	0.96	13.03	1.98
LightGBM	8.81	5.04	0.96	12.40	0.20
CatBoost	8.97	5.32	0.95	13.44	0.89
Transformer	43.16	34.16	-0.06	105.41	130.54

相关成果 (多模态与时序预测)

- [1] Computers, Materials & Continua (SCI, IF=2) Sentiment Analysis Using E-Commerce Review Keyword-Generated Image with a Hybrid Machine Learning-Based Model (导师一作, 被引: 2)
Code: <https://github.com/Yorkson-huang/Text-to-image-SA>
- [2] 黄月胜 (一作) Engineering Letters(ESCI), An Improved Hybrid CNN-LSTM-Attention Model with Kepler Optimization Algorithm for Wind Speed Prediction.(被引: 3) Stars 21
Code: <https://github.com/Yorkson-huang/KOA-CNN-LSTM-Attention> (Github 21 stars)
- [3] 黄月胜 (一作) EIECT'2023 (IEEE, Poster), An Ensemble Learning Approach for Wind Power Forecasting.
- [4] 基于深度学习的多模态数据融合细粒度情感分析方法研究, 广东省科技创新资金自然科学基金类重点资助项目, 4.5万元, 2024-2026 (本人主持)
- [5] 神经侦探——用于神经退行性疾病诊断的可解释多模态对比学习框架, 大创, 2024-2025 (本人主持, 结项)
- [6] Kaggle Silver Medal

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

■ 进度总结



■ 不足与未来展望

硬件设计较为分散

➤ 硬件集成 ➤

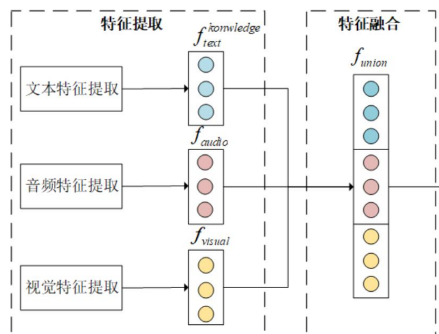
- ✓ 借鉴ESP-SparkBot设计，硬件集成桌宠形态，适配教育场景空间需求



融合算法较简单

➤ 算法迭代 ➤

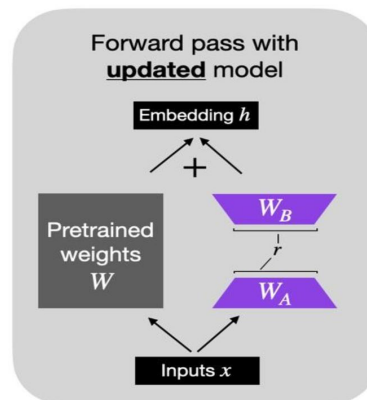
- ✓ 未来探索特征级融合方法，如基于注意力机制的多模态深度学习模型，提升情感识别的准确性



大模型简单调用API

➤ 大模型优化 ➤

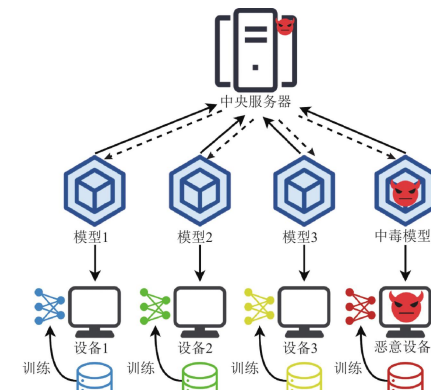
- ✓ 引入知识图谱和引入大模型微调技术，构建更加智能化的情感干预本地大模型



隐私保护不够牢固

➤ 隐私保护 ➤

- ✓ 进一步强化数据隐私保护措施。探索联邦学习的应用，减少敏感数据的传输与存储





廣東技術師範大學
Guangdong Polytechnic Normal University

敬請各位老師批評指正

謝謝!