

# EASRL-MH: 基于强化学习的智能精神健康诊断系统

## EASRL-MH: Enhanced Adaptive Symptom Reasoning Learning for Mental Health

刘沛卓  
西北工业大学附属中学, 西安 710072  
pzliu@stanford.edu\*

姜凯然  
西北工业大学附属中学, 西安 710072  
r18061178166@163.com

### 1. 研究背景 (Background)

目前, 全球有超过十亿人受精神健康问题困扰, 传统访谈需要大约 30-40 个症状线下询问, 并根据医生经验给出诊断。这导致了诊断效率低, 诊断成本高, 诊断正确率不稳定等多种问题。

Share of the population with depression, 2016

Prevalence of depressive disorders in a given population. This is measured as the age-standardized prevalence, which assumes a constant age structure to compare between countries and through time. Figures attempt to provide a true estimate (going beyond reported diagnosis) of depression prevalence based on medical, epidemiological data, surveys and meta-regression modelling.

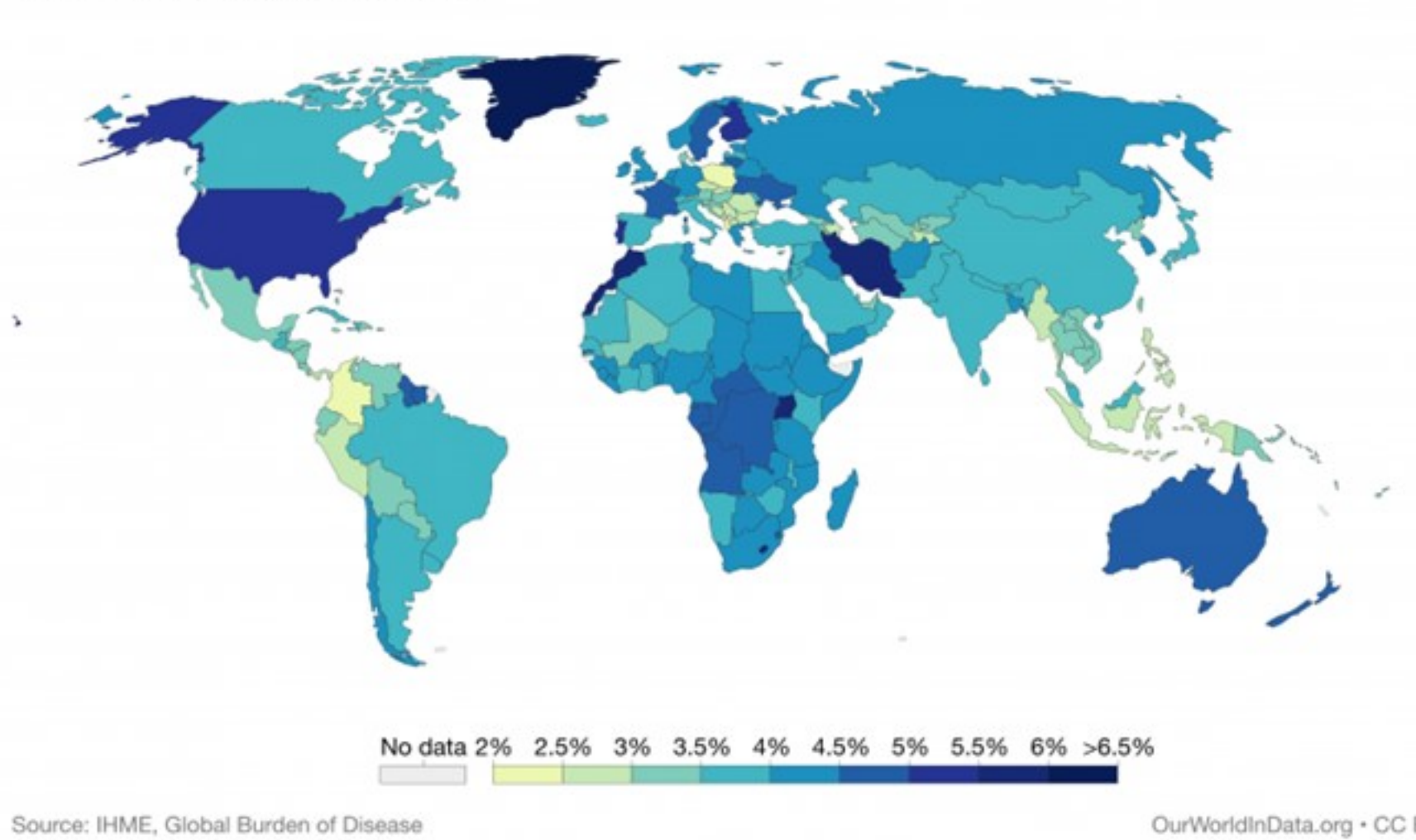


图: 2016年  
全球抑郁人  
口百分比热  
力图

### 2. 研究目标

**精准:** 在遵循 DSM-IV 诊断逻辑的前提下将准确率维持在 85 % 以上  
(平均人类医生诊断准确率);

**高效:** 平均提问控制在 10 题内;

**友好:** 流程模块化, 医生可接入, 设计简约友好的用户界面;

**可解释性:** 心理健康作为高度敏感领域可解释性是模型评估中的重要指标。

### 3. 数据选择(Data Selection)

我们选择了NCS-R(National Comorbidity Survey-Replication)作为研究数据集。NCS-R数据集中包含了≈ 9,282 名美国 18 岁以上成人, 并采用面对面访谈形式进行。覆盖抑郁症、双相障碍、焦虑障碍、恐惧症、创伤后应激障碍 (PTSD)、物质使用障碍、强迫症、饮食障碍等多项精神疾病。

NCS-R使用 DSM-IV 标准通过CIDI(Composite International Diagnostic Interview) 结构式工具进行评估, 该数据集已经匿名化处理, 保证了伦理和隐私合规。

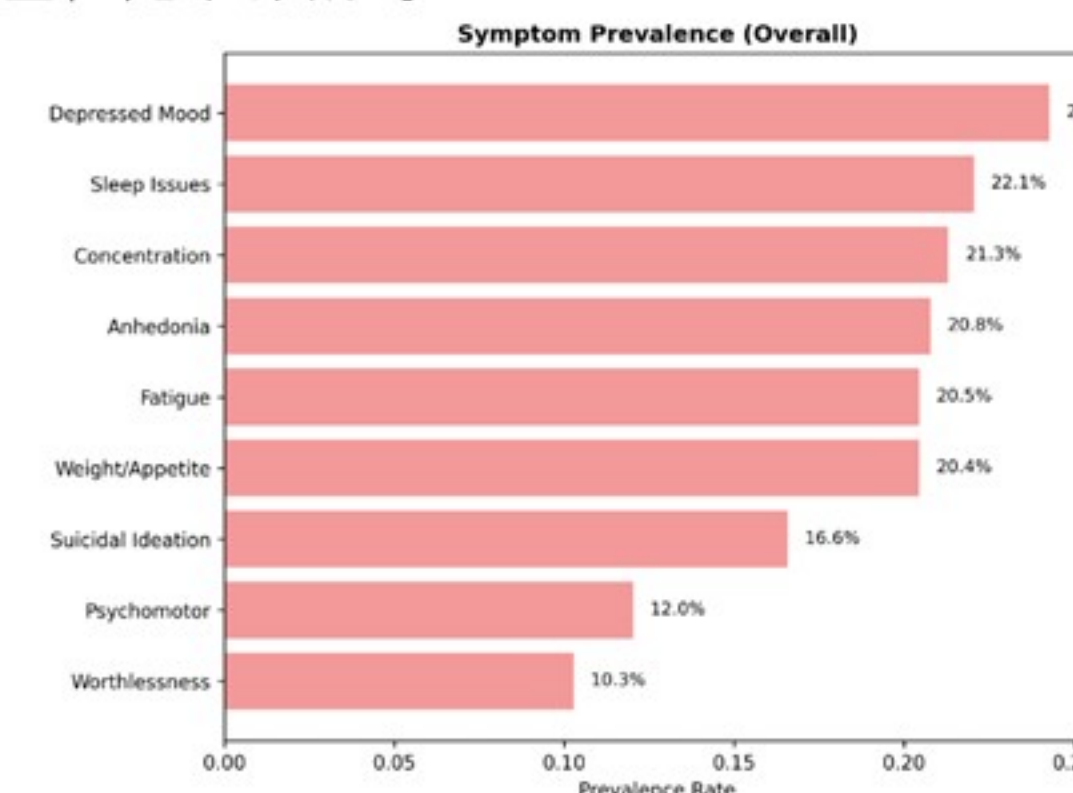


图: NCS-R数据集中MDE相关症状出现比例

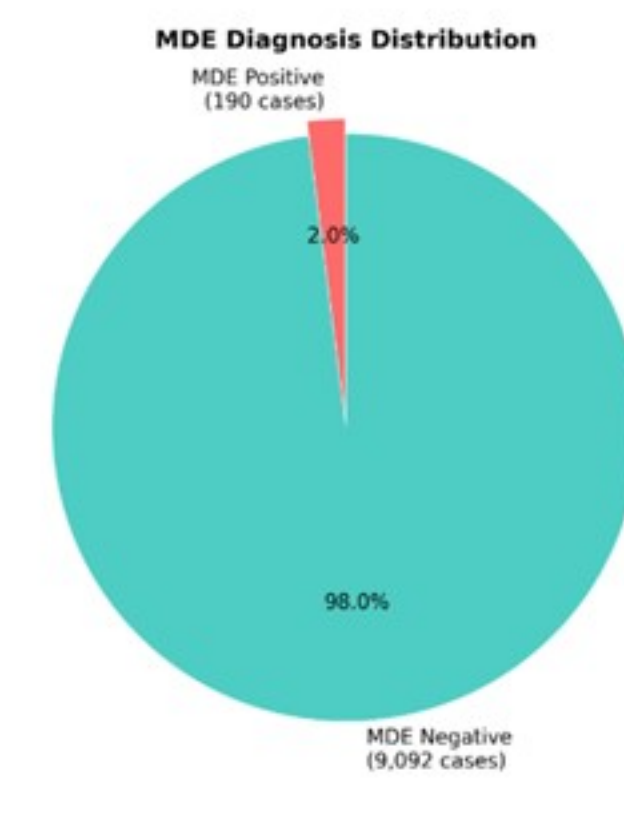


图: NCS-R数据集中MDE阳性比

### 4. 方法概述 (System Overview)

本方法采用两阶段学习架构, 包括:

1. 症状选择策略学习模块 (DQN-based Agent): 结合 DSM-IV 症状图谱与不确定性驱动的询问策略, 动态选择待询问症状;
2. 症状推理诊断模块 (GNN-based Diagnosis Model): 基于症状-诊断二分图构建多头注意力图神经网络, 实现精准分类。

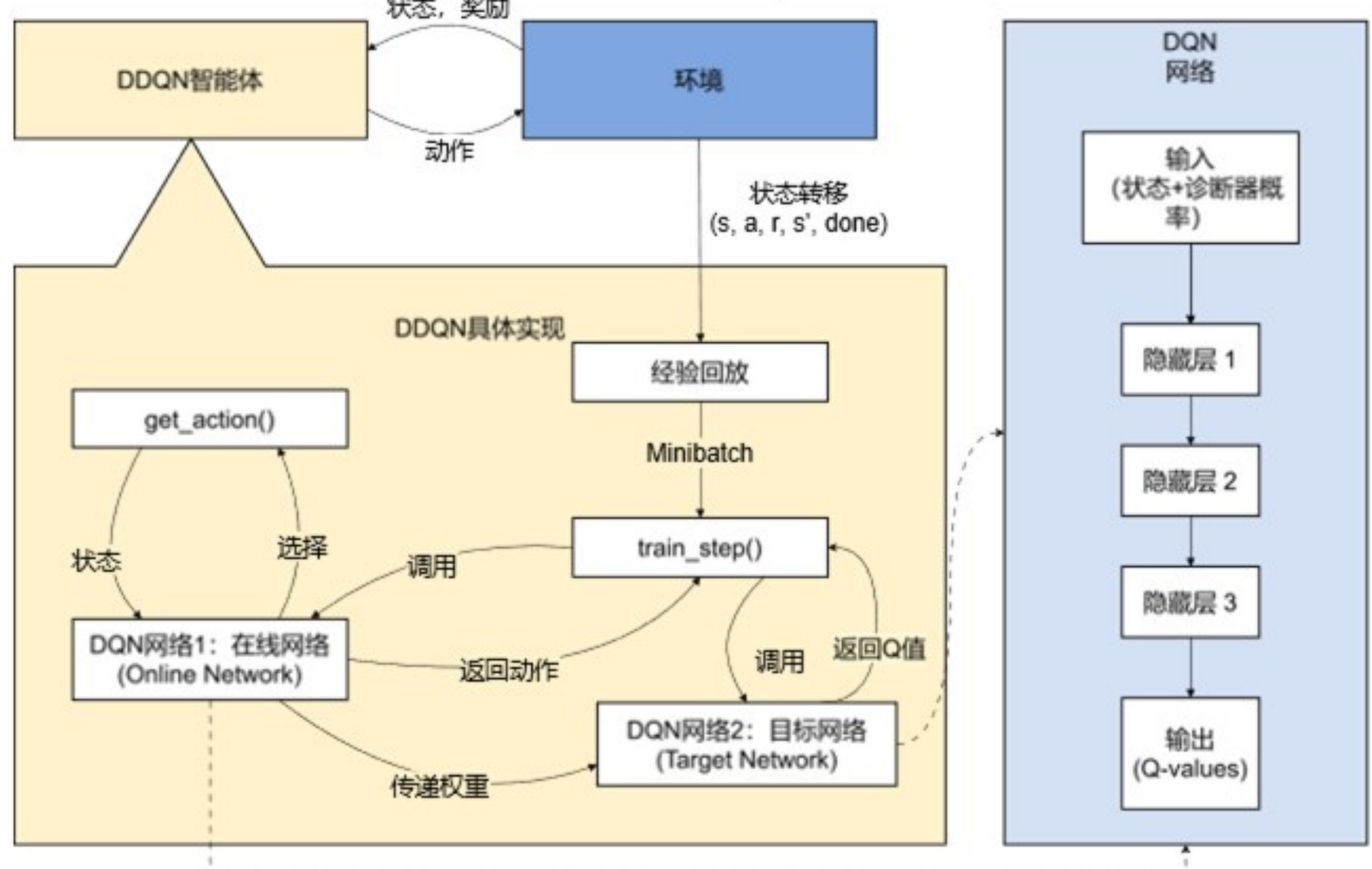


图: 简化模型架构图

### 5. 技术细节 (Implementation Details)

- 1) 症状询问策略优化基于 DQN 结构, 结合  $\epsilon$ -greedy 策略, 引入 70% 结构性知识引导 + 30% 不确定性驱动探索机制, 优化询问效率。
- 2) 图神经网络诊断模型设计基于多头注意力机制的图神经网络 (GAT), 充分利用症状与诊断之间的关联性; 其中核心症状边权被置为 1.0, 以保持 DSM-IV 的医学解释性。
- 3) 联合训练策略采用交替训练方法, 实现问诊策略与诊断模型的相互优化, 避免梯度冲突或模式崩塌现象。

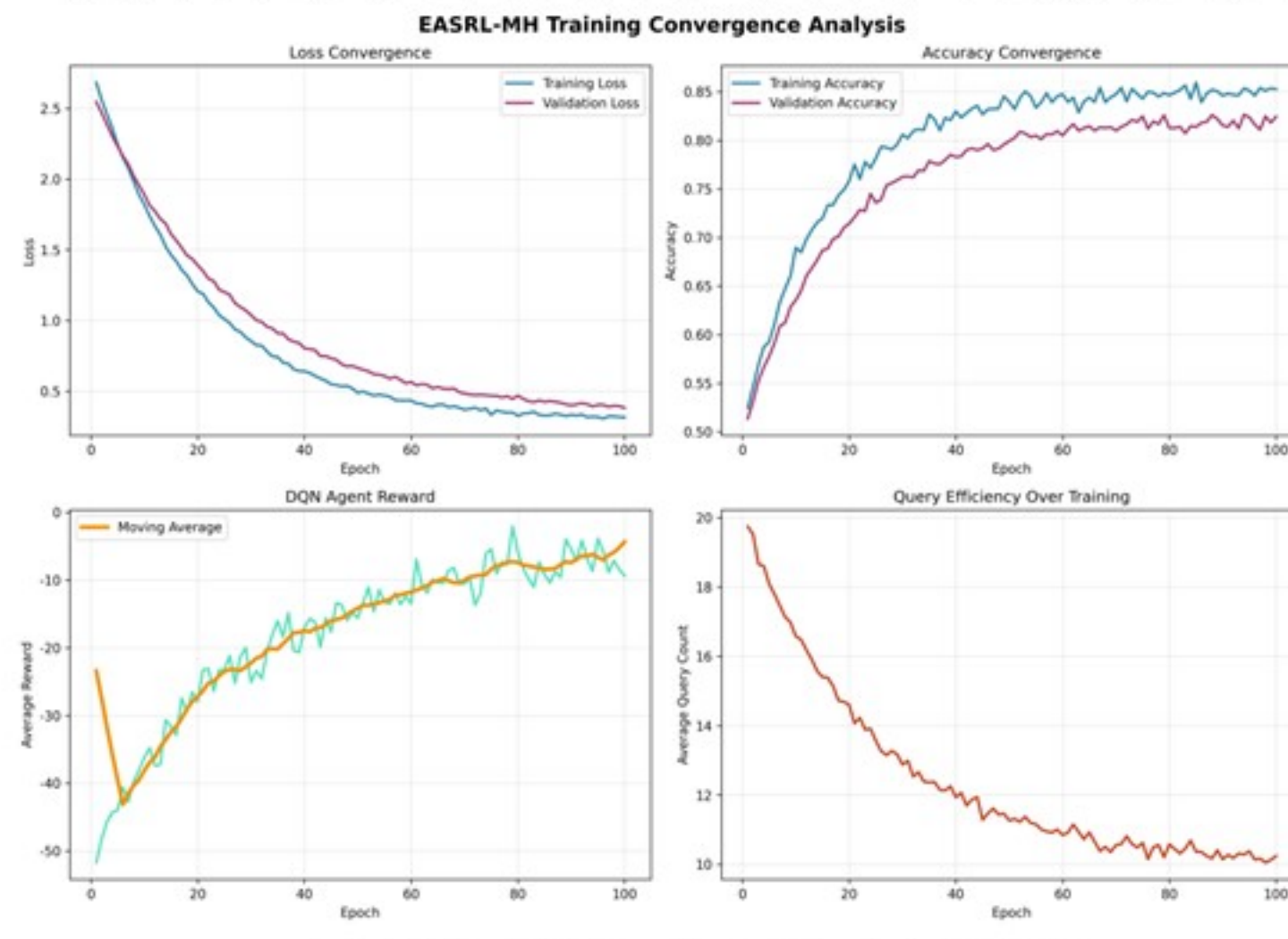


图: 训练过程追踪, 由图可见, 收敛效果较好

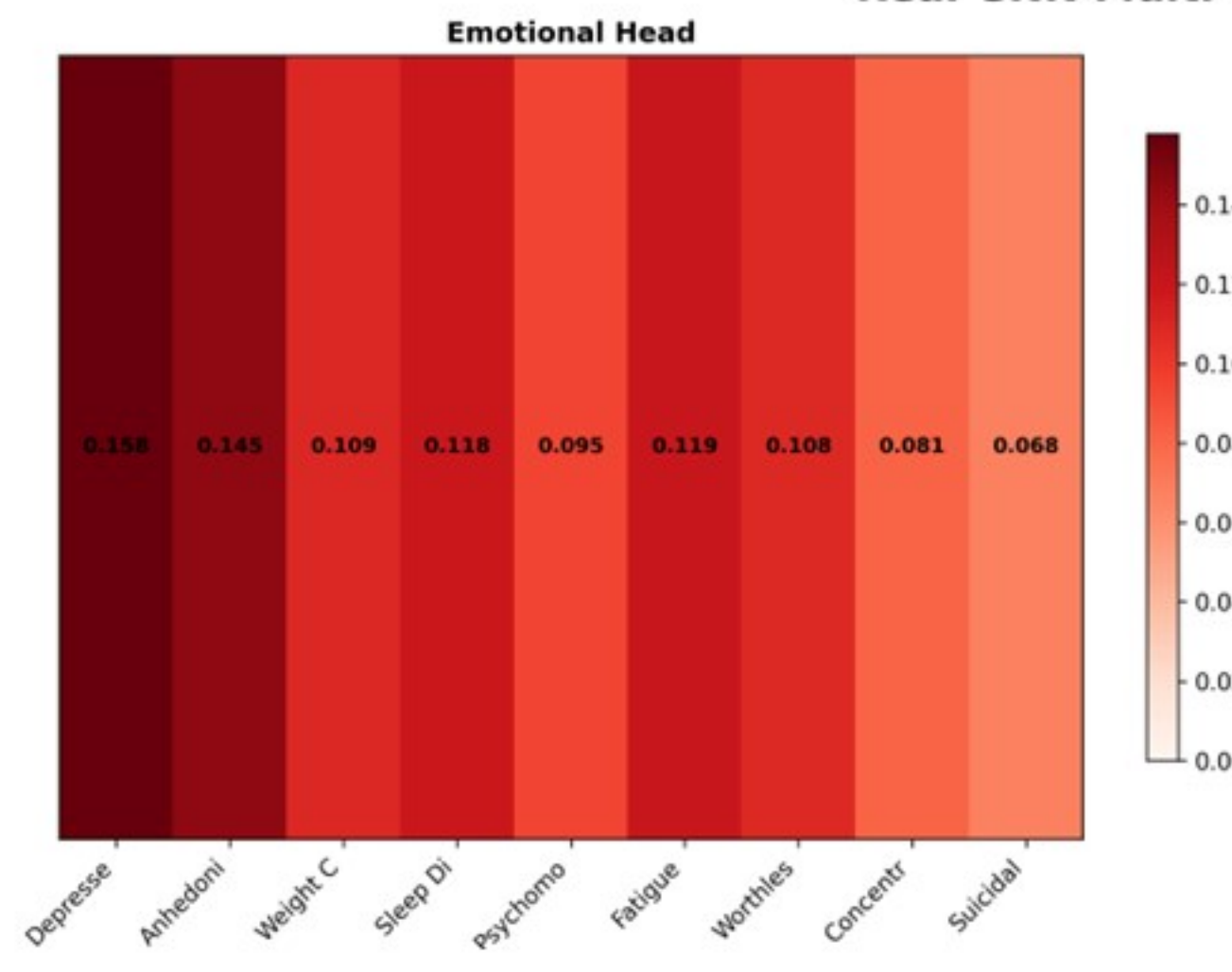
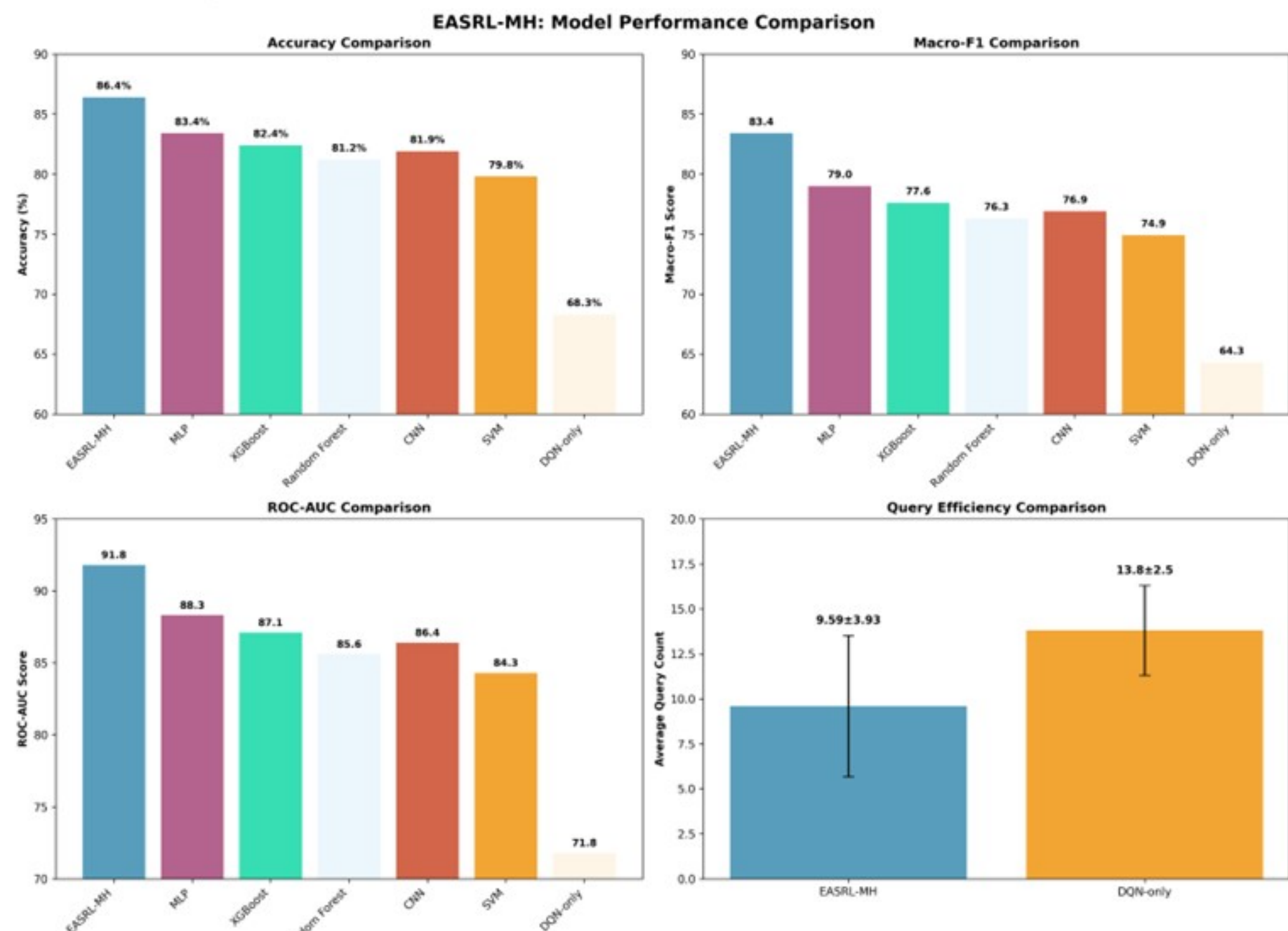


图: 注意力头可视化

### 6. 实验结果(Baseline Comparison)

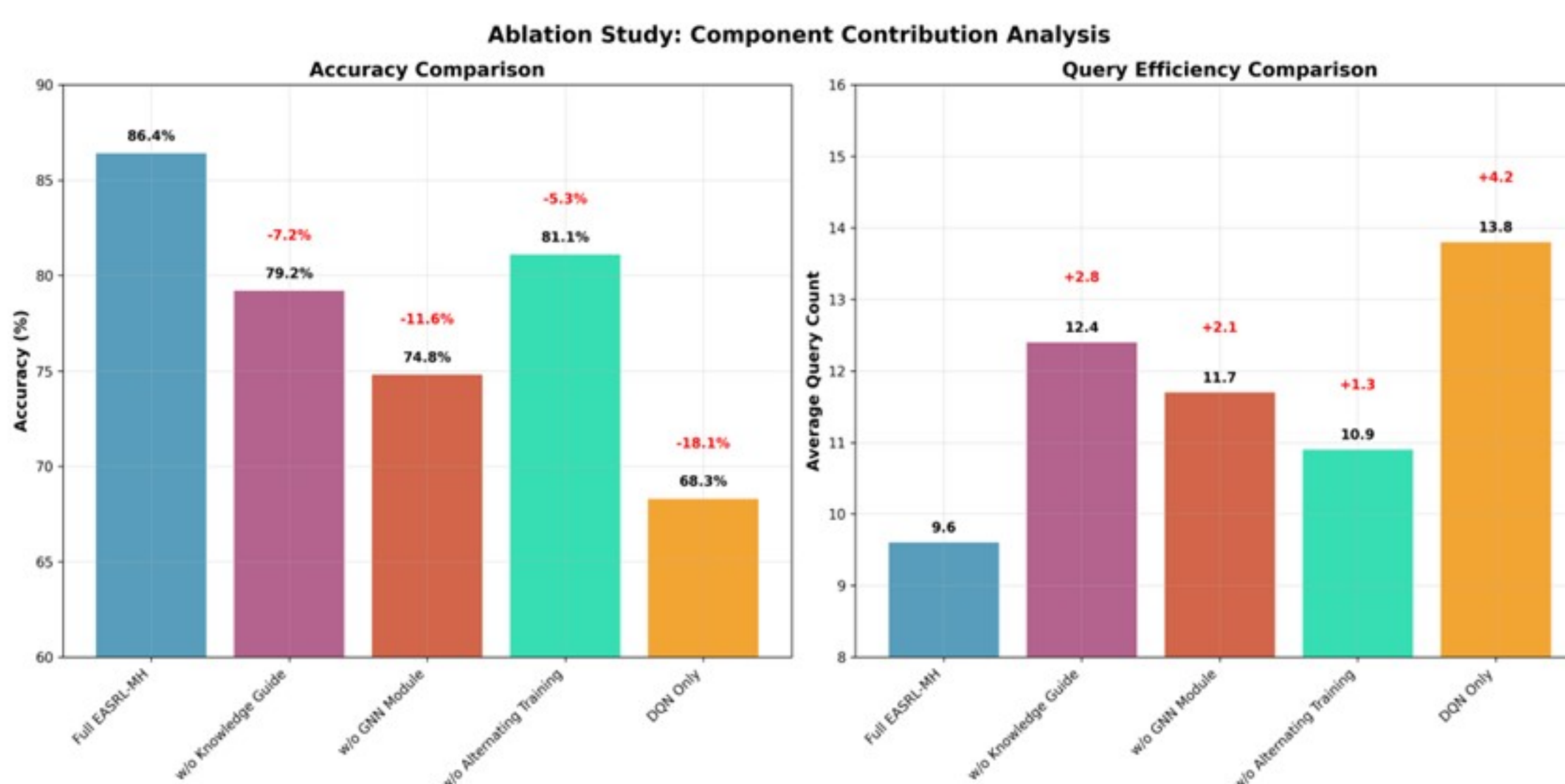
EASRL-MH 显著优于基线模型, 在保证高诊断准确率的同时显著减少了平均询问数量。



### 7. 消融研究 (Ablation Study)

为验证各模块作用, 我们进行了逐项消融实验:

去除知识引导 (Knowledge-guided Querying) 后, 准确率下降 7.2%;  
去除 GNN 模块, 准确率下降至 74.8%;  
若仅使用 DQN, 准确率下降至 68.3%。



### 8. 应用前景

1. 辅助临床决策: 自动化预诊断模型可节省医生 30-50% 访谈时间;
2. 公共健康场景: 适用于校园、企业、社区等大规模初筛;
3. 远程诊断服务: 为偏远地区提供智能辅助问诊支持。

### 9. 现存不足

1. DSM-IV 症状集合的时代滞后;
2. 临床真实世界落地验证不足;
3. 支持诊断的精神健康障碍仍较少。

### 10. 未来工作(Future Work)

1. 标准升级: 支持 DSM-5 及 ICD-11 编码体系;
2. 联邦学习集成: 保障隐私前提下实现跨机构模型优化;
3. 疾病类型扩展: 覆盖更多精神疾病种类 (如焦虑、PTSD、双相障碍等)

### 11. 团队成员(About Us)

刘沛卓 负责项目的代码编写、论文撰写、实验和测试、海报制作以及对展示PPT的大纲制作等;  
姜凯然 负责PPT的实际制作、论文排版美化等工作, 并完成了项目视频的制作;  
团队成员均为西北工业大学附属中学高2026届B7班学生。