

命题项目组（本科生）

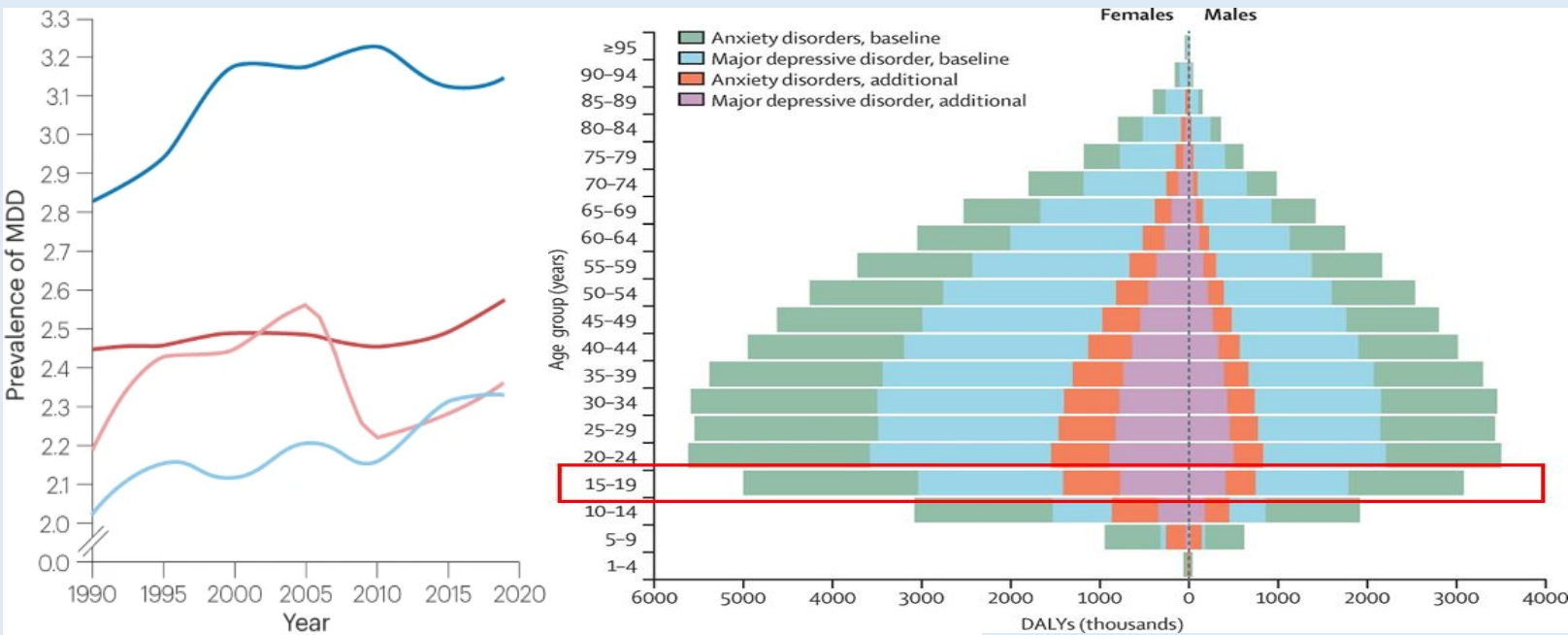
作品ID：1654

赛道名称：心理疾病诊疗与康复

作品名称：基于机器学习的青少年早期抑郁症检测

研究背景

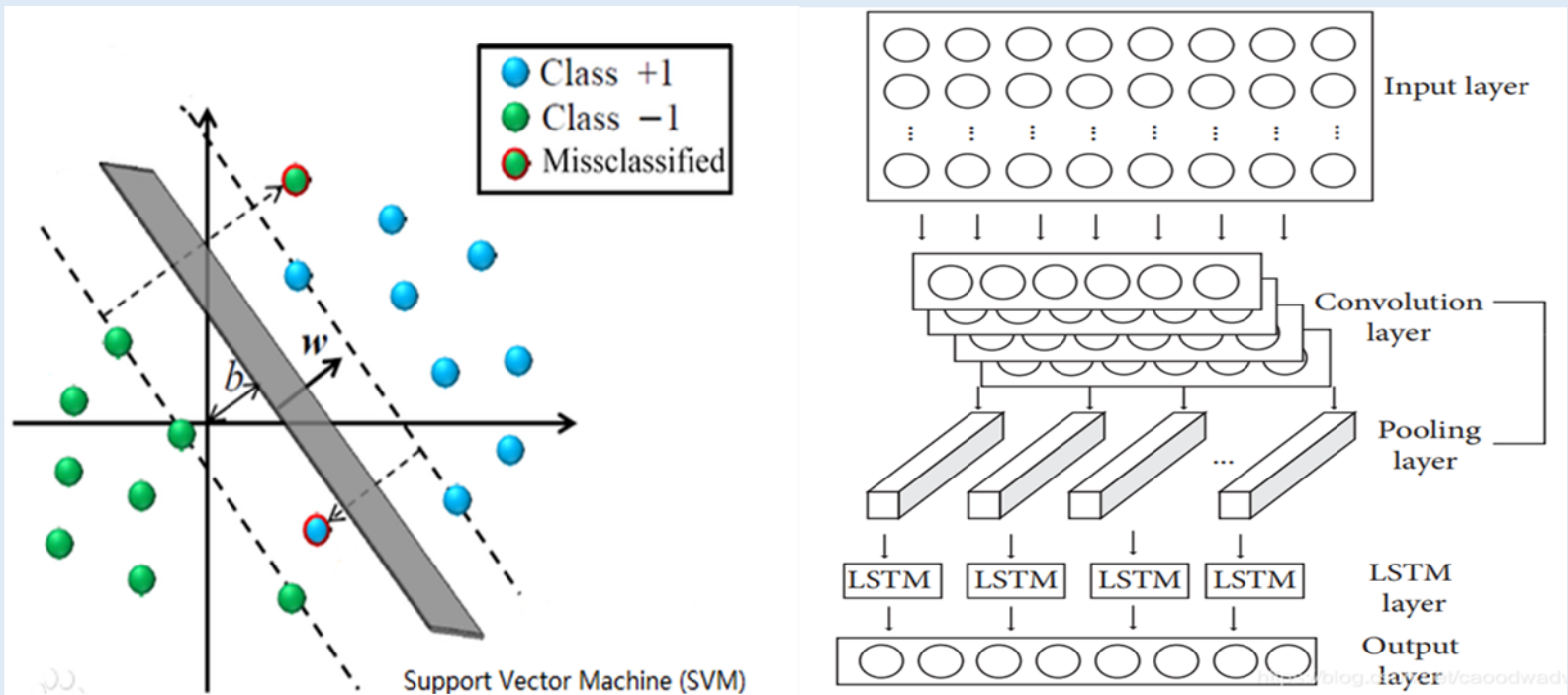
抑郁症是一种发病机理尚不明确的精神心理障碍疾病，其发病率高、治疗周期长、复发率高，但疾病识别率与诊疗率低。近些年来，抑郁症发病群体正呈现年轻化的趋势。当前，抑郁症的诊断主要依靠精神科医生的经验方法，误诊率高。及时在抑郁症早期进行筛查诊断，对于改善患者预后、减轻家庭和社会负担具有重要意义。本项目期望通过为青少年早期抑郁症大规模筛查提高一种客观有效、检出率高的抑郁症辅助诊断方法。



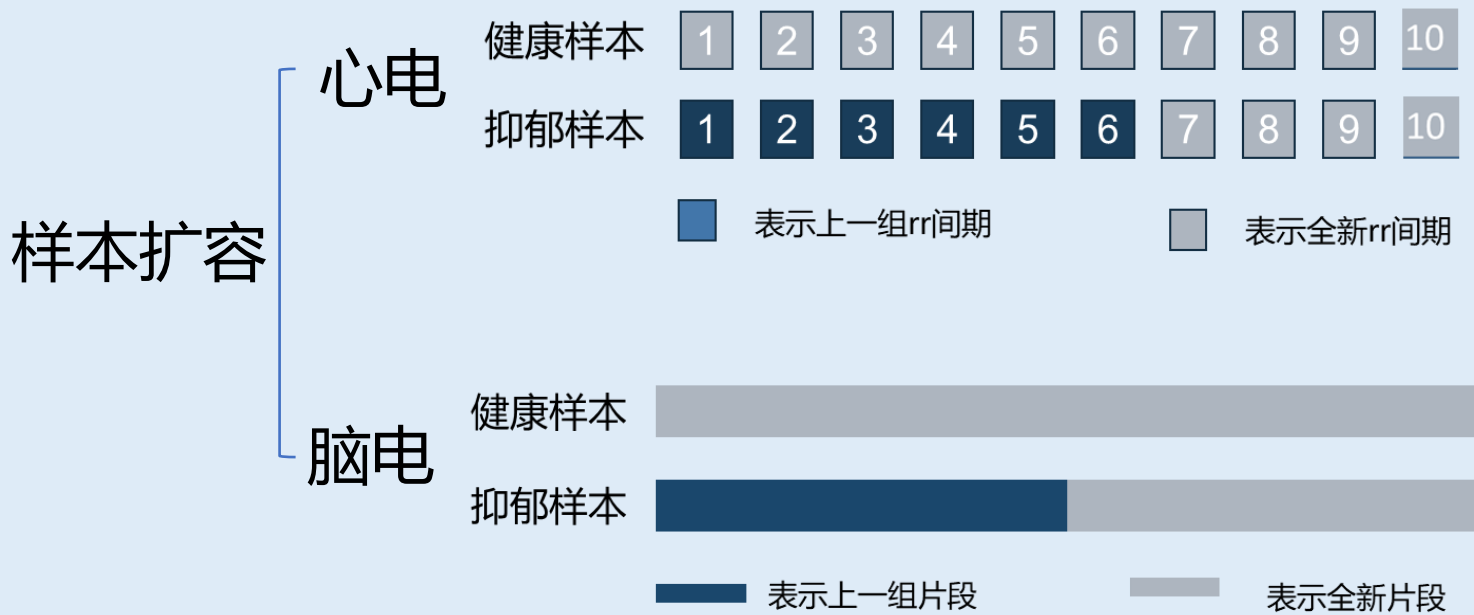
模型内容

支持向量机（SVM）主要是通过寻找一个最优超平面来最大化不同类别之间的边际，从而实现数据分类。SVM在处理小样本、高维特征数据时表现出色。

卷积神经网络擅长在序列数据中提取局部特征，长短期记忆网络擅长处理序列数据中的长期依赖问题。CNN-LSTM结合两者优点，在处理具有时间序列特性的大规模数据时表现出强大的能力。

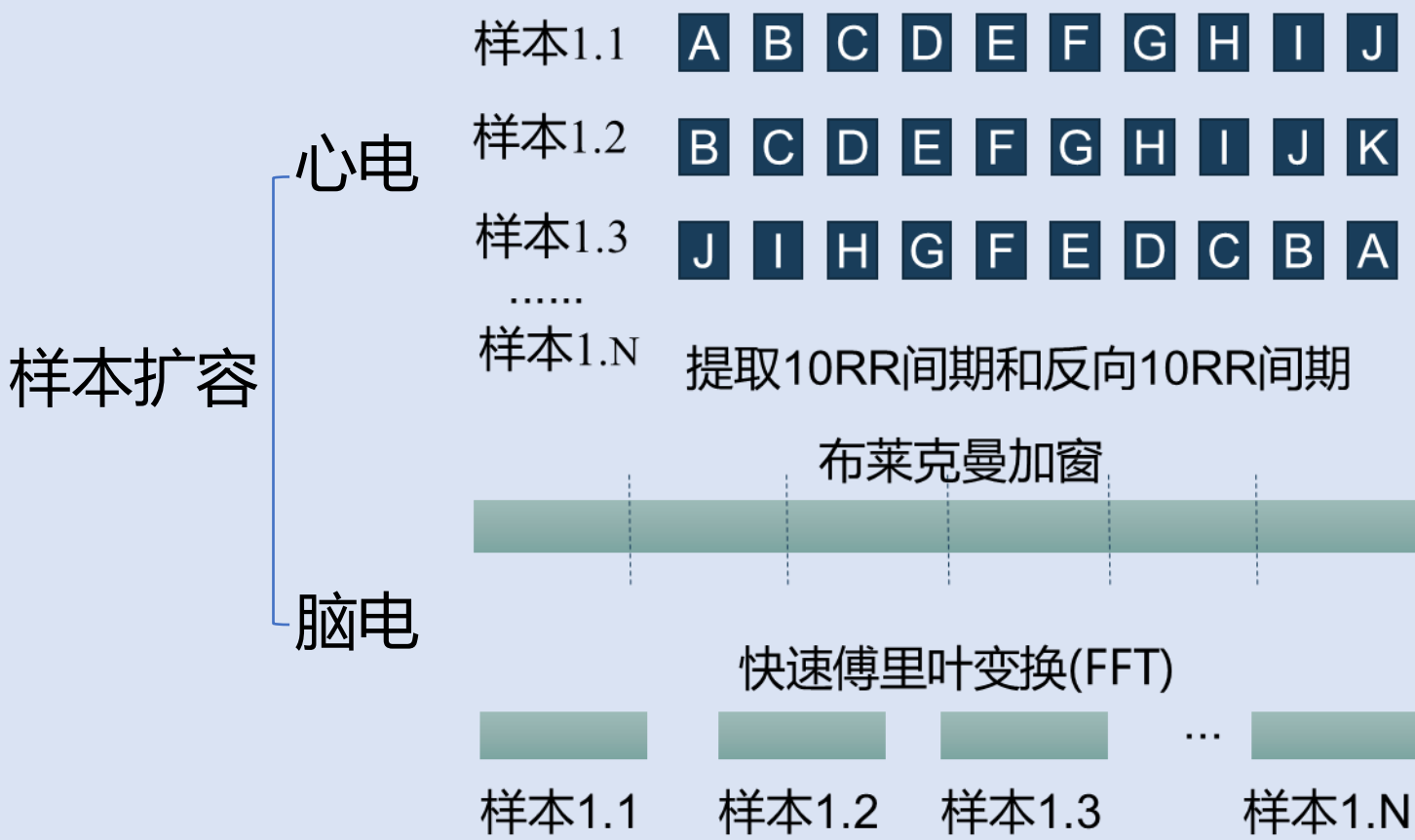


① 数据预处理(SVM)



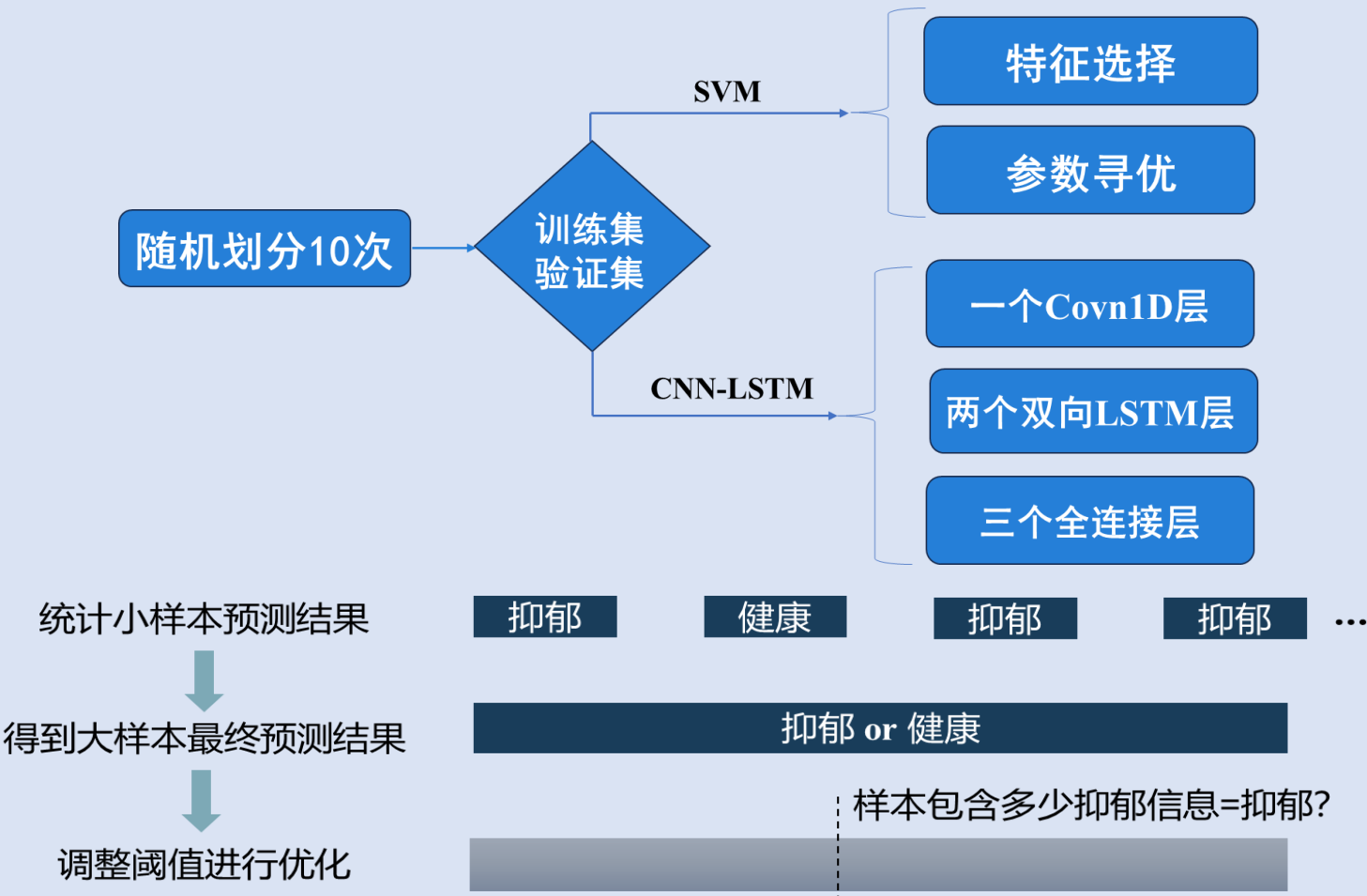
心电提取心脏变异率（HRV）的七个特征，脑电提取均值、方差、五种频带的能量和样本熵、近似熵，再通过互信息排序和循环比较选择模型最优的n个特征。

② 数据预处理(CNN-LSTM)



心电信号提取10RR间期，再通过时间反转和时间平移扩大心电数据量,达到数据增强的效果。脑电信号通过带通滤波以及ICA去伪迹后，输入到Blackman Window进行加窗处理以扩大脑电数据量，并进行FFT变换提取频域信息。

③ 模型训练与优化

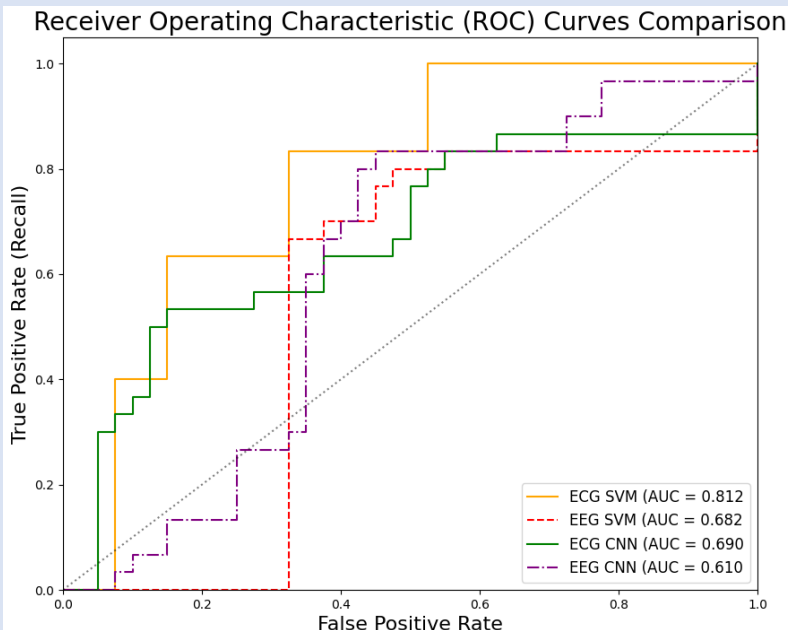


划分10次split以增加数据的容错性。SVM通过对多个特征的优化选取和对C、 γ 以及核函数（rbf、linear、poly）逐次尝试并选择最优参数。CNN-LSTM通过各层进行特征提取，时间依赖性分析和二分类。

结果分析

利用SVM和CNN-LSTM两种模型，选取心电和脑电两种模态，得到四种预测结果。对于验证集，SVM处理心电脑电的抑郁检出率均超过80%，F1分别达到0.64和0.74。CNN-LSTM的平均抑郁检出率达到了73.3%，F1达到了0.65。

以上结果显示出其在早期诊断中的潜力，为青少年早期抑郁症的筛查与预警提供了有效的辅助诊断方法。



数据扩容解决样本少问题

设置重复解决不平衡分类问题

抑郁片段比率提供诊断新思路