

基于语音特征的抑郁症 AI 筛查模型的研究与设计

张宵¹, 白雪俊¹, 唐琳¹, 张乐伊¹, 苏雪¹, 王金社^{1*}, 沈宇星^{2*}, 陈彦华^{3*}

1. 宁夏医科大学, 宁夏 银川 750004

2. 成都好麦科技有限公司, 四川 成都 610000

3. 宁夏医科大学总医院心理卫生中心, 宁夏 银川 750004

DOI: 10.61369/TACS.2025010025

摘要 : 针对传统抑郁量表^[2]筛查效率低的问题, 本研究提出基于语音特征的自动筛查模型。通过采集200例临床患者和健康个体的语音样本, 经预处理提取特征后, 构建结合 LSTM 时间建模与 Attention 机制的深度学习模型。测试显示模型准确率达84.62%, F1分数0.86, 在效率和一致性上优于传统量表。

关键词 : 抑郁症; 语音特征; LSTM-Attention 机制; 深度学习; 心理健康筛查

Research and Design of AI Screening Model for Depression Based on Speech Features

Zhang Xiao¹, Bai Xuejun¹, Tang Lin¹, Zhang Leyi¹, Su Xue¹, Wang Jinshe^{1*}, Shen Yuxing^{2*}, Chen Yanhua^{3*}

1.Ningxia Medical University, Yinchuan, Ningxia 750004

2.Chengdu Haomai Technology Co., Ltd., Chengdu, Sichuan 610000

3.Mental Health Center of Ningxia Medical University General Hospital, Yinchuan, Ningxia 750004

Abstract : To address the low efficiency of traditional depression scale screening^[2], this study proposes an automatic screening model based on speech features. By collecting voice samples from 200 clinical patients and healthy individuals, we constructed a deep learning model integrating LSTM temporal modeling and Attention mechanism after preprocessing and feature extraction. The model achieved an accuracy of 84.62% and F1-score of 0.86 in testing, outperforming traditional scales in efficiency and consistency.

Keywords : depression; speech features; lstm-attention mechanism; deep learning; mental health screening

引言

据世界卫生组织 (WHO) 2022 年统计, 全球抑郁症患者已超过3.8亿人, 但其中仅半数获得有效诊断^[1]。传统筛查抑郁症患者的方法主要根据临床医生的主观评估和患者自述量表 (如 PHQ-9^[3]、HAMD^[4]), 耗时长、误诊率高、覆盖人群有限。

语音特征是一种非侵入性的生物标记, 在多项研究中被证实与情绪状态高度相关^[11]。抑郁症患者常表现出语调单调、语速缓慢、音量降低及频谱能量异常等语音特征, 这些特征可通过算法自动提取并进行量化分析。

因此, 本研究提出了一种基于 LSTM-Attention 机制的深度学习模型, 为抑郁症的早期识别提供了非侵入性的解决方案。

一、数据集及预处理

(一) 数据来源及组成

本研究的语音数据包括100例宁夏医科大学总医院心理卫生中心确诊的抑郁症患者语音样本和100例宁夏医科大学健康学生志愿者语音样本。为保证语音信号没有背景噪音干扰, 在隔音室内使用高质量的录音设备进行录音。

(二) 数据预处理

先处理掉背景噪音及干扰讯号的原始语音讯号降噪。为了能够有效地保留语音信号的主要特征, 我们采用了基于小波变换的降噪算法^[9]。然后, 将降噪后的语音信号进行切片处理。最后, 通过规范处理切片后的语音信号, 杜绝音量大小在不同样本间的差别。

(三) 特征提取

在本研究中, 从语音信号中提取了三类特征: 韵律特征、音

基金项目: 名称: 2024年大学生创新创业训练计划立项项目

题目: 心理“特警”——基于语音特征的 AI 抑郁症筛查模型

编号: 202410752017

频特征和音谱特征，并对这些特征进行了深入研究。频谱特征是通过快速傅里叶变换（FFT）^[8] 绘制频谱图来识别谐波结构的频率特性。这些特征的提取为后续帮助分类性能增强的模型提供了有效支撑。

深度学习模型的 BiLSTM-Attention 机制是由 LSTM 网络^[6] 可以捕捉长期依赖的语音信号关系，避免梯度消失的问题，保留关键信息。Attention 机理^[7] 像一个专注于器一样，将不同的部分的权重分配到语音特征中去，突显出与抑郁症有关的特定节奏、频率变化等关键特征，从而提高识别准确率。在模型训练过程中，不断调整模型参数，逐步构建从语音特征到诊断结果的精确映射，采用交叉熵损失函数和 Adam 优化算法^[9]。

（四）模型评价

本次研究所采用的指标有准确度，敏感性，特异性及 F1 分值。准确率体现了模型整体分类的正确性，敏感度衡量的是模型对抑郁症患者识别能力的灵敏度，而特异性则体现的是模型的健康人群识别能力。F1 分数的综合考虑的是模型的精确度和召回率。

二、实验与结果

本研究所有实验都在一台 GPU 的电脑上进行，使用 Python 编程语言和 Pytorch 深度学习框架来实现，将数据集按 8:2 的比例随机划分为训练集和测试集。

（一）模型训练过程分析

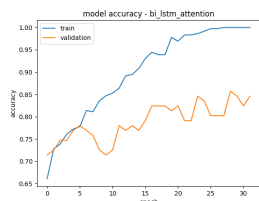


图1 模型训练与验证准确率变化

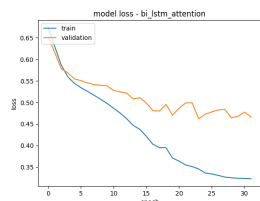


图2 模型训练与验证损失变化

如图1所示，模型训练准确率（蓝）随轮次从65%快速升至99%，验证准确率（橙）由71%缓升至85%，两者保持约14%差距，表明模型存在过拟合但具备实用潜力。图2中训练损失（蓝）由70%降至35%，验证损失（橙）由65%降至45%，始终维持差异，进一步印证模型泛化能力需提升。

（二）模型分类效果验证

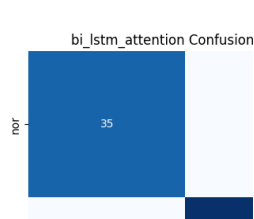


图3 混淆矩阵

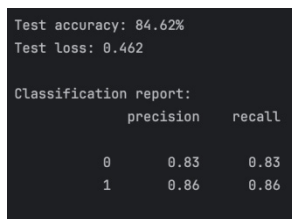


图4 分类报告

如图3所示，测试集含91例样本（健康42例，抑郁49例），比例均衡。健康人误判7例，患者误判7例，模型灵敏度85.7%、特异性83.3%，错误分布均衡，符合筛查需求。图4显示健康与抑郁 F1 分数分别为0.83和0.86，整体准确率84.62%，表明模型对两类样本的分类一致性高，平衡漏诊与误诊能力良好。

三、讨论

本模型在测试集中表现出较高的分类性能（准确率84.62%，灵敏度85.7%，特异性83.3%），显著优于传统量表筛查方法，可作为提高筛查效率的辅助诊断工具。此外，模型在健康人群与患者间的错误分布均衡（误判率均为14.3%），能够平衡漏诊与误诊风险，适用于大规模筛查场景。

然而，该筛查模型仍具有局限性。第一，资料集的规模有限、多样性受限，影响模型的泛化能力。未来要构建大规模数据集。同时可借鉴残差网络（ResNet）的深层特征提取能力^[12]。第二，模型的可解释性差。今后的研究引入可解释的 AIM 技术。第三，本研究只关注语音特征，没有考虑到其它模态资料。今后的研究可探索多模态数据融合^{[14][15]}的方法，如，引入 Transformer 架构^{[10][13]}。

四、结论

本研究构建了一种高效的、非侵入式的抑郁症 AI 筛查模型。实验表明，该模型在保持较高灵敏度和特异性的同时，还保持较高的准确率，为基层医疗机构的早期筛查提供了可行的技术方案。

未来将重点扩大数据集的范围和数量，提高模型的泛化能力，并联合多模态数据提升筛查精度，推动该技术向临床辅助诊断工具的转化，为抑郁症的早期干预提供更全面的技术支持。

参考文献

- [1] 世界卫生组织. 抑郁症及其他常见精神障碍：全球卫生估算报告[R]. 瑞士：世界卫生组织，2022.
- [2] World Health Organization. The ICD-10 classification of mental and behavioural disorders: Clinical descriptions and diagnostic guidelines[M]. Geneva: WHO, 1992.
- [3] Kroenke K, Spitzer R L. The PHQ-9: A new depression diagnostic and severity measure[J]. Psychiatric Annals, 2002, 32(9): 509-515.
- [4] Hamilton M. A rating scale for depression[J]. Journal of Neurology, Neurosurgery & Psychiatry, 1960, 23(1): 56-62.
- [5] Donoho D L, Johnstone I M. Ideal spatial adaptation by wavelet shrinkage[J]. Biometrika, 1994, 81(3): 425-455.
- [6] Hochreiter S, Schmidhuber J. Long short-term memory[J]. Neural Computation, 1997, 9(8): 1735-1780.
- [7] Bahdanau D, Cho K, Bengio Y. Neural machine translation by jointly learning to align and translate[J]. arXiv preprint arXiv:1409.0473, 2014.
- [8] Cooley J W, Tukey J W. An algorithm for the machine calculation of complex Fourier series[J]. Mathematics of Computation, 1965, 19(90): 297-301.
- [9] Kingma D P, Ba J. Adam: A method for stochastic optimization[J]. arXiv preprint arXiv:1412.6980, 2014.
- [10] Vaswani A, et al. Attention is all you need[C]. Advances in Neural Information Processing Systems, 2017: 5998-6008.
- [11] Cummins N, et al. A review of depression and suicide risk assessment using speech analysis[J]. Speech Communication, 2015, 71: 10-49.
- [12] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, NV, USA: IEEE, 2016: 770-778.
- [13] Devlin J, Chang M W, Lee K, et al. BERT: Pre-training of deep bidirectional transformers for language understanding[J]. arXiv:1810.04805, 2018.
- [14] Valstar M, Schuller B, Smith K, et al. AVEC 2016: Depression, mood, and emotion recognition workshop and challenge[C]//Proceedings of the 6th International Workshop on Audio/Visual Emotion Challenge. Amsterdam, Netherlands: ACM, 2016: 3-10.
- [15] Li X, Pang T, Liu Y, et al. Multimodal fusion for mental health assessment[J]. IEEE Transactions on Affective Computing, 2021, 12(3): 582-595.