



ScienceDirect®

情感障碍杂志

第323卷，2023年2月15日，第299-308页

研究论文

使用融合语音产生和感知特征的拟议语音链模型进行抑郁识别

Minghao Du^{a1}, Shuang Liu^{a1} , 王涛, 张文泉^a, Yufeng Ke^a, 龙晨^a, 董明^{a b} [显示更多](#)

轮廓 | 共享 引用

<https://doi.org/10.1016/j.jad.2022.11.060> [获取权利和内容](#) [在知识共享许可下](#)

公开权限

聚焦

- 基于机器语音链，提取LPC和MFCC，以互补捕获独立于文本的语音特征。
- 提出了一种分割和融合方法，从可变长度的语音中提取分段内和分段间特征。
- 识别的改善表明，患者的声道变化值得关注抑郁症诊断。

摘要

背景

越来越多的抑郁症患者给临床诊断带来了巨大的压力。基于音频的诊断是早期大规模筛查的有用辅助工具。然而，目前的方法只考虑语音感知特征，忽略了患者的声道变化，这可能部分导致识别不良。

方法

这项工作提出了一种新的抑郁识别机器语音链模型（MSCDR），该模型可以捕获从说话人的嘴到听众耳朵的与文本无关的抑郁语音表示，以提高识别性能。在拟议的MSCDR中，提取了线性预测编码（LPC）和Mel频率cepstral系数（MFCC）特征，分别描述语音生成和语音感知的过程。然后，一维卷积神经网络和长短期记忆网络依次捕获分段内和分段间动态抑郁特征进行分类。

结果

我们在两个具有不同语言和范式的公共数据集上测试了MSCDR，即困扰分析访谈奥兹博士和精神障碍分析的多模态开放数据集。MSCDR在两个数据集中的准确性为0.77和0.86，平均F1得分为0.75和0.86，优于其他现有方法。这一改进揭示了语音产生和感知特征在携带抑郁信息时的互补性。

限制

样本量相对较小，这可能会在一定程度上限制在临床翻译中的应用。

结论

这项实验证明了拟议的MSCDR的良好概括能力和优越性，并表明抑郁症患者的声道变化值得关注基于音频的抑郁症诊断。



关键词

抑郁症; 深度学习; 音频; 特征融合; 辅助诊断

1.介绍

抑郁症是一种常见但严重的心理障碍，其特点是持续悲观主义、认知能力下降和社会功能障碍

(Hammar等人, 2022年)。为了科学地预防抑郁症, 及时诊断以确保适当的治疗是必要的 (Costantini等人, 2021年)。世界卫生组织估计, 目前有3.22亿人患有抑郁症 (组织, 2017年), 这严重增加了诊断负担。因此, 需要自动方法来提高诊断能力。尽管基于脑电图 (Saeedi等人, 2021年)、基于心率 (Hartmann等人, 2019年) 和基于血液 (Sealock等人, 2021年) 的方法由于生理信号的客观性, 在抑郁症诊断中表现出良好表现, 但由于设备成本高昂和繁琐的收集过程使其难以普及。相比之下, 基于音频的抑郁症诊断更适合早期大规模筛查。这种方法捕捉辅助语言学差异来诊断抑郁症, 如韵律和语音质量。辅助语言学不关注有意识和主观的语义信息, 是一种无意识的人类交流现象, 包含丰富的态度、主题和情感内容 (Madhavi等人, 2020年)。与神经运动系统和生理状态相关, 言语中的副语言信息对神经退行性疾病的影响非常敏感 (Gómez-Rodellar等人, 2020年), 这可以作为与抑郁症相关的客观标志。此外, 音频收集只需要一个麦克风, 这个过程可以是非接触的, 这两者都增加了可访问性。重要的是, 由于音频的开放性, 主体隐私可以受到保护。因此, 基于音频的抑郁症诊断可以成为提高当前诊断能力的补充方法。

以前的研究证实了抑郁症患者的语音差异。患者通常呈现少说话、少说话、犹豫不决的临床语音表现 (Sahu和Espy-Wilson, 2014年)。经过进一步量化, 发现抑郁组和非抑郁组在声门、基本频率、抖动和闪烁方面的振幅频率特征存在显著差异 (Jia等人, 2019年; Silva等人, 2021年; Simantiraki等人, 2017年)。为了解释这一现象, 一些研究假设患者的神经肌肉协调因认知能力下降而受损, 他们进一步推测声道受到抑郁症的影响 (Espy-Wilson等人, 2019年; Seneviratne等人, 2020年)。基于这项研究, 将语音特征作为抑郁症诊断线索的可行性可以被认为是确认的。基于音频的抑郁症诊断的目标是通过发音特征来识别抑郁症, 无论语言、内容或说话习惯如何。为了实现这一目标, 最近的努力主要涉及两个方面: 语音特征提取和优化建模。

对于语音特征, 考虑到抑郁语音的感知差异, 手工描述词, 如速度、修音特征和光谱特征已被广泛使用。关于情绪感知, 短时间能量、强度、响度和零交叉率被提取为手工描述符, 它们在抑郁分类任务中显示出稳健性 (Long等人, 2017年)。关于音调感知, Lam-Cassettari和Kohlhoff (2020) 以及Patil和Wadhai (2021) 分析了抑郁组和非抑郁组之间的音高差异, 并证明了音高作为分类标记的可行性。关于听觉感知, 抑郁症患者的Mel频率cepstral系数 (MFCC-2) 的第二维度明显高于非抑郁症患者的, 这反映了2000-3000赫兹左右的频率能量差 (Taguchi等人, 2018年)。基于这些差异, 反映时间频率信息的MFCC、Mel光谱图和光谱图已被用于抑郁症诊断, 并显示出积极的性能 (He等人, 2022年; Rejaibi等人, 2022年; Vázquez-Romero和Gallardo-Antolín, 2020年; Yadav和Sharma, 2021年)。然而, 这些特征仅从语音感知过程中提取, 基于抑郁语音声音的感官差异, 而不是其产生方式。由于言语差异可能源于声道的变化, 仅从言语感知过程中提取特征会导致语音链的信息丢失 (Denes等人, 1993年; Tjandra等人, 2020年)。最近, 一项关于说话人身份识别的研究建立了一个语音链模型, 该模型可以从语音产生和语音感知过程中捕获语音身份特征 (Chowdhury和Ross ,

2020年)。这项工作使用线性预测编码 (LPC) 对说话者的声道进行建模, 并使用MFCC来描述人耳的感知规律。证明了拟议模型与现有方法相比的有效性, 这可能代表了抑郁症诊断的进步。因此, 我们认为从语音产生和语音感知过程中提取语音特征可以进一步提高抑郁识别。

自动抑郁症诊断模型广泛采用两种关键方法: 传统机器学习和神经网络。传统机器学习的代表, 如支持向量机 (Dai等人, 2021年; Valstar等人, 2016年)、线性回归 (Jiang等人, 2018年; Pan等人, 2018年) 和决策树 (Liu等人, 2020年; Pampouchidou等人, 2016年) 经常被选中进行分类, 但它们有一些局限性。根据这些模型的输入维度的要求, 从整个语音中提取的语音特征的统计函数 (平均值、中位数、方差) 通常被用作输入, 这忽略了与抑郁症密切相关的语音动态变化 (Wichers, 2014年)。相比之下, 神经网络不受输入维度的限制, 可以在时间或频域中提取动态信息。Srimadthur和Lalitha (2020) 提出了一个端到端卷积神经网络 (CNN) 框架, 以根据处理的音频识别抑郁症, 并取得了比传统机器学习模型更好的分类结果。Muzammel等人 (2020年) 将整个演讲分成几个部分, 并提取光谱特征, 然后建立了一个音素级CNN架构来捕捉元音和辅音声学特征。这种方法在语音片段上提供了出色的结果, 但整个语音没有经过测试。Zhao等人 (2021年) 专注于情感突出区域, 并提出了一个基于注意力的长短期记忆网络 (LSTM) 网络, 以获取关键的抑郁信息进行分类的时间信息。这些研究表明, 神经网络对动态信息很敏感。

分类模型的另一个挑战是阶级不平衡, 例如数量和语音持续时间不一致。以前的研究使用随机抽样 (He和Cao, 2018年; Zhao等人, 2021年)、调整大小 (Dong和Yang, 2021年; Othmani等人, 2021年) 和裁剪 (Ma等人, 2016年; Negi等人, 2018年) 来确保模型的无偏见, 但与抑郁症相关的信息可能会丢失。换句话说, 在几分钟的语音中只诊断几秒钟是不合理的, 光谱图的含义会在压缩后改变。幸运的是, (Rejaibi等人, 2022年) 提出了一个合奏系统, 该系统将语音分为片段, 以单位长度进行检测, 并通过硬投票分类器进行最终分类。然而, 抑郁症不仅反映在分段的分类比例上, 例如, 非抑郁症的受试者也会说负面的话, 抑郁症的受试者也表现出更少的积极性表达。因此, 需要探索分段之间更复杂的关系。

在这项工作中, 我们提出了一种新的抑郁识别 (MSCDR) 机器语音链模型。它有三个主要步骤。首先, 将原始语音预处理为片段, 然后提取40维LPC和39维MFCC特征, 分别描述语音生成和语音感知的过程。其次, 提出了一个一维卷积神经网络 (1D-CNN) 来提取段内抑制特征, 该网络由两个网络组成, 并行处理LPC和MFCC特征作为输入。最后, 使用特征级融合算法进行时间特征的融合, 并提出LSTM来捕获分段间抑郁相关特征进行分类。我们在英语数据集“奥兹遇险分析访谈” (DAIC-WOZ) 和中国数据集多模态开放数据集 (MODMA) 上使用了拟议的MSCDR, 并将分类结果与现有方法进行了比较, 以证明MSCDR的优越性和推广性。

这项工作的贡献可以总结如下:

- 基于机器语音链, LPC和MFCC特征从说话人的嘴提取到听众的耳朵, 以互补表示

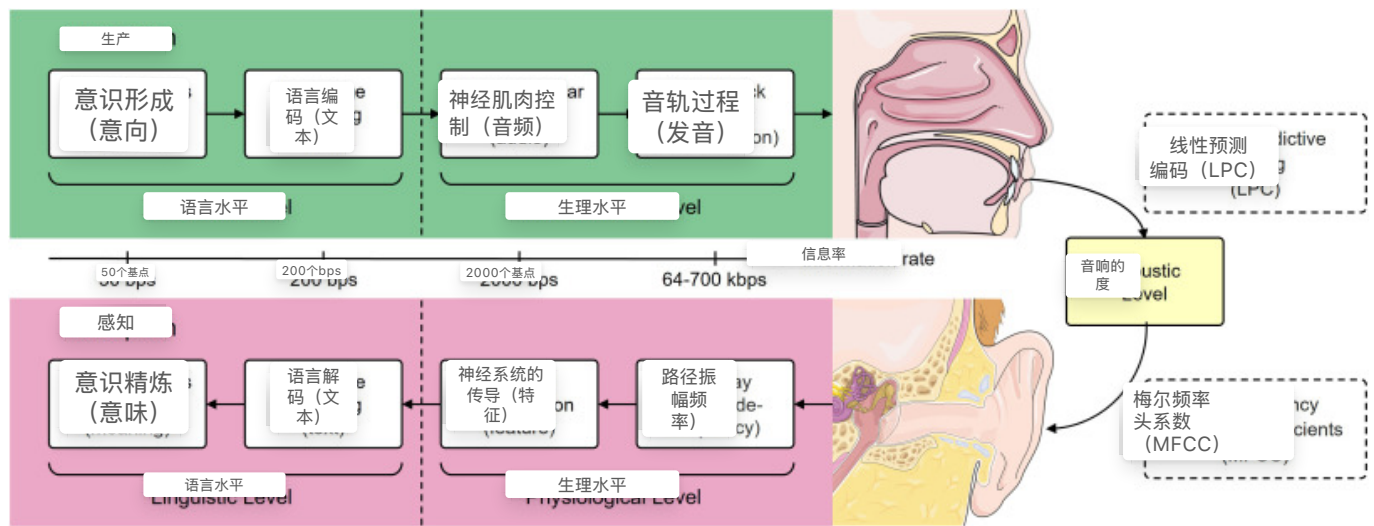
发音表示。

- 提出了一种分割和融合方法，从可变长度的语音中提取分段内和分段间特征，而不进行裁剪和冗余。
- 构建了一个框架来捕捉与文本无关的抑郁特征进行识别，这表明患者的声道变化也值得关注基于音频的抑郁诊断。
- 本文的其余部分组织如下。[第2节](#)介绍了拟议的MSCDR的理论基础和细节。[第3节](#)报告了实验设置和识别结果，[第4节](#)对此进行了讨论。最后，结论和未来的工作方向在[第5节](#)。

2.材料和方法

2.1.语音生成和感知特征

语音链概念最早由[Denes等人 \(1993年\)](#)引入，它解释了从说话人到听众的信息产生、传播和感知的闭环过程中所涉及的物理学和生物学。基于这一理论，[Chowdhury和Ross \(2020\)](#)更进一步，首先开发了一个基于深度学习的闭环语音链模型，该模型集成了人类语音感知和生产行为进行身份识别。与独立系统相比，这项工作提高了性能。我们根据以前研究中的描述再现了语音链的可视化 ([Denes等人, 1993年](#); [Tjandra等人, 2020年](#)) ([图. 1](#))。在生产过程中，信息在语言层面被编码为文本，声道从发音中产生声音，从而在生理层面上传递口语的声学特性。在感知过程中，声学metus提取了生理层面上必不可少的语音特征，文本在语言层面上被解码为意义。值得注意的是，传输的语音信息中包含的信息速率明显高于文本信息本身的基本信息速率。因此，抑郁症患者的语音差异可以客观地反映在语音产生和生成过程中，而不是以前研究中使用的分离过程。我们首先实施了抑郁症诊断的语音链模型，该模型集成了人类语音感知和生产语音行为。参考之前对说话人身份识别的研究 ([Chowdhury和Ross, 2020年](#))，我们使用LPC对语音生成过程进行建模，使用MFCC对语音感知过程进行建模。该组合补充覆盖闭环语音链，并有效地提取与抑郁症相关的信息。



下载：下载高分辨率图像 (707KB)

下载：下载全尺寸图像

图。1.根据Denes等人的说法，语音链。(1993) ; Tjandra等人。(2020)。

2.1.1.线性预测编码

根据语音的源滤波器模型，人类的声音从肺部作为能量源激发，并通过声道作为滤波器进行处理 (Guzman等人，2020年；Mittal和Sharma，2021年)。语音信号中包含的信息是由声道作为时变滤波器的调制形成的，而不是能量源。线性预测编码 (LPC) 是模拟声道以反映说话人特征的数字滤波器参数。因为人的声音是一个高度相关的序列， p 过去语音样本的线性组合可以预测下一个语音样本 $\hat{x}(n)$ ，由以下方式给出：

$$\hat{x}(n) = \sum_{k=1}^p a_k x(n-k).$$

a_k 是声道滤波系数； $x(k)$ ($k = 1, 2, 3, \dots, p$)是 k 个过去语音样本。真正的 n 个语音样本是 $x(n)$ ，预测误差 e_n 可以给出为：

$$e_n = x(n) - \hat{x}(n) = x(n) - \sum_{k=1}^p a_k x(n-k).$$

通过最小化 e_n 的均方误差，可以得到滤波系数 a_k ($k = 1, 2, 3, \dots, p$)作为 p 阶LPC，这提供了人类声道滤波系数的估计。

2.1.2.Mel频率Cepstral系数

如图1所示，听觉通路根据声音的频率内容分离声音，并将声波转换为大脑的神经信号。Mel频率Cepstral系数 (MFCC) 模拟人类外周听觉系统，并广泛用于语音识别 (Rejaibi等人，2022

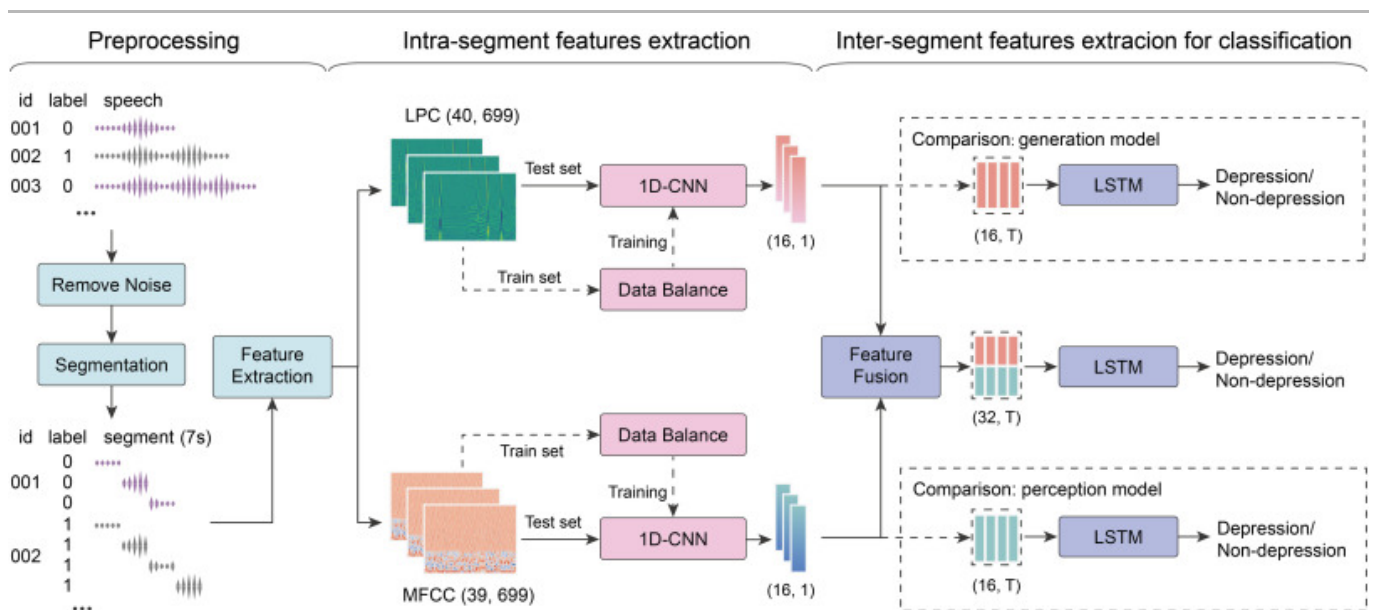
年)。MFCC以非线性尺度,即梅尔尺度描述了头肌的能量。这个音阶反映了人耳的特性,它对低频声音比对高频声音更敏感。Mel尺度和频率之间的关系可以通过以下方式近似:

$$Mel(f) = 2595 \times \lg\left(1 + \frac{f}{700}\right).$$

MFCC提取方式如下:1)从频率计算快速傅里叶变换频谱,2)提取分配在梅尔刻度上的滤波器库输出,3)通过离散余弦变换获得分频系数(Guzman等人,2020年)。

2.2.拟议的MSCDR

拟议的MSCDR的整个过程由三部分组成:预处理、分段内特征提取和分段间特征提取进行分类,如图2所示。每个主题的原始演讲依次分为部分,然后在预处理后提取LPC和MFCC特征。建立1D-CNN是为了从LPC和MFCC特征中提取段内高水平抑郁特征。之后,每个受试者的所有分段特征在时域中融合,并通过LSTM提取分段之间的抑郁相关性信息进行分类。为了进一步验证机器语音链的改进,构建了两个单一模型:生成模型仅从语音生成过程中提取语音特征,感知模型仅从语音感知过程中提取。



下载: 下载高分辨率图像 (567KB)

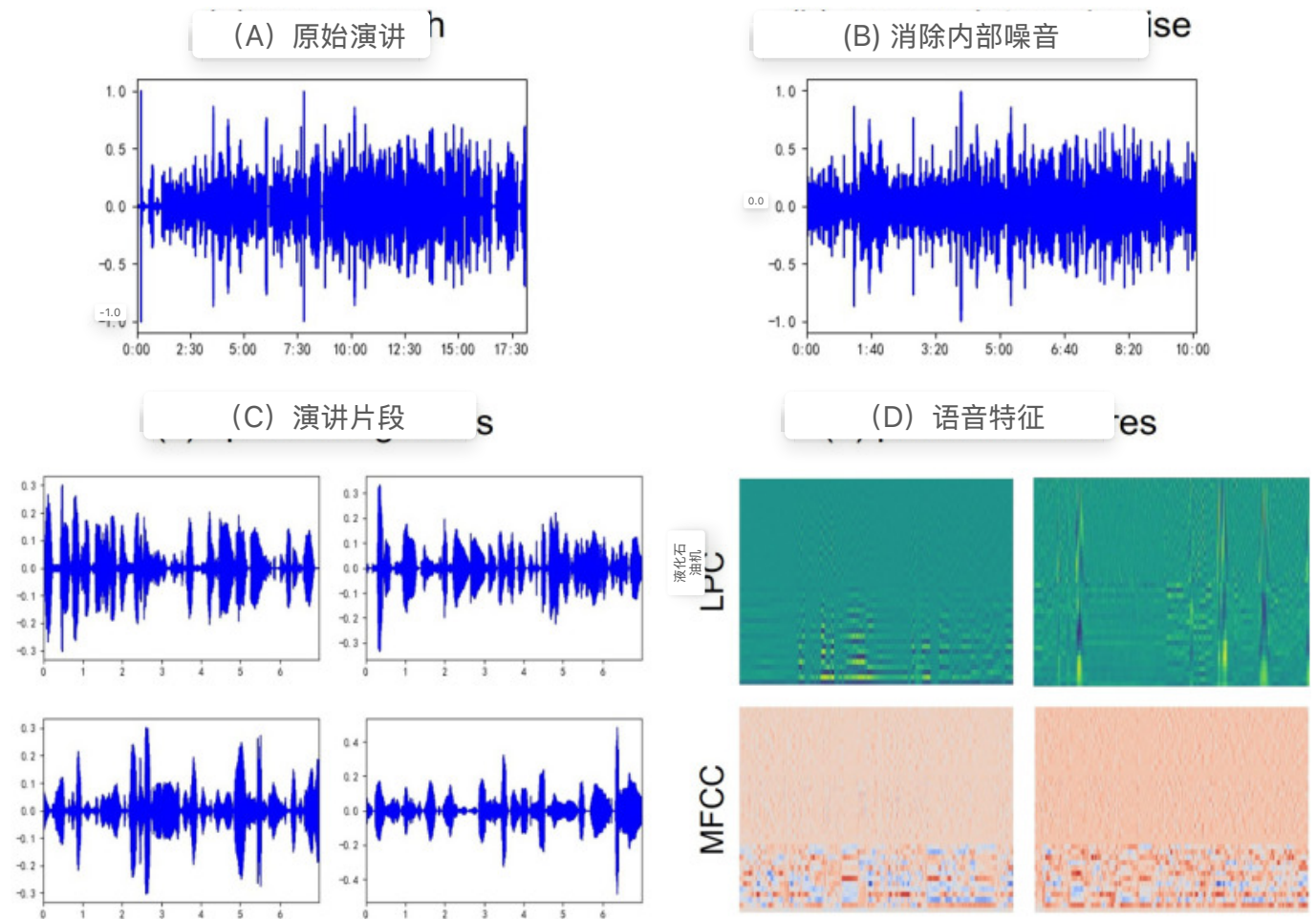
下载: 下载全尺寸图像

图2.拟议的MSCDR整体过程的可视化,包括三个部分:预处理、分段内特征提取和分段间特征提取,用于分类。黑色虚线代表训练阶段。在1D-CNN的训练阶段,正样本数和负样本数有一个平衡过程。虚拟框中有两个单一模型供比较。

2.2.1.预处理

原始语音包含在收集过程中捕获的内部噪声,如面试官的声音和静音剪辑,这些与抑郁无关

联，因此会影响识别性能。我们删除了噪音部分，然后将整个语音分成7秒的片段，没有重叠，并记录了它们的顺序，如图3所示。3.如前所述，这种分割用于统一不同长度的演讲，并增加训练样本的数量。分割过程中的语义破坏不影响与文本无关的分类模型。7 s作为分段长度基于枚举实验的结果，与以前的研究一致（Alghifari等人，2019年）。之后，使用相同长度 $[0.025 \times fs]$ 的滑动窗口（ fs 是采样频率）和哈明窗口提取LPC和MFCC特征。LPC特征包括20个滤波系数和20个一阶delta系数，MFCC特征包括13个Mel-cepstral系数、13个一阶和13个二阶delta系数。



下载：下载高分辨率图像 (610KB)

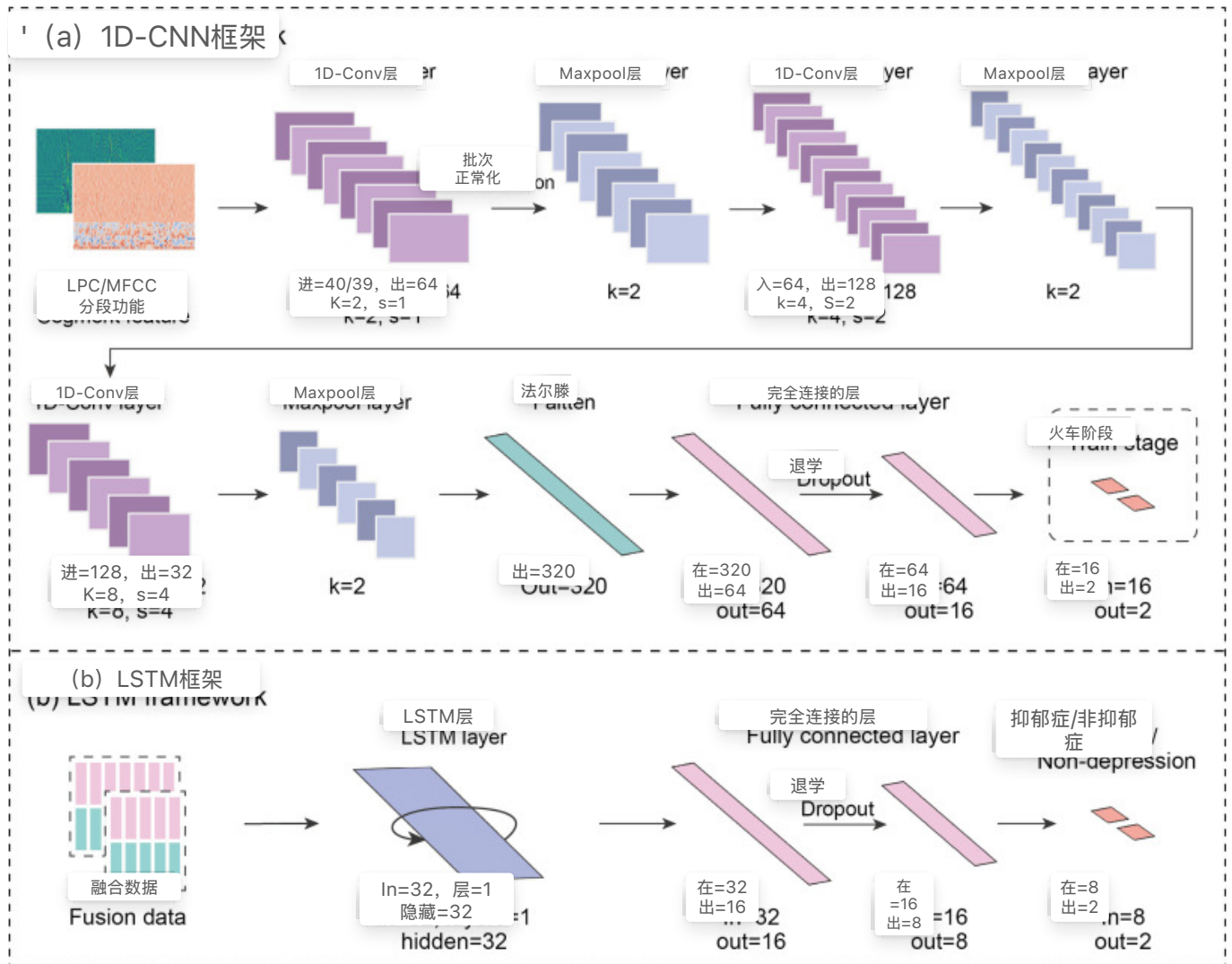
下载：下载全尺寸图像

图。3.一个主体的部分预处理结果，具有 (a) 原始语音，(b) 无内部噪声的语音，(c) 7秒的语音片段和 (d) 语音特征。

2.2.2.段内特征提取

经过预处理后，每个受试者的所有片段都被混合，以消除受试者身份的影响，并建立了1D-CNN，从片段中提取与抑郁症相关的高级特征。由于LPC和MFCC的轴代表不同的大小，并且对应于具有完全不同含义的时间和频率维度，因此1D-CNN对所有频率的卷积提高了模型对频域

的灵敏度 (Vázquez-Romero和Gallardo-Antolín, 2020年)。图4 (a) 显示了1D-CNN框架的结构。1D-CNN和1D最大池的组合使该模型能够有效地捕获短期时间动态信息和频率相关性。批量归一化和辍学层提高了训练速度, 并防止过度拟合。密集层进一步提取特征并减少输出维度。为了避免班级不平衡, 在训练阶段, 训练组中抑郁和非抑郁部分的数量以1:1的比例平衡。在测试阶段, 每个段的倒数第二密度层的16维输出仅保留给下一个会话。



下载：下载高分辨率图像 (782KB)

下载：下载全尺寸图像

图. 4.建议的深度学习架构包括 (a) 用于段内特征的1D-CNN框架和 (b) 用于段间特征和分类的LSTM框架。虚线中的密集层仅在训练期间用于计算网络权重。对于LSTM, MSCDR的输入维度为32, 单个模型的输入维度为16。K代表内核大小, s代表步长。

2.2.3.用于分类的跨段特征提取

段内特征提取后, 获得16维输出, 包括短段的凹陷相关信息。MSCDR将每个LPC和MFCC特征的输出从每个段连接到32维段特征, 以整合语音生成和感知过程。然后, 每个主题的所有特征

都按照原始时间域顺序从分段级别拼接到个人级别。对于单个模型，所有分段特征都直接拼接到个人层，没有维度串行。最后，构建了一个单层LSTM，以捕捉个人层面的片段之间的短期和长期时间相关特征。如图4 (b) 所示，包括两个密集(layer)的层来减少维度并完成分类。LSTM的递归层消耗可变长度的输入，最终在最后一个顺序步骤中仅产生该层的输出，这有效地处理了不同主体的语音长度不一致。

2.2.4.绩效指标

分类性能是使用混淆矩阵、准确性和F1分数确定的，这与之前的研究类似 (Valstar等人, 2016年; Zhao等人, 2021年)。F1分数是精度和召回的谐波平均值，它是不平衡分类问题的有用评估标准。

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN},$$

$$Precision = \frac{TP}{TP+FP},$$

$$Recall = \frac{TP}{TP+FN},$$

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall}.$$

TP/FP表示真/假阳性样本，TN/FN表示真/假阴性样本。更大的F1分数意味着更好的辨别能力。

3.结果

拟议的MSCDR可以通过发音表示来识别抑郁症，无论语言、内容或说话习惯如何。为了验证该方法的文本独立性，我们在两个具有不同范式和语言的数据集上进行了测试，并将分类结果与以前的研究进行了比较。

3.1.两个公共数据集

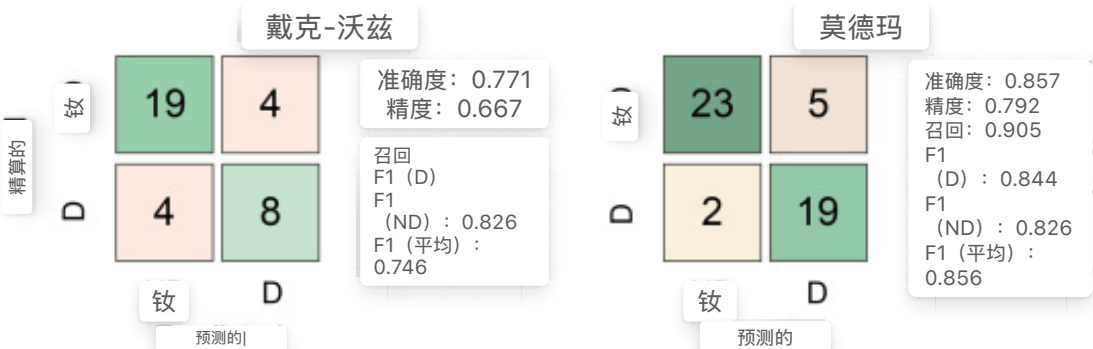
DAIC-WOZ (Gratch等人, 2014年) 得到了AVEC2017挑战 (Ringeval等人, 2017年) 的支持，有助于了解几种典型的精神障碍：例如焦虑、抑郁和创伤后应激障碍。它记录了189名受试者的临床访谈音频，并由专业人士结合PHQ-8二进制来区分抑郁症和非抑郁症患者。由一个名为Ellie的人类控制的虚拟面试官主持的范式。Ellie有一组固定的发语，它根据受试者的实时反应提供反馈。英语是提问和回答的语言。录音的平均长度为15分钟，采样频率为16千赫。与DAIC-WOZ的公开划分一致，训练集 (30抑郁与77非抑郁) 和发展集 (12抑郁与23非抑郁) 用于训练和测试。

由中国兰州大学支持的MODMA (Cai等人, 2022年) 适用于精神障碍分析。它包含52名受试者 (23名抑郁症门诊患者，29名非抑郁症患者和1名抑郁症数据缺陷)。抑郁症患者是从符合

《[精神障碍诊断和统计手册](#)》（DSM）重度抑郁症诊断标准的住院患者和门诊患者中招募的。健康对照组由海报招募，并被排除在其他疾病之外。每个受试者被要求以不同的口语模式完成29项录音任务：采访、单词和段落阅读，以及在三种情感价下进行图片描述：积极、中性和消极。由于归纳不力，段落阅读录音被排除在外，这在讨论部分进行了解释。口语是中文，录音长度从几秒钟到几分钟不等。为了避免录制时间太短，我们结合了同一主题的几段录音，并以8:2的比例随机分为一组训练集（89个抑郁与117个非抑郁）和一个测试集（21个抑郁与28个非抑郁）。

3.2.识别性能

在MSCDR的训练阶段，PyTorch用于实现系统，学习率为0.01，批次大小为16，交叉熵为损失函数，以及防止过度拟合的早期停止机制。[图5](#)显示了DAIC-WOZ和MODMA测试集上生成的混淆矩阵。我们计算了评估指标，并将其与以前的模型进行了比较（[表1](#)）。拟议的MSCDR在DAIC-WOZ和MODMA上的准确性分别为0.77和0.86，平均F1得分分别为0.75和0.86，比现有方法高出0.09和0.07。特别是，DAIC-WOZ上抑郁症的F1评分显著提高了0.17，表明MSCDR显著增强了捕捉抑郁症信息的能力。



[下载：下载高分辨率图像 \(184KB \)](#)
[下载：下载全尺寸图像](#)

图。5.在DAIC-WOZ和MODMA的测试集上生成了拟议的MSCDR的混淆矩阵。ND代表非抑郁，D代表抑郁。

表1. 拟议的MSCDR与现有方法的比较。MODMA的第二行和第三行分别是基于RNN方法（[Rejaibi等人，2022年](#)）和DepAudioNet（[Ma等人，2016年](#)）的再现结果。ND代表非抑郁，D代表抑郁。

数据集	方法	特点	模型	F1得分 (D/ND)	F1得分 (平均)
戴克-沃	Valstar等人，2016年	F0、VUV和	SVM	0.41/0.58	0.50

兹		MFCC			
	马等人，2016年	梅尔光谱图	美国有线电视新闻网-LSTM	0.52/0.70	0.61
	黄等人，2020年	MFCC	FVTC-CNN	0.40/0.84	0.62
	Rejaibi等人，2022年	MFCC	RNN	0.46/0.85	0.64
	Othmani等人，2021年	MFCC，光谱图	美国有线电视新闻网	0.49/0.82	0.66
	Dumpala等人，2021年	开放式微笑	LSTM	0.50/0.82	0.66
	MSCDR (我们的)	LPC-MFCC	美国有线电视新闻网-LSTM	0.67/0.83	0.75
莫德玛	陈和潘，2021年	eGe地图	决策树	0.80/-	-
	Rejaibi等人，2022年	MFCC	RNN	0.66/0.73	0.70
	马等人，2016年	梅尔光谱图	美国有线电视新闻网-LSTM	0.82/0.75	0.79
	MSCDR (我们的)	LPC-MFCC	美国有线电视新闻网-LSTM	0.84/0.87	0.86

DAIC-WOZ的基线 (Valstar等人，2016年) 和MODMA的方法 (Chen和Pan，2021年) 仅对整个语音特征使用统计函数 (例如，平均值、中位数)，这些函数对时间变化不敏感。相比之下，研究 (Ma等人，2016年；Huang等人，2020年；Rejaibi等人，2022年；Othmani等人，2021年) 提取了包含动态信息的语音 (Mel-Spectrogram，MFCC) 的时频特征，并获得了更好的结果。然而，这些方法使用的特征仅来自语音感知过程，没有考虑到患者声道的变化。与这些相比，拟议的MSCDR从语音生成和语音感知过程中提取，并进一步有效地提高了分类性能。抑郁症F1评分的提高反映了1D-CNN和LSTM在提取抑郁症相关信息方面的有效性。考虑到并非抑郁症患者的全部言语都是抑郁症，非抑郁症患者也可能有抑郁症片段，1D-CNN首先在片段水平上提取抑郁症相关信息，而不是整个言语，因为它对频域敏感。然后，LSTM网络通过捕获段间相关性进行分类。与基于全语音的分类相比，这种方法为抑郁症诊断提供了新的想法。此外，英语数据集DAIC-WOZ和中文数据集MODMA的优秀结果证明了MSCDR的文本独立性，这符合基于音频的抑郁症诊断的要求，即对发音特征进行分类，无论语言、内容或说话习惯如何。

4.讨论

4.1.与单一模型的比较

为了进一步验证机器语音链的改进，我们在相同条件下将MSCDR与两个单一模型进行了比较，如图 2 所示。生成模型仅从语音生成过程中提取了与抑郁相关的特征，而感知模型仅从语音感知过程中提取了它们。表 2 中的结果表明，与单一模型相比，使用MSCDR有显著改善。拟议的MSCDR从生成过程中提取LPC，从感知过程中提取MFCC。正如LPC描述说话者的声道，MFCC描述人耳的感知规律一样，它们的组合代表发音表示的互补性。这种改善也在一定程度上证明了抑郁症患者的声道发生了变化，这与假设一致 (Espy-Wilson等人，2019年；Seneviratne等人，2020年)。因此，抑郁症患者生理结构的变化也值得关注基于语音的抑郁症自动诊断，与手动诊断相比，这将导致改善。

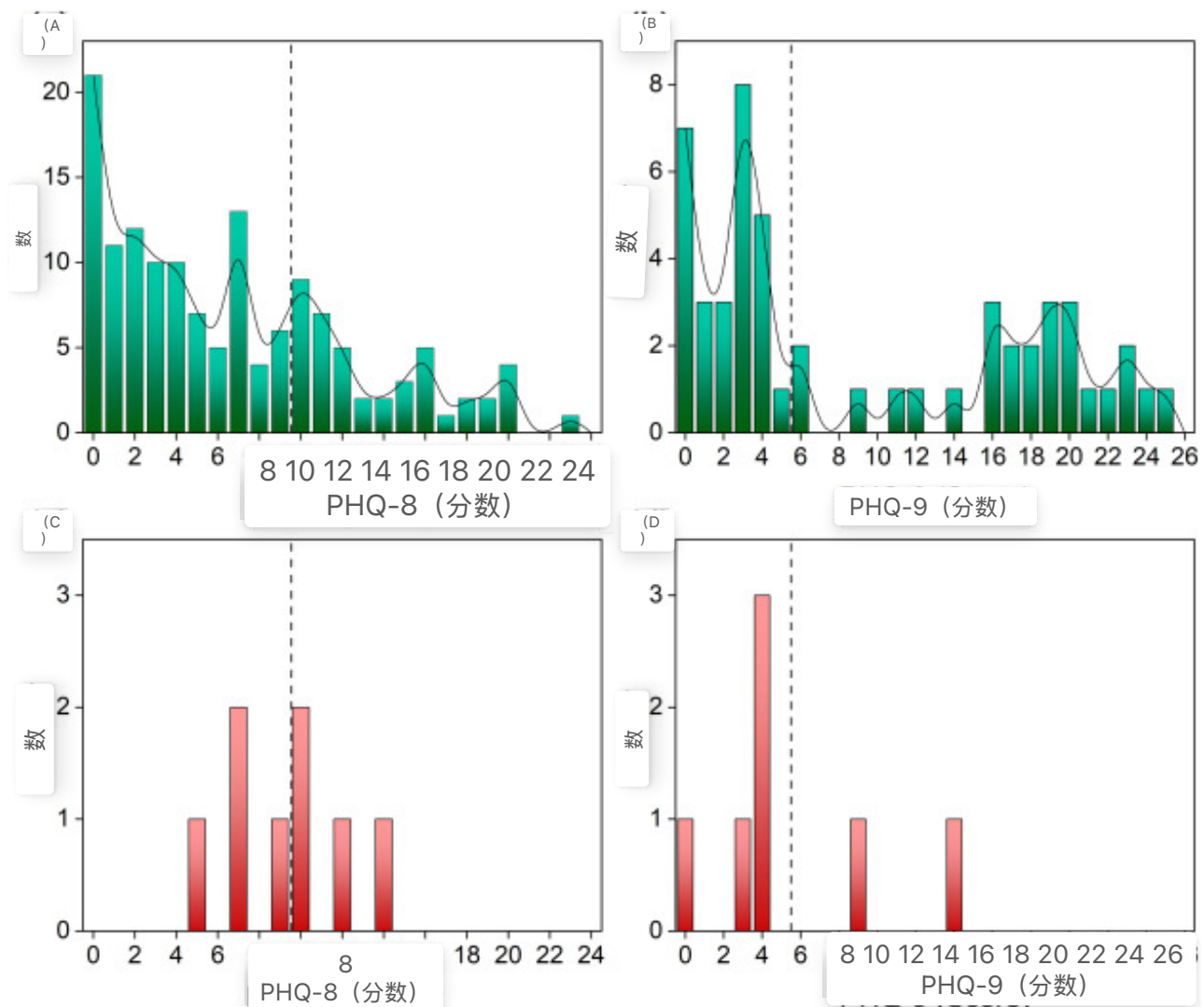
表2。将拟议的MSCDR与两种单一方法进行比较：生成模型仅来自语音生成过程，感知模型仅来自语音感知过程。ND代表非抑郁，D代表抑郁。

数据集	模型	准确度	F1得分		
			D	ND	平均
DAIC-沃兹	世代模式	0.771	0.500	0.852	0.676
	感知模型	0.714	0.583	0.783	0.683
	MSCDR	0.771	0.667	0.826	0.746
莫德玛	世代模式	0.837	0.789	0.867	0.828
	感知模型	0.816	0.790	0.836	0.814
	MSCDR	0.857	0.844	0.868	0.856

4.2.错误分类分析

尽管MSCDR在两个数据集中的分类结果都很出色，但DAIC-WOZ的平均F1得分明显低于MODMA，明显差距为0.11。我们分析了可能影响DAIC-WOZ性能的三个因素。第一个因素是DAIC-WOZ中可能的标签错误。MODMA的标签通过多个尺度和医生诊断科学地共同标记，包括PHQ-9 (Kroenke和Spitzer，2002年)、 (Gerdner和Allgulander，2009年) 和GAD-7 (Spitzer等人，2006年)。然而，DAIC-WOZ仅使用PHQ-8量表 (Kroenke等人，2009年) 对受试者的标签进行二元化，这在临床诊断中并不严格，很有可能出现错误标签。第二个因素是抑郁症状在DAIC-WOZ中并不突出。如图 6 (a) 和 (b) 所示，DAIC-WOZ的评分分布集中在非抑郁或轻度患者中，而MODMA的评分分布更分散。DAIC-WOZ的这种集中分布可能会影响模型的训练。此外，图 6 (c) 显示，DAIC-WOZ中分类错误的受试者集中在抑郁症和非抑郁

症的阈值上。如图6 (d) 所示，MODMA数据集中也发生了同样的现象。因此，第三个因素是，抑郁症的音频特征可能受到锯齿，即一些具有高评分的非抑郁症患者也可能表现出抑郁语音特征，一些轻度患者可能没有抑郁症语音症状。这种现象也值得注意，以前没有提到过。



下载：下载高分辨率图像 (561KB)

下载：下载全尺寸图像

图. 6. (a) DAIC-WOZ受试者的PHQ-8分数分布。(b) MODMA受试者的PHQ-9分数分布。(c) DAIC-WOZ分类错误受试者的PHQ-8分数分布。(d) MODMA分类错误受试者的PHQ-9分数分布。垂直虚线表示抑郁和非抑郁的阈值。

4.3.演讲任务选择

与只有采访任务的DAIC-WOZ不同，MODMA由五个演讲任务组成：采访、段落阅读、单词阅读、图片描述和主题感知测试 (TAT)，它们具有不同的诱导效果。与 (Taguchi等人, 2018 年) 类似，我们对抑郁组和非抑郁组之间MFCC特征的每个维度进行了统计分析，以分析不同任务的诱导效果。首先，我们使用Kolmogorov-Smirnov检验来验证样品是否符合正态分布。

然后，Levene检验被用来检验方差的同质性。如果满意，则使用学生的 t 检验；否则，使用Welch的 t 检验。我们使用错误发现率（FDR）计算来纠正我们的假设测试发现。[表3](#)显示了每个任务的39维统计分析结果。正如我们所看到的，段落阅读任务中存在显著差异的特征数量远远低于其他任务，这表明该任务的诱导效应很低，这与之前的结果一致（[Long等人，2017年](#)；[Rejaibi等人，2022年](#)）。我们认为，这是由于其固定内容和受试者的一致熟悉，因此将MODMA中的段落阅读录音排除在这项研究之外。

表3.MODMA.p < 0.05中五个语音任务的39维MFCC特征的 p 值表明存在显著差异。

特点	采访	段落阅读	单词阅读	图片描述	纹身
MFCC-0	0.01	0.02	0.07	0.01	0.01
MFCC-1	0.01	0.08	0.04	0.00	0.00
MFCC-2	0.84	0.59	0.79	0.92	0.79
MFCC-3	0.84	0.90	0.53	0.34	0.49
MFCC-4	0.85	0.98	0.85	0.58	0.94
MFCC-5	0.28	0.97	0.44	0.24	0.00
MFCC-6	0.77	0.62	0.48	0.75	0.68
MFCC-7	0.00	0.09	0.00	0.00	0.45
MFCC-8型	0.22	0.61	0.41	0.25	0.14
MFCC-9	0.04	0.04	0.07	0.01	0.01
MFCC-10	0.70	0.62	0.48	0.73	0.92
MFCC-11	0.39	0.27	0.02	0.19	0.52
MFCC-12型	0.14	0.58	0.59	0.04	0.21
MFCC-13	0.35	0.63	0.43	0.85	0.02
MFCC-14	0.49	0.49	0.87	0.66	0.19
MFCC-15	0.12	0.37	0.70	0.20	0.32
MFCC-16型	0.12	0.37	0.70	0.20	0.32
MFCC-17号	0.81	0.88	0.60	0.31	0.35
MFCC-18	0.83	0.71	0.43	0.46	0.36

MFCC-19	0.54	0.13	0.87	0.86	0.46
MFCC-20	0.72	0.70	0.61	0.79	0.36
MFCC-21	0.49	0.10	0.86	0.05	0.56
MFCC-22	0.52	0.35	0.42	0.84	0.65
MFCC-23	1.00	0.90	0.66	0.68	0.86
MFCC-24	0.12	0.53	0.02	0.05	0.77
MFCC-25型	0.83	0.76	0.03	0.47	0.19
MFCC-26	0.45	0.62	0.39	0.85	0.02
MFCC-27号	0.44	0.51	0.85	0.60	0.20
MFCC-28	0.12	0.42	0.91	0.21	0.34
MFCC-29	0.17	0.05	0.53	0.10	0.62
MFCC-30型	0.81	0.94	0.60	0.42	0.51
MFCC-31	0.90	0.61	0.54	0.31	0.28
MFCC-32	0.66	0.11	0.91	0.85	0.44
MFCC-33	0.52	0.60	0.53	0.87	0.28
MFCC-34号	0.59	0.07	0.89	0.03	0.78
MFCC-35	0.41	0.33	0.27	0.76	0.51
MFCC-36	1.00	0.98	0.69	0.73	0.94
MFCC-37	0.03	0.41	0.03	0.05	0.83
MFCC-38	0.63	0.66	0.04	0.38	0.12
p的数量<0.05	5	2	7	6	6

4.4.简要总结

我们的结果表明了辅助语言表达与抑郁症之间的潜在关联，并进一步表明，言语可以作为早期发现精神障碍的有力工具。目前，有许多关于精神障碍的生理项目来探索其认知和病理机制。例如，对新陈代谢（[Bocchio-Chiavetto等人，2018年](#)）、基因（[Li等人，2021年](#)）、脑电图（[Saedi等人，2021年](#)）、磁共振成像（[Squarcina等人，2017年](#)）的研究在探索其生理机制

方面取得了一些重要进展。我们相信，言语可以成为理解精神障碍的重要补充，因为它的获取成本低，受欢迎程度高。此外，基于音频的诊断技术可以应用于手机和手环等智能设备，以积极检测人们的心理健康，这可以应对社会中潜在的心理风险，并具有广泛的应用前景。

5. 结论

在这项研究中，我们提出了一个MSCDR，该MSCDR从语音感知和生产过程中提取语音特征，用于自动抑郁识别。具有不同范式和语言的两个数据集的出色分类结果证明了拟议的MSCDR的良好泛化能力和优越性。由于样本量小，我们无法将MSCDR应用于抑郁症水平的诊断。接下来，我们将扩大样本大小，并在临床翻译前进行进一步验证。我们认为，这项研究表明，抑郁症患者声道的变化值得关注，也为基于音频的抑郁症诊断研究提供了理论基础和灵感。

CRedit作者贡献声明

M. 杜，S. 刘和D. Ming设计了这项研究。T.王和L.Chen做了数据策划和正式分析。M. 杜，W. 张和Y. Ke构建了模型，进行了测试和可视化。M. 杜和S. 刘起草了手稿。所有作者都为结果的解释、手稿修订做出了贡献，并批准了手稿的最终版本。

资金

由中国国家自然科学基金会在Grant81925020和81801786下以及中国天津总计划在Grant19JCYBJC29200下支持的研究。

利益冲突

作者没有报告潜在的利益冲突。

致谢

作者衷心感谢DAIC-WOZ和MODMA的收藏家和参与者为这项研究提供了音频数据。图1包含由知识共享署名3.0未移植许可证许可的Servier Medical Art的修改图像。

推荐文章

参考文献

Alghifari等人，2019年 M.F. Alghifari，T.S.古纳万，M.A.W.诺丁，M. 卡蒂维，L. 博尔汉
关于抑郁检测的最佳语音段长度

2019年IEEE智能仪器、测量和应用国际会议 (ICSIMA) , IEEE (2019) , 第1-5页 ,
[10.1109/ICSIMA47653.2019.9057319 ↗](#)

[谷歌学术 ↗](#)

[Bocchio-Chiavetto等人, 2018年](#) L. Bocchio-Chiavetto , R. 扎纳尔迪尼 , S. 托佐藤 , M. Ventriglia , C. Ferrari , C. 博内托 , A. 拉萨尔维亚 , F. Giubilini , A. Fioritti , F. Pileggi , M. 普拉泰利 , M. 帕瓦纳蒂 , A. 法瓦罗 , G. De Girolamo , G.B. 弗里索尼 , M. 鲁格里 , M. 吉纳雷利

初发作精神病 (FEP) 患者的免疫和代谢变化

大脑行为。免疫 , 70 (2018) , 第315-324页 , [10.1016/j.bbi.2018.03.013 ↗](#)

 [查看PDF](#) [查看文章](#) [在Scopus中查看 ↗](#) [谷歌学术 ↗](#)

[蔡等人, 2022年](#) H. 蔡 , Z. 元 , Y. 高 , S. 太阳 , N. 李 , F. 田 , H. 肖 , J. 李 , Z. 杨 , X. 李 , Q. 赵 , Z. 刘 , Z. 姚 , M. 杨 , H. 彭 , J. Zhu , X. Zhang , G. Gao , F. Zheng , R. 李 , Z. 郭 , R. 马 , J. 杨 , L. 张 , X. 胡 , Y. 李 , B. 胡

用于精神障碍分析的多模态开放数据集

科学。数据 , 9 (2022) , p. 178 , [10.1038/s41597-022-01211-x ↗](#)

[在Scopus中查看 ↗](#) [谷歌学术 ↗](#)

[陈和潘, 2021年](#) X. 陈 , Z. 潘

基于语音数据用于公共心理健康的抑郁筛查和预警的方便和低成本模型

内。J. 环境。Res. 公共卫生 , 18 (2021) , p.6441 , [10.3390/ijerph18126441 ↗](#)

[在Scopus中查看 ↗](#) [谷歌学术 ↗](#)

[Chowdhury和Ross, 2020年](#) A. Chowdhury , A. 罗斯

融合MFCC和LPC功能, 使用1D三联体CNN在严重退化的音频信号中进行扬声器识别

IEEE Trans. Inf. Forensics Secur. , 15 (2020) , 第1616-1629页 ,
[10.1109/TIFS.2019.2941773 ↗](#)

[在Scopus中查看 ↗](#) [谷歌学术 ↗](#)

[Costantini等人, 2021年](#) L. 康斯坦蒂尼 , A. 科斯坦扎 , A. 奥多内 , A. 阿古利亚 , A. Escelsior , G. Serafini , M. Amore , A. 阿梅里奥

抑郁症筛查研究的突破：从验证到疗效研究

Acta Biomed. Ateneo Parmense , 92 (2021) , [10.23750/abm.v92i3.11574 ↗](#)
e2021215–e2021215

[谷歌学术 ↗](#)

[Dai等人, 2021年](#) Z. Dai , H. 周 , Q. 巴 , Y. 周 , L. 王 , G. 李

使用新颖的特征选择算法与上下文感知分析相结合，改善抑郁预测

J。影响。Disord.，295（2021），第1040-1048页，[10.1016/j.jad.2021.09.001](#) ↗



[查看PDF](#)

[查看文章](#)

[在Scopus中查看](#) ↗

[谷歌学术](#) ↗

Denes等人，1993年 P.B.德内斯，P. 德内斯，E. 平森

语音链

麦克米兰（1993）

[谷歌学术](#) ↗

董和杨，2021年 Y. 董，X.杨

基于语音和情绪线索的分层抑郁检测模型

神经计算，441（2021），第279-290页，[10.1016/j.neucom.2021.02.019](#) ↗



[查看PDF](#)

[查看文章](#)

[在Scopus中查看](#) ↗

[谷歌学术](#) ↗

Dumpala等人，2021年 S.H.Dumpala，S. 罗德里格斯，S. Rempel，R. Uher，S. 奥尔
演讲者嵌入和时间背景对抑郁症检测的意义

（2021），[10.48550/arXiv.2107.13969](#) ↗

阿尔Xiv

[谷歌学术](#) ↗

Espy-Wilson等人，2019年 C.埃斯皮-威尔逊，A.C.拉默特，N. Seneviratne，T.F. Quatieri

使用倒声道变量评估抑郁症中的神经运动协调

2019年演讲间。在2019年Interspeech上发表，ISCA（2019），第1448-1452，

[10.21437/Interspeech.2019-1815](#) 页 ↗

[在Scopus中查看](#) ↗

[谷歌学术](#) ↗

Gerdner和Allgulander，2009年 A.格德纳，C. 阿奥尔古兰德

瑞典版童年创伤问卷的心理测量属性——简表（CTQ-SF）

北方。J。精神病学，63（2009），第160-170页，[10.1080/08039480802514366](#) ↗

[在Scopus中查看](#) ↗

[谷歌学术](#) ↗

Gómez-Rodellar等人，2020年 A.戈麦斯-罗德拉尔，D. Palacios-Alonso，J.M.费兰德斯-维森

特，J.梅基斯卡，A. Álvarez-Marquina，P.戈麦斯-维尔达

基于声门流声学分析的区分帕金森病和衰老语音的方法

内。J。神经系统，30（2020），p.2050058，[10.1142/S0129065720500586](#) ↗

[在Scopus中查看](#) ↗

[谷歌学术](#) ↗

格拉奇等人，2014年 J. 格拉奇，R. Artstein，G. Lucas，G. Stratou，S.舍勒，A. 纳扎里安，

R.伍德，J.博伯格，D. DeVault，S. 马塞拉

人类和计算机访谈的苦恼分析访谈语料库

第九届国际语言资源和评价会议 (LREC'14) 会议记录 (2014年) , 第3123-3128页

[在Scopus中查看](#) [谷歌学术](#)

Guzman等人, 2020年 M。Guzman , T. Bertucci , C.Pacheco , F. Leiva , F. Quintana , R.Ansaldi , C.Quezada , D.穆诺兹

基于不同半闭塞声道练习的生理语音治疗方案在行为失声患者中的有效性：一项随机对照试验

J。公社。Disord. , 87 (2020) , 第106023条 , [10.1016/j.jcomdis.2020.106023](#)

 [查看PDF](#) [查看文章](#) [在Scopus中查看](#) [谷歌学术](#)

Hammar等人, 2022年 Å。哈马 , E.H.罗纳德 , G.Å.雷克达尔

重度抑郁症的认知障碍和神经认知特征——临床视角

前面。精神病学 , 13 (2022) , [10.3389/fpsyt.2022.764374](#)

[谷歌学术](#)

Hartmann等人, 2019年 R。哈特曼 , F.M.施密特 , C。桑德 , U。赫格尔
心率变异性作为抑郁症临床状态的指标

前面。精神病学 , 9 (2019) , p.735 , [10.3389/fpsyt.2018.00735](#)

[在Scopus中查看](#) [谷歌学术](#)

他和曹, 2018年 L。他 , C。曹

使用语音卷积神经网络进行自动抑郁分析

J。生物医学。通知 , 83 (2018) , 第103-111页 , [10.1016/j.jbi.2018.05.007](#)

 [查看PDF](#) [查看文章](#) [在Scopus中查看](#) [谷歌学术](#)

He等人, 2022年 L。他 , M。牛 , P。蒂瓦里 , P。马丁宁 , R。苏 , J。姜 , C。郭 , H。王 , S。丁 , Z.王

用视听线索进行抑郁症识别的深度学习：回顾

Inf。融合 , 80 (2022) , 第56-86页 , [10.1016/j.inffus.2021.10.012](#)

 [查看PDF](#) [查看文章](#) [在Scopus中查看](#) [谷歌学术](#)

黄等人, 2020年 Z.黄 , J.Epps , D。约阿希姆

利用扩张的CNNS利用声道协调在自然环境中检测抑郁

ICASSP 2020-2020 IEEE声学、语音和信号处理国际会议 (ICASSP) , 在ICASSP 2020-2020 IEEE声学、语音和信号处理国际会议 (ICASSP) (2020) 上发表 , 第6549-

6553, [10.1109/ICASSP40776.2020.9054323](#)页

[在Scopus中查看](#) [谷歌学术](#)

Jia等人, 2019年 Y. 贾, Y. 梁, T. 朱

中国抑郁症患者语音质量分析

2019年第22届东方COCOSDA语音数据库和评估技术协调与标准化国际委员会 (O-COCOSDA) 会议, IEEE (2019), 第1-6页, [10.1109/O-COCOSDA46868.2019.9060848](#) ↗



[查看PDF](#) [查看文章](#) [在Scopus中查看](#) ↗ [谷歌学术](#) ↗

Jiang等人, 2018年 H. 江, B. Hu, Z. Liu, G. Wang, L. 张, X. 李, H. 康

使用基于多种语音特征的合奏逻辑回归模型检测抑郁症

计算。数学。医学方法, 2018年 (2018年), p.6508319, [10.1155/2018/6508319](#) ↗

[在Scopus中查看](#) ↗ [谷歌学术](#) ↗

Kroenke和Spitzer, 2002年 K. 克罗恩克, R.L. 斯皮策

PHQ-9: 一种新的抑郁症诊断和严重程度措施

精神病学家。Ann., 32 (2002), 第509-515页, [10.3928/0048-5713-20020901-06](#) ↗

[在Scopus中查看](#) ↗ [谷歌学术](#) ↗

Kroenke等人, 2009年 K. 克罗恩克, T.W. 斯特林, R.L. 斯皮策, J.B. 威廉姆斯, J.T. 贝里, A.H. 莫克达德

PHQ-8作为当前普通人群抑郁的衡量标准

J. 影响。Disord., 114 (2009), 第163-173页, [10.1016/j.jad.2008.06.026](#) ↗



[查看PDF](#) [查看文章](#) [在Scopus中查看](#) ↗ [谷歌学术](#) ↗

Lam-Cassettari和Kohlhoff, 2020年 C. Lam-Cassettari, J. 科尔霍夫

母亲抑郁对语言学前婴儿的婴儿说话的影响: 对语言发育的影响

PLOS ONE, 15 (2020), 文章e0236787, [10.1371/journal.pone.0236787](#) ↗

[在Scopus中查看](#) ↗ [谷歌学术](#) ↗

李等人, 2021年 X. 李, X. 苏, J. 刘, H. 李, M. 李, W. 李, X.-J. 罗

转录组范围的关联研究确定了新的易感性基因和抑郁症的途径

翻译。精神病学, 11 (2021), p. 306, [10.1038/s41398-021-01411-w](#) ↗

[谷歌学术](#) ↗

刘等人, 2020年 Z. 刘, D. 王, L. 张, B. 胡

语音中抑郁症识别的新型决策树

ArXiv Prepr (2020)

ArXiv200212759

[谷歌学术](#) ↗

Long等人, 2017年 H. Long, Z. Guo, X. Wu, B. 胡, Z. 刘, H. 蔡

检测言语抑郁：不同言语类型的比较和组合

2017年IEEE生物信息学和生物医学国际会议 (BIBM)。在2017年IEEE生物信息学和生物医学国际会议 (BIBM) 上发表, IEEE, 密苏里州堪萨斯城 (2017年), 第1052-1058、[10.1109/BIBM.2017.8217802](#)页 ↗

[在Scopus中查看](#) ↗ [谷歌学术](#) ↗

马等人, 2016年 X. 马, H. 楊, Q. 陳, D. 黄, Y. 王

DepAudioNet: 基于音频的抑郁分类的高效深度模型

第六届音频/视觉情感挑战国际研讨会的会议记录。在MM '16上发表: ACM多媒体会议, ACM, 荷兰阿姆斯特丹 (2016), 第35-42页, [10.1145/2988257.2988267](#) ↗

[谷歌学术](#) ↗

Madhavi等人, 2020年 我。马达维, S. 查米什卡, R. 纳瓦拉特内, V. Nanayakkara, D. Alahakoon, D. 德席尔瓦

一种深度学习方法, 用于从网络物理环境中的音频流中检测与工作相关的压力

2020年第25届IEEE新兴技术和工厂自动化国际会议 (ETFA)。在2020年第25届IEEE新兴技术和工厂自动化 (ETFA) 国际会议上发表, IEEE, 奥地利维也纳 (2020), 第929-936, [10.1109/ETFA46521.2020.9212098](#)页 ↗

[在Scopus中查看](#) ↗ [谷歌学术](#) ↗

Mittal和Sharma, 2021年 V. 米塔尔, R.K. 夏尔马

基于语音信号分析和综合的帕金森病分类

内。J. 健康c. Inf. 系统。通知, 16 (2021), [10.4018/IJHISI.20211001.0a30](#) ↗

[谷歌学术](#) ↗

Muzammel等人, 2020年 M. 穆扎梅尔, H. 萨拉姆, Y. 霍夫曼, M. 切图阿尼, A. 奥斯曼尼

AudVowelConsNet: 用于临床抑郁症诊断的基于音素水平的深度CNN架构

马赫。学习。申请, 2 (2020), 第100005条, [10.1016/j.mlwa.2020.100005](#) ↗



[查看PDF](#) [查看文章](#) [谷歌学术](#) ↗

Negi等人, 2018年 H. 内基, T. 博拉, M.S. 皮莱, D. 库马尔

一种使用音频情绪分析检测抑郁症的新方法

内。J. Inf. 系统。管理。科学。 (2018), p. 1

[谷歌学术](#) ↗

Othmani等人, 2021年 A. 奥斯曼尼, D. 卡多奇, K. Bentounes, E. 雷贾比, R. 阿尔弗雷德, A. 哈迪德

走向强大的深度神经网络, 用于语音影响和抑郁识别

国际模式识别会议，斯普林格 (2021)，第5-19页，[10.1007/978-3-030-68790-8_1](https://doi.org/10.1007/978-3-030-68790-8_1) ↗

[在Scopus中查看](#) ↗ [谷歌学术](#) ↗

Pampouchidou等人，2016年 A.Pampouchidou，O. Simantiraki，A.法兹洛拉希，M. 足部炎，D.Manousos，A.Roniotis，G. Giannakakis，F. Meriaudeau，P.西莫斯，K. Marias，F. Yang，M.Tsiknakis

通过融合音频、视频和文本的高级和低级功能来评估抑郁症

第六届音频/视觉情感挑战国际研讨会会议记录，AVEC '16，计算机协会，美国纽约州纽约 (2016年)，第27-34页，[10.1145/2988257.2988266](https://doi.org/10.1145/2988257.2988266) ↗

[在Scopus中查看](#) ↗ [谷歌学术](#) ↗

Pan等人，2018年 W. 潘，J. 王，T.刘，X.刘，M.刘，B. 胡，T.朱

基于语音分析的抑郁症识别

下巴。科学。Bull.，63 (2018)，第2081-2092页，[10.1360/N972017-01250](https://doi.org/10.1360/N972017-01250) ↗

[在Scopus中查看](#) ↗ [谷歌学术](#) ↗

Patil和Wadhai，2021年 M. 帕蒂尔，V. 瓦达伊

使用声学特征选择用于凹陷检测的分类器

2021年计算智能和计算应用国际会议 (ICCICA)。在2021年计算智能和计算应用国际会议 (ICCICA) (2021年) 上发表，第1-4,[10.1109/ICCICA52458.2021.9697240](https://doi.org/10.1109/ICCICA52458.2021.9697240) ↗

[谷歌学术](#) ↗

Rejaibi等人，2022年 E. Rejaibi，A. Komaty，F. Meriaudeau，S.阿格雷比，A. 奥斯曼尼

基于MFCC的递归神经网络，用于自动识别和评估语音临床抑郁症

生物医学。信号过程。控制，71 (2022)，第103107条，[10.1016/j.bspc.2021.103107](https://doi.org/10.1016/j.bspc.2021.103107) ↗

 [查看PDF](#) [查看文章](#) [在Scopus中查看](#) ↗ [谷歌学术](#) ↗

Ringeval等人，2017年 F. Ringeval，B.舒勒，M. 瓦尔斯塔，J. 格拉奇，R. 考伊，S. 舍勒，S.Mozgai，N.康明斯，M. 施密特，M. 潘提克

Avec 2017：现实生活中的抑郁症，以及影响识别研讨会和挑战

第七届音频/视觉情感挑战年度研讨会 (2017年)，第3-9页，[10.1145/3133944.3133953](https://doi.org/10.1145/3133944.3133953) ↗

[在Scopus中查看](#) ↗ [谷歌学术](#) ↗

Saeedi等人，2021年 A.赛迪，M. 赛迪，A. Maghsoudi，A. 沙尔巴夫

基于脑电图信号的有效连接性的重度抑郁症诊断：卷积神经网络和长短期记忆方法

认知。Neurodyn.，15 (2021)，第239-252页，[10.1007/s11571-020-09619-0](https://doi.org/10.1007/s11571-020-09619-0) ↗

[在Scopus中查看](#) ↗ [谷歌学术](#) ↗

Sahu和Espy-Wilson , 2014年 S。萨胡 , C。埃斯皮-威尔逊

抑郁症对言语的影响

J。声。索克。上午 , 136 (2014)

2312-2312

[谷歌学术 ↗](#)

Sealock等人 , 2021年 J.M.西洛克 , Y.H.李 , A。莫斯卡蒂 , S。Venkatesh , G. Voloudakis , P.斯特劳布 , K.辛格 , Y.-C.A.Feng , T. Ge , P.鲁索斯

使用PsycheMERGE网络来调查抑郁症多基因评分与白细胞计数之间的关联

JAMA精神病学 , 78 (2021) , 第1365-1374页 , [10.1001/jamapsychiatry.2021.2959 ↗](#)

[在Scopus中查看 ↗](#) [谷歌学术 ↗](#)

Seneviratne等人 , 2020年 N。塞内维拉特内 , J.R.威廉森 , A.C.Lammert , T.F. Quatieri , C.埃斯皮-威尔逊

关于使用声道变量量化抑郁症中神经运动协调的扩展研究

2020年演讲间。在2020年Interspeech上发表 , ISCA (2020) , 第4551-

4555,[10.21437/Interspeech.2020-2758](#)页 [↗](#)

[在Scopus中查看 ↗](#) [谷歌学术 ↗](#)

Silva等人 , 2021年 W.J.席尔瓦 , L。洛佩斯 , M.K.C.加尔迪诺 , A.A.阿尔梅达

声音声学参数作为抑郁症的预测因素

J。语音 , S0892199721002058 (2021) , [10.1016/j.jvoice.2021.06.018 ↗](#)

[谷歌学术 ↗](#)

Simantiraki等人 , 2017年 O. Simantiraki , P.Charonyktakis , A。潘普奇杜 , M.Tsiknakis , M.Cooke

基于语音的自动抑郁评估的语音源功能

2017年演讲间。在2017年Interspeech上发表 , ISCA (2017) , 第

2700-2704,[10.21437/Interspeech.2017-1251](#)页 [↗](#)

[在Scopus中查看 ↗](#) [谷歌学术 ↗](#)

Spitzer等人 , 2006年 R.L.斯皮策 , K。Kroenke , J.B.威廉姆斯 , B。洛韦

评估广泛性焦虑症的简要措施 : GAD-7

拱门。实习生。Med. , 166 (2006) , 第1092-1097页 , [10.1001/archinte.166.10.1092](#)页 [↗](#)

[在Scopus中查看 ↗](#) [谷歌学术 ↗](#)

Squarcina等人 , 2017年 L。斯夸奇纳 , U。卡斯泰拉尼 , M。贝拉尼 , C。佩里尼 , A。拉萨尔

维亚 , N。杜西 , C。博内托 , D。克里斯托法洛 , S。Tosato , G. Rambaldelli , F.

Alessandrini , G. Zoccatelli , R.Pozzi-Mucelli , D.拉莫纳卡 , E。Ceccato , F. Pileggi , F.

Mazzi , P.桑托纳斯塔索 , M. 鲁杰里 , P. 布兰比拉

使用支持向量机和多核学习技术对大量患者队列的第一发作精神病进行分类

NeuroImage , 145 (2017) , 第238-245页 , [10.1016/j.neuroimage.2015.12.007](https://doi.org/10.1016/j.neuroimage.2015.12.007) ↗

 [查看PDF](#) [查看文章](#) [在Scopus中查看](#) ↗ [谷歌学术](#) ↗

Srimadhur和Lalitha , 2020年 N.S.Srimadhur , S. 拉利莎

使用语音检测和评估抑郁程度的端到端模型

普罗塞迪亚计算。科学 , 第三届计算和网络通信国际会议 (CoCoNet'19) , 171 (2020) , 第12-21页 , [10.1016/j.procs.2020.04.003](https://doi.org/10.1016/j.procs.2020.04.003) ↗

 [查看PDF](#) [查看文章](#) [在Scopus中查看](#) ↗ [谷歌学术](#) ↗

Taguchi等人 , 2018年 T. Taguchi , H.立川 , K. 尼莫托 , M. 铃木 , T.长野 , R.立花 , M.西村 , T. 新井

使用声学特征辨别重度抑郁症

J. 影响。Disord. , 225 (2018) , 第214-220页 , [10.1016/j.jad.2017.08.038](https://doi.org/10.1016/j.jad.2017.08.038) ↗

 [查看PDF](#) [查看文章](#) [在Scopus中查看](#) ↗ [谷歌学术](#) ↗

Tjandra等人 , 2020年 A.Tjandra , S. 萨克蒂 , S. 中村

机器语音链

IEEEACM翻译。音频语音语言。过程。 , 28 (2020) , 第976-989, [10.1109/TASLP.2020.2977776](https://doi.org/10.1109/TASLP.2020.2977776) 页 ↗

[在Scopus中查看](#) ↗ [谷歌学术](#) ↗

Valstar等人 , 2016年 M. 瓦尔斯塔 , J. 格拉奇 , B. 舒勒 , F.林格瓦尔 , D.拉兰 , M. 托雷斯·托雷斯 , S. 舍勒 , G.斯特拉图 , R.考伊 , M. 潘提克

Avec 2016 : 抑郁、情绪和情绪识别研讨会和挑战

第六届音频/视觉情感挑战国际研讨会会议记录 (2016年) , 第3-10页 , [10.1145/2988257.2988258](https://doi.org/10.1145/2988257.2988258) ↗

[在Scopus中查看](#) ↗ [谷歌学术](#) ↗

Vásquez-Romero和Gallardo-Antolín , 2020年 A.Vázquez-Romero , A. 加利亚多-安托林

使用合奏卷积神经网络自动检测语音中的抑郁

熵 , 22 (2020) , p. 688 , [10.3390/e22060688](https://doi.org/10.3390/e22060688) ↗

[在Scopus中查看](#) ↗ [谷歌学术](#) ↗

巫師 , 2014年 M. 巫師

抑郁症的动态性质 : 应对当前挑战的精神障碍的新微观视角

心理。Med. , 44 (2014) , 第1349-1360页 , [10.1017/S0033291713001979](https://doi.org/10.1017/S0033291713001979) ↗

[在Scopus中查看](#) ↗ [谷歌学术](#) ↗

[世界卫生组织, 2017年](#) 世界卫生组织

抑郁症和其他常见精神障碍：全球健康估计

世界卫生组织 (2017)

[谷歌学术](#) ↗

[Yadav和Sharma, 2021年](#) 你。亚达夫, A.K.夏尔马

使用情绪分析从视听线索中自动检测抑郁症的回顾

2021年第二届电子和可持续通信系统国际会议 (ICESC) , IEEE (2021) , 第1462-1467、
[10.1109/ICESC51422.2021.9532751](#)页 ↗

[在Scopus中查看](#) ↗ [谷歌学术](#) ↗

[赵等人, 2021年](#) Y。赵, Z.梁, J.杜, L。张, C。刘, L。赵

基于注意力的多头长短期记忆, 用于从言语中检测抑郁症

前面。神经机器人, 111 (2021) , [10.3389/fnbot.2021.684037](#) ↗

[谷歌学术](#) ↗

引用者 (55)

[基于MFCC和CNN生成的光谱图特征的深度学习模型用于检测抑郁症](#)

2024年, 生物医学信号处理和控制

引用摘录：

...Valstar等人[44]为DAIC-WOZ和Chen等人[51]为MODMA提出的模型仅对整个语音特征使用统计函数 (如平均值、中位数) , 这些函数对时间变化没有反应。相比之下, [24,46–49]等研究检索了语音的时间频率特征, 如Mel光谱图和MFCC, 其中包括动态信息, 并产生了更好的结果。拟议的模型通过从语音创作和感知过程中提取, 进一步提高了分类性能。...

[显示摘要](#) ✓

[基于图形神经网络的音频信号检测抑郁症的新研究](#)

2024年, 生物医学信号处理和控制

引用摘录：

... (i) 非图神经网络方法的比较：对于数据集DAIC-WOZ, 非图卷积方法的最佳分类性能之一是DEPA [40], F1得分为90%。与其他传统机器学习[17]和神经网络[36,38,39]比较方法相比, 性能提高了33.8%, 效果是除我们之外最好的。它说明了DEPA [40]通过训练编码器-解编码器模型来预测重建的光谱图中心段, 使其能够很好地学习给定光谱图中的上下文信息, 从而为下游分类任务获得音频嵌入的更有效的表

示。...

[显示摘要](#) ✓

[人工智能在青少年心理健康障碍的诊断、预后和治疗](#) ↗

2023年，公共卫生前沿

[在低资源环境中使用wav2vec 2.0的转移学习来改善言语抑制检测](#) ↗

2024年，科学报告

[KWHO-CNN：基于优化注意力驱动CNN的混合元启发式算法，用于自动临床抑郁症识别](#) ↗

2024年，国际计算与实验科学与工程杂志

[考虑我们需要的所有因素：具有不确定性近似的多模态抑郁分类](#) ↗

2023年，IEEE访问



[查看关于Scopus的所有引用文章](#) ↗

1 这些作者对这项工作做出了同等贡献，应被视为共同第一作者。

© 2022 作者。由爱思唯尔出版。

All content on this site: Copyright © 2026 Elsevier B.V., its licensors, and contributors. All rights are reserved, including those for text and data mining, AI training, and similar technologies. For all open access content, the relevant licensing terms apply.