arXiv:2601.07999v1 [cs.SD] 12 Jan 2026

VoxCog：通过方言知识实现端到端多语言认知障碍分类

TiantianFeng[a]，Anfeng Xu[a]，Jinkook Lee[b]，Shrikanth Narayanan[a,1]

[a]南加州大学明希电气与计算机工程系，美国加利福尼亚州洛杉矶市麦克林托克大道*3740*号，邮编*90089*[b]南加州大学多恩西夫文理学院，美国加利福尼亚州洛杉矶市，邮编*90089*

摘要

本研究通过整合能够明确识别语音方言的语音基础模型，提出了从言语中进行认知障碍分类的一种新视角。我们的研究动机源于这一观察：阿尔茨海默病（AD）或轻度认知障碍（MCI）患者往往会表现出可测量的言语特征，例如发音速率减慢、声音延长；这些特征与言语中可见的方言语音变异相似。基于这一理念，我们提出 VoxCog，这是一个端到端框架，利用预训练方言模型检测 AD 或 MCI，无需依赖文本或图像等额外模态。通过在多个用于检测 AD 和 MCI 的多语言数据集上开展实验，我们证明，在语音基础模型上以方言分类器进行模型初始化，能够持续提升对 AD 或 MCI 的预测性能。与先前采用不同信号模态集成多种计算方法的方案相比，我们训练的模型表现相当，且往往更优。尤其是，我们的端到端语音模型在 ADReSS 2020 挑战赛和 ADReSSo 2021 挑战赛测试集上的准确率分别达到 87.5% 和 85.9%，优于采用多模态集成计算或大语言模型的现有方案。

关键词：阿尔茨海默病、语音基础模型、语音分析、深度学习、方言建模

1. 引言

认知障碍是老年人口中一个重大且日益普遍的挑战。尤其是阿尔茨海默病（AD），它是老年人口中最常见的神经退行性疾病之一，也是导致痴呆的首要原因（Park 等人（2003）；Ritchie 等人（2017）；Livingston 等人（2020））。其特征是认知衰退逐渐加重、广泛的神经元退化，最终导致神经元死亡。除阿尔茨海默病外，轻度认知障碍（MCI）（Gauthier 等人（2006））是指认知衰退程度超过个体实际年龄所预期的水平。检测轻度认知障碍对痴呆的早期检测具有重要意义，因为超过 30% 的轻度认知障碍患者会在 5 年内发展为某种类型的痴呆。由于言语是一项基本的神经功能，涉及记忆、语言和运动计划，因此人们越来越关注开发基于言语的评估任务，以识别认知衰退和阿尔茨海默病的早期迹象。例如，DementiaBank（Lanzi 等人（2023））是一个广泛共享的资料库，提供认知障碍者和健康对照组的言语样本。它收录了参与标准化言语任务的数据，例如通过言语进行“偷饼干”图片描述或朗读任务。

尽管人们已付出大量努力来改进阿尔茨海默病的预测，但许多现有方法依赖于整合多种数据模态的知识（例如语音音频、语言），以取得更好的性能（Luz et al. (2021年, 2020)）。这增加了建模（例如，需要构建说话人分离和自动语音识别等构件）及结果解释的复杂性，领域专家也可能难以理解究竟是哪种模态驱动了预测。这些额外的计算阶段还会增加系统实现细节的复杂性，从而带来可复现性方面的挑战。受这些挑战启发，本文提出一个研究问题：端到端语音建模方法能否在不依赖自动语音识别（*ASR*）或语言建模等额外计算阶段的情况下，有效检测认知障碍？如果可以，哪些先验知识对这项任务有帮助？

为了解答这一问题，我们研究整合方言语音模式变异这一信息来源——它似乎与认知障碍预测正交——能否提高对轻度认知障碍或阿尔茨海默病的检测效果。我们利用方言知识的动机在于，存在认知障碍的个体在进行语音任务时可能会……

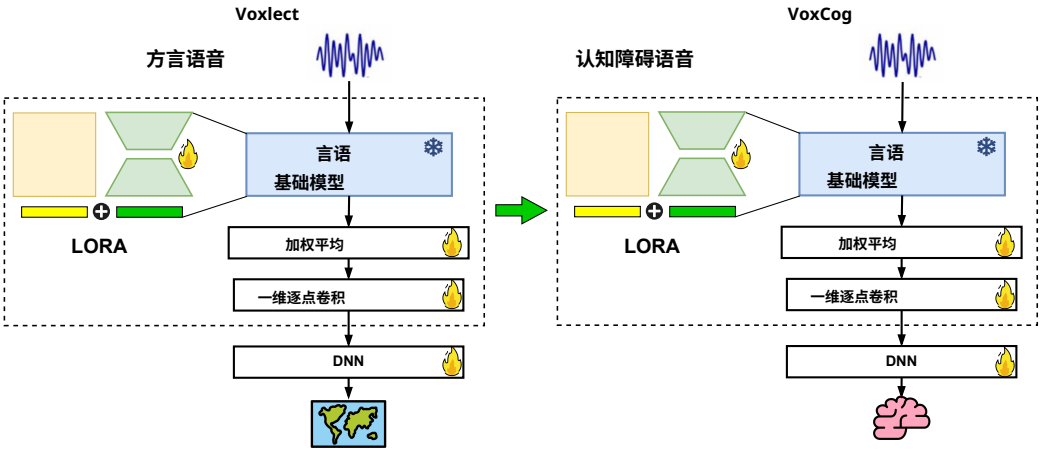

图1：VoxCog框架概览。我们使用预测相应方言所得的模型权重初始化VoxCog模型的训练。图中使用了来自OpenMoji网站的素材。

.....出现发音、语调和用词各异的言语，这些模式与方言建模中的发现相似。因此，基于这一观察，我们假设，具备方言变异先验知识的计算模型有助于识别与认知障碍相关的语音标记。有趣的是，实验结果表明，预训练的方言分类器为阿尔茨海默病和轻度认知障碍检测提供了强有力的先验，从而实现端到端的纯语音方法，减少了整合文本或图像等其他模态的需求。

具体而言，我们在本研究中的贡献如下：（1）我们提出了一种名为 VoxCog 的端到端建模方法，利用能够捕捉方言知识的预训练模型对认知障碍进行分类。（2）我们在多个广泛使用的认知障碍分类数据集上开展了大量实验，这些数据集涉及参与者执行各种语音任务。结果表明，在英语和非英语语料数据集中，VoxCog 的表现始终优于未显式建模方言信息的基线骨干网络。（3）我们还将 VoxCog 与现有文献中的代表性研究进行比较，结果表明，VoxCog 的表现具有竞争力，且在大多数情况下更胜一筹。

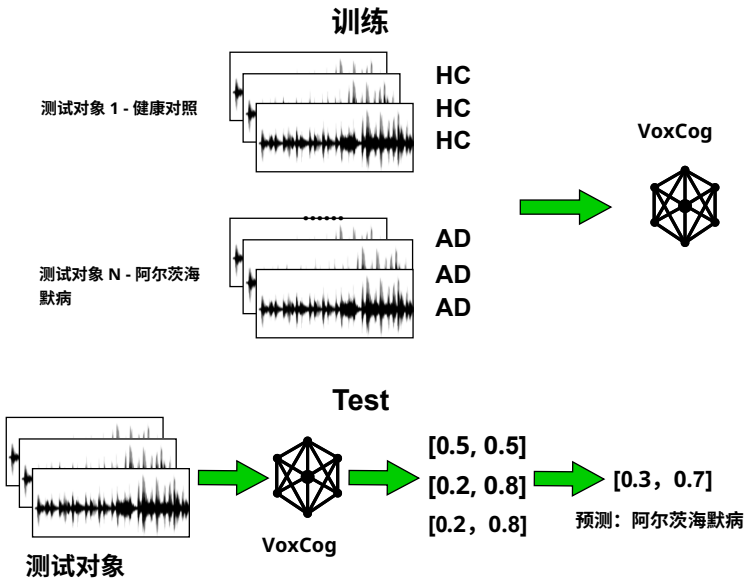

图2：VoxCog 中数据处理的概览。

2. VoxCog建模

本文介绍了 VoxCog，这是一种用于认知障碍分类的端到端多语言框架，它会显式利用从大规模方言语音数据中学习到的方言知识，如图 1 所示。该方法使用名为 Voxlect 的预训练方言识别模型（Feng 等人（2025b））的权重来初始化认知障碍分类模型。我们的目标是将方言表征迁移到涉及认知障碍分类的下游任务中。我们证明，VoxCog 仅使用原始语音信号，无需依赖转录文本或手工设计的副语言特征，即可直接达到最先进的性能。这些结果表明，方言感知语音表征可作为一种数据高效先验，用于跨语言分类认知健康衰退。

2.1. Voxlect

方言分类模型来自我们近期推出的 Voxlect 基准，该基准系统地评估语音基础模型预测世界多种语言中方言变异的能力。在这项工作中，

我们采用 Voxlect 模型来预测英语、西班牙语和普通话方言。训练方言模型时，我们使用预训练的 Whisper-Large（Radford 等人（2023））和 MMS-LID-256（Pratap 等人（2024））作为骨干模型。所有模型均采用低秩适配（LoRA）（Hu 等（2022））进行训练，并使用统一架构：先经过一维卷积神经网络，再经过 DNN 层，以完成方言分类。方言模型的详细信息如下：

2.1.1. 英语方言

具体而言，英语方言模型可识别全球 16 种口语方言，包括北美、英格兰、威尔士语、苏格兰语、北爱尔兰语、爱尔兰语、日耳曼语、罗曼语、斯拉夫语、闪米特语、大洋洲、南非、东南亚、东亚、南亚以及上述未提及的其他变体。该英语方言模型使用 10 多个数据集组合进行训练，包括 TIMIT（Garofolo 等（1993））、CommonVoice（Ardila 等（2020））、EdAcc（Sanabria 等（2023））、British Isles Speaker（Demirsahin 等（2020））、L2-Arctic（Zhao 等（2018））、VoxPopuli（Wang 等（2021））、Fair-Speech（Veliche 等（2024））、ESLTTS（Wang 等（2024））、Hispanic-English（Byrne 等（2014））和 ParaSpeechCaps（Diwan 等（2025））。由于短于 3 秒的语句无法提供足够的信息来可靠地表征方言，因此方言建模时排除了这些语句。

2.1.2. 西班牙语方言

西班牙语方言模型可预测 6 种地区方言：伊比利亚半岛方言、墨西哥方言、智利方言、安第斯—太平洋方言、中美洲和加勒比地区方言，以及拉普拉塔河方言。我们使用拉丁美洲西班牙语（Guevara-Rukoz 等（2020））和 CommonVoice-es（Ardila 等（2020））的数据集来训练西班牙语方言模型。

2.1.3. 汉语方言

中文方言模型采用 KeSpeech（Tang 等（2021年））提供的普通话方言标签。鉴于北京官话和东北官话与标准普通话具有共同的互通性，我们将这两种方言的样本归为标准普通话。除预测普通话方言外，我们还纳入了 CommonVoice（Ardila 等（2020））中的粤语样本，以扩大中文语言的覆盖范围。最后，中文方言模型预测标准普通话、冀鲁官话、西南官话

、江淮官话、兰银官话、中原官话、胶辽官话和粤语。

2.2. VoxCog

2.2.1. 数据准备

大多数用于认知障碍分类的基于语音的数据集，都是在受限实验环境中采集的，参与者需要完成预先设定的引导任务（例如图片描述）。因此，录音时长相对较短，通常约为 1–2 分钟。为适应这一录音条件，我们使用长度为 15 秒、步长为 5 秒的滑动窗口对每位受试者的录音进行分段，并将相应的认知障碍标签分配给每个片段，以用于训练。这样做是因为录音时长相对较短（约 1–2 分钟），其中包含足够信息的 15 秒言语片段可以用于认知障碍分类。在推理时，我们聚合片段级预测，以生成单个受试者的认知障碍预测结果。数据处理流程以及推理时的聚合过程概览见图2。

2.2.2. 模型架构

我们使用 Voxlect 预训练权重（Feng 等人（2025b））初始化模型，并通过 LORA模块和一维逐点卷积层进行适配；下游 DNN层则针对认知障碍分类进行随机初始化。在微调过程中，我们将预训练的 LORA模块和卷积层与 DNN分类器一同解冻，从而实现针对认知障碍分类的任务特定适配。

3. 数据集

我们使用 6 个基于语音的数据集来预测认知障碍（包括阿尔茨海默病或轻度认知障碍状态），具体如表1所述。除 2021-NCMMSC（Chen 等人（2023））外，所有数据集均从 DementiaBank（Lanzi 等人（2023））网站下载。

3.1. 2020 ADReSS 挑战数据集

我们的第一个实验数据集是 INTERSPEECH 2020 ADReSS挑战赛数据集（Luz 等人（2020））。该数据集的设计采用了

表1：本文所用数据集的组统计数据摘要。（HC：健康对照；MCI：轻度认知障碍；AD：阿尔茨海默病；DM：痴呆）

数据集	语言	组别	Task
ADReSS	英语	78 名健康对照，78 名阿尔茨海默病患者	描述任务
ADReSSo	英语	115 名健康对照，122 名阿尔茨海默病患者	描述任务
VAS	英语	30名健康对照，30名糖尿病患者	与 Alexa 交互
2021-NCMMSC	中文	44名健康对照，26名轻度认知障碍患者	描述任务
TAUKADIAL-zh	中文	43 名健康对照，44 名轻度认知障碍患者	描述任务
Ivanova	西班牙语	197 名健康对照、91 名轻度认知障碍、74 名阿尔茨海默病	阅读

阿尔茨海默病和健康对照（HC）受试者的均衡分布。训练集包含108名受试者，其中54名为阿尔茨海默病患者，54名为健康对照组成员。同样，测试集包含48名受试者，两个群体各占一半。言语数据采集自以英语为母语的受试者，他们根据 Cookie Theft 图片（Kaplan 等人（2001））完成描述任务。

3.2. 2021 ADReSSo 挑战赛数据集

本文使用的第二个数据集是 ADReSSo 挑战数据集，该数据集最初作为 INTERSPEECH ADReSSo 2021 挑战赛的一部分发布（Luz 等人（2021））。该完整数据集包括 237 名参与者的言语录音，其中 166 名受试者属于训练集，71 名受试者属于测试集。在这些参与者中，87 人被标记为阿尔茨海默病受试者，79 人为健康对照受试者。与 2020 ADReSS 挑战数据集类似，所有参与者均为英语母语者，并被要求使用 Cookie Theft 图片（Kaplan 等人（2001））完成图片描述任务。

3.3. VAS 数据集

我们的第三个实验数据集使用 VAS 语料库（Kurtz 等人（2023）），该语料库通过邀请参与者与 Alexa Echo 设备互动而收集。由于 COVID-19 疫情，本分析所用数据是远程收集的。在本文的实验中，我们重点对健康对照和阿尔茨海默病受试者进行分类。

3.4. 西班牙语 *Ivanova* 数据集

本文中我们尝试的第一个非英语数据集是西班牙语 Ivanova 数据集（Ivanova 等人（2022））。该数据集包含 361 位西班牙语母语者的语音录音。其中，74 人患有阿尔茨海默病，90 人患有轻度认知障碍，其余 197 人为健康对照受试者。所有受试者均被要求朗读塞万提斯《堂吉诃德》的前两句话。

3.5. 2021-NCMMSC 数据集

第二个非英语实验数据集使用 2021-NCMMSC 挑战赛数据集（ Chen 等人（2023））。在本实验中，我们重点关注从健康对照受试者中检测阿尔茨海默病患者，数据集包含 26 名阿尔茨海默病患者和 44 名健康对照受试者。大多数参与者是普通话母语者，并在录音过程中按要求完成图片描述、流利度测试和自由对话任务。

3.6. TAUKADIAL-zh 数据集

我们使用的最后一个数据集是 TAUKADIAL 挑战赛数据集，旨在对轻度认知障碍（MCI）患者和健康参与者进行分类。该数据集包含参与者完成图片描述等任务时的语音录音。参与者的母语为英语或中文。我们使用 Whisper-Large 推断每段录音所使用的语言，并选取数据集中的中文子集，在非英语数据评估中评估 VoxCog。

4. 实验

4.1. 数据划分

具体而言，我们在 2020 ADReSS、TAUKADIAL-zh 和 2021 NCMMSC 挑战赛数据集上进行 3 折交叉验证，在 2021年 ADReSSo 挑战赛数据集上进行 5 折交叉验证。对于 ADReSS、ADReSSo 和 NCMMSC 数据集，我们报告留出测试集上的宏平均 F1 分数和准确率的平均值。对于 TAUKADIAL 挑战赛数据集，我们计算非加权平均召回率（UAR）。这些指标是根据各自的挑战赛规则选定的。折数根据各数据集中的受试者数量通过启发式方法确定。此外，我们在西班牙语

表 2：英语数据集上 VoxCog 与基线方法在认知障碍分类方面的比较。报告了不同训练折和验证折上的宏平均 F1 和准确率均值。

数据集	预测	Whisper-Large		VoxCog-Whisper-Large	
		宏平均 F1	Acc	宏平均 F1	Acc
2020-ADReSS	健康对照 /阿尔茨海默病	0.801	80.55	**0.846**	**84.72**
2021-ADReSSo	健康对照 /阿尔茨海默病	0.763	76.61	**0.819**	**81.69**
VAS	健康对照 / 糖尿病	0.852	86.21	**0.913**	**91.51**
数据集	**预测**	**MMS-LID-256**		**VoxCog-MMS-LID-256**	
		宏平均 F1	Acc	宏平均 F1	Acc
2020-ADReSS	健康对照 /阿尔茨海默病	0.704	70.83	**0.705**	**70.83**
2021-ADReSSo	健康对照 /阿尔茨海默病	0.638	64.22	**0.696**	**69.85**
VAS	健康对照 / 糖尿病	0.896	89.84	**0.896**	**90.00**

Ivanova 数据集和 VAS 数据集上进行 5 折交叉验证，并报告各测试折上宏平均 F1 和准确率的平均值。对于所有实验，我们均报告受试者级预测分数。

4.2. 训练细节

在训练过程中，我们采用一系列数据增强方法，包括添加高斯噪声、背景噪声、时间拉伸和极性反转，以提升模型的鲁棒性。在实验中，我们对 Whisper-Large（Radford 等人（2023））和 MMS-LID-256（Pratap 等人（2024））模型的原始版本及其适配 Voxlect 的版本进行微调。所有实验最多训练 10 个 epoch，学习率从 [0.0001、0.0002、0.0005、0.001 和 0.002] 中选取。

5. 结果

5.1. VoxCog 在英语数据上能否优于基线模型？

我们首先将提出的 VoxCog 模型与未显式建模英语方言的基线模型进行比较，英语数据集上的结果汇总于表 2。总体而言，无论采用何种骨干模型，我们都观察到 VoxCog 始终优于相应的基线模型。例如，VoxCog–Whisper-Large 模型在 2020 ADReSS 挑战赛和 2021年 ADReSSo 挑战赛数据集上的平均准确率分别达到 84.72% 和 81.69%。相比之下，

表3：在非英语数据集上比较 VoxCog 与基线模型对认知障碍的分类效果。TAUKADIAL 报告非加权平均召回率，Ivanova 和 2021-NCMMSC 数据集使用 Macro-F1分数。

数据集	预测组	Whisper-Large	VoxCog Whisper-Large
2021-NCMMSC	健康对照 / 阿尔茨海默病	0.854	**0.940**
TAUKADIAL-zh	健康对照 / 轻度认知障碍	0.571	**0.603**
Ivanova	健康对照 / 轻度认知障碍	**0.631**	0.626
Ivanova	健康对照 / 阿尔茨海默病	0.715	**0.718**
数据集	**预测**	**MMS-LID-256**	**VoxCog MMS-LID-256**
2021-NCMMSC	健康对照 / 阿尔茨海默病	0.894	**0.952**
TAUKADIAL-zh	健康对照 / 轻度认知障碍	0.517	**0.520**
Ivanova	健康对照 / 轻度认知障碍	0.585	**0.634**
Ivanova	健康对照 / 阿尔茨海默病	0.694	**0.738**

基线 Whisper-Large 模型在同一数据集上的准确率仅为 80.55% 和 76.61%。这些结果表明，显式建模英语口语方言特征有助于提高认知障碍检测的准确率。

5.2. VoxCog 在非英语数据上的表现

除了在英语数据集上评估 VoxCog 外，我们还将 VoxCog 与基线模型在西班牙语和中文等非英语语音数据集上的表现进行了比较。具体而言，表3中的结果显示，在这些多语言场景下，VoxCog 的表现始终优于基线模型。例如，在 Ivanova 数据集上，VoxCog 在健康对照与阿尔茨海默病分类任务中取得了 0.738 的宏平均 F1 分数，高于基线模型的 0.694。同样，在 TAUKADIAL 数据集的中文子集上，与基线模型相比，VoxCog 将宏平均 F1 分数从 0.571 提高到了 0.603。这些结果表明，所提出的 VoxCog 能够有效地泛化到不同语言场景。

5.3. 将 *VoxCog* 与现有文献进行比较

我们进一步将所提出的VoxCog模型与文献中报道的代表性方法进行基准比较。为确保比较公平，我们将评估范围限定于ADReSS 2020和ADReSSo 2021挑战赛数据集，

表4：将 VoxCog 与现有文献中报道的代表性系统进行比较。所报告的数值基于在不同折上训练的模型预测结果的平均值。特别是，粗体表示各项指标的最佳结果，下划线表示次佳结果。

数据集	模型	宏平均 F1	准确率
2020-ADReSS	挑战基线（Luz 等人（2020））	0.745	75.00
	INESC-ID（Pompili 等人（2020））	0.808	81.25
	JHU x-vector（Pappagari 等人（2020））	0.745	75.00
	RMIT 系统（Syed 等人（2020））	-	<u>85.42</u>
	VoxCog（我们的方法）	**0.875**	**87.50**
2021-ADReSSo	挑战基线（Luz 等人（2021年））	0.789	78.87
	PPGs-BERT（Sun 等人（2025））	-	73.20
	Whisper-MJT（Jia 等人（2025））	-	74.65
	MUET-RMIT（Syed 等人（2021））	<u>0.845</u>	<u>84.51</u>
	WavBERT（Zhu 等（2021））	0.830	83.10
	CogBench（Feng 等（2025a））	0.661	66.48
	VoxCog（本文）	**0.859**	**85.92**

这两个数据集均提供标准化测试集，可确保跨研究比较。鉴于这些数据集具有基于挑战赛的特点，许多已发表的研究通常采用集成方法来优化性能。类似地，我们采用一种简单的集成方法，对使用不同交叉验证折训练的模型所产生的预测概率进行汇总。这使我们的评估与挑战赛文献中的常见做法保持一致，同时避免设计过于复杂的集成方法。

2020-ADReSS 表4中的结果显示，我们提出的 VoxCog 准确率达到 87.5%，与文献中报道的代表性方法相比具有竞争力，其中包括采用声学模型与基于语言的模型集成方案的挑战赛提交系统。尤其是，VoxCog 的表现优于大多数主要依赖声学表征的基线方法，例如基于 x-vector 的系统（Pappagari 等人（2020））。值得注意的是，VoxCog 是一个统一的端到端框架，只需输入原始语音信号即可，从而简化了训练和推理流程。

2021-ADReSSo 在规模更大的 2021-ADReSSo 数据集上，所提出的 VoxCog 模型也始终优于所有已报告的系统，即使其中许多系统在模型训

练过程中融合了语言表征。具体而言，VoxCog 优于 WavBERT（Zhu 等（2021）），后者融合了语义和非语义语言信息。值得注意的是，VoxCog 也优于 VoxBench（Feng 等（2025a））报告的大语言模型基线。这表明，VoxCog 模型无需额外计算（如自动语音识别或语言建模），即可有效达到与文献中顶尖解决方案相当的性能。

6. 结论

在本研究中，我们提出了 VoxCog，这是一种端到端建模方法，通过融入方言知识来提升基于言语的认知障碍预测能力。具体而言，VoxCog 建立在我们此前针对多种语言开展的大规模方言分类基准评测工作之上。我们在多种语言中广泛使用的多个基于言语的认知障碍分类数据集上评估了所提出的方法。实验结果表明，与未明确建模口语方言的基线模型相比，融入预训练方言知识能够持续提升性能。这从实证角度验证了我们的假设：方言数据包含某种语言系统性的语音、韵律和发音模式，有助于检测认知衰退者的异常言语行为；这些人的言语往往偏离正常言语模式。此外，所提出的端到端 VoxCog 框架取得了与文献中具有代表性的研究相当的性能。当前录音中的一项实际挑战是包含访谈者语音。在未来的工作中，我们计划整合稳健的说话人日志技术，在下游分类之前更准确地提取感兴趣语音片段，以减少潜在的混杂因素。

7. 致谢

T本研究由美国国家老龄研究所/美国国立卫生研究院（R01 AG080473）和美国国立卫生研究院（R01 AG051125）资助

参考文献

Ardila, R., Branson, M., Davis, K., Kohler, M., Meyer, J., Henretty, M., Morais, R., Saunders, L., Tyers, F., Weber, G., 2020. Common Voice：一个大规模多语言语音语料库，载于：第十二届语言资源与评估会议论文集，第 4218–4222 页。

Byrne, W., Knodt, E., Bernstein, J., Emami, F., 2014. 西班牙裔英语数据库（ldc2014s05）。语言数据联盟。

Chen, X., Zhang, W.Q., Ma, Y., 2023. 基于原始波形的端到端阿尔茨海默病检测方法。《电子学报》51, 3582– 3590.

Demirsahin, I., Kjartansson, O., Gutkin, A., Rivera, C., 2020. 不列颠群岛英语口音的开源多说话人语料库，载于：第十二届语言资源与评估会议论文集，第6532–6541页。

Diwan, A., Zheng, Z., Harwath, D., Choi, E., 2025. 扩展丰富风格提示的文本转语音数据集。arXiv预印本 arXiv:2503.04713。

Feng, R., Luo, Z., Wang, W., Song, Y., Liu, Y., Zhu, T., Li, J., Wang, X., 2025a. Cogbench：用于多语言言语认知障碍评估的大语言模型基准测试。arXiv 预印本 arXiv:2508.03360。Feng, T., Huang, K., Xu, A., Shi, X., Lertpetchpun, T., Lee, J., Lee, Y., Byrd, D., Narayanan, S., 2025b. Voxlect：用于对全球方言和区域语言建模的语音基础模型基准测试。arXiv 预印本 arXiv:2508.01691。Garofolo, J.S., Lamel, L.F., Fisher, W.M., Fiscus, J.G., Pallett, D.S., 1993. DARPA TIMIT 连续语音声学语音学语料库 CD-ROM。NIST 语音光盘 1-1.1。NASA STI/Recon 技术报告 n 93, 27403。Gauthier, S., Reisberg, B., Zaudig, M., Petersen, R.C., Ritchie, K., Broich, K., Belleville, S., Brodaty, H., Bennett, D., Chertkow, H., 等，2006. 轻度认知障碍。《柳叶刀》367, 1262–1270。Guevara-Rukoz, A., Demirsahin, I., He, F., Chu, S.H.C., Sarin, S., Pipatsri-sawat, K., Gutkin, A., Butryna, A., Kjartansson, O., 2020. 为低资源文本转语音众包拉丁美洲西班牙语，载于：第十二届语言资源与评估会议论文集，第 6504–6513 页。

Hu, E.J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L., Chen, W., 等，2022. LORA：大语言模型的低秩适配。ICLR 1, 3。

Ivanova, O., Meilán, J.J.G., Martínez-Sánchez, F., Martínez-Nicolás, I., Llorente, T.E., González, N.C., 2022。通过西班牙语朗读任务语料库评估阿尔茨海默病的言语特征并进行区分。《计算机语音与语言》73, 101341。Jia, K., Li, J., Li, K., Zhang, W.Q., 2025。基于 Whisper 的多语言阿尔茨海默病检测及针对低资源语言的改进，载于：Interspeech 2025会议论文集，第549–553页。Kaplan, E., Goodglass, H., Weintraub, S., 2001。波士顿命名测验。《临床神经心理学家》。Kurtz, E., Zhu, Y., Driesse, T., Tran, B., Batsis, J.A., Roth, R.M., Liang, X., 2023。使用语音助手指令早期检测认知衰退，载于：ICASSP 2023 IEEE国际声学、语音与信号处理会议，IEEE，第1–5页。Lanzi, A.M., Saylor, A.K., Fromm, D., Liu, H., MacWhinney, B., Cohen, M.L., 2023。DementiaBank：理论依据、方案及说明性分析。《美国言语语言病理学杂志》32，426–438。Livingston, G., Huntley, J., Sommerlad, A., Ames, D., Ballard, C., Banerjee, S., Brayne, C., Burns, A., Cohen-Mansfield, J., Cooper, C., 等，2020。痴呆预防、干预与照护：《柳叶刀》委员会2020年报告。《柳叶刀》396，413–446。Luz, S., Haider, F., Fromm, D., MacWhinney, B., 等，2020。通过自发语音进行阿尔茨海默病性痴呆识别：ADReSS挑战赛，载于：Interspeech 2020会议论文集，第2172–2176页。Luz, S., Haider, F., de la Fuente, S., Fromm, D., MacWhinney, B., 2021。仅通过语音检测认知衰退：ADReSSo挑战赛，载于：Interspeech 2021，第3780–3784页。doi:10.21437/Interspeech.2021-1220。Pappagari, R., Cho, J., Moro-Velázquez, L., Dehak, N., 2020。利用最先进的说话人识别和自然语言处理技术

检测阿尔茨海默病并评估其严重程度，载于：Interspeech 2020会议论文集，第2177–2181页。Park, H.L., O’Connell, J.E., Thomson, R.G., 2003。一般老年人群认知衰退的系统综述。《国际老年精神病学杂志》18，1121–1134。Pompili, A., Rolland, T., Abad, A., 2020。用于ADReSS 2020 挑战赛的 INESC-ID多模态系统，载于：Interspeech 2020会议论文集，第2202–2206页。Pratap, V., Tjandra, A., Shi, B., Tomasello, P., Babu, A., Kundu, S., Elkahky, A., Ni, Z., Vyas, A., Fazel-Zarandi, M., 等，2024。将语音技术扩展至1,000+种语言。《机器学习研究杂志》25，1–52。Radford, A., Kim, J.W., Xu, T., Brockman, G., McLeavey, C., Sutskever, I., 2023。通过大规模弱监督实现稳健的语音识别，载于：国际机器学习会议，PMLR，第28492–28518页。Ritchie, K., Carriere, I., Su, L., O’Brien, J.T., Lovestone, S., Wells, K., Ritchie, C.W., 2017。高迟发型阿尔茨海默病风险成年人群的中年认知特征：PREVENT研究。《阿尔茨海默病与痴呆》13，1089–1097。Sanabria, R., Bogoychev, N., Markl, N., Carmantini, A., Klejch, O., Bell, P., 2023。爱丁堡国际英语口音语料库：迈向英语自动语音识别民主化，载于：ICASSP 2023 IEEE国际声学、语音与信号处理会议，IEEE，第1–5页。Sun, Q., Qiu, Z., Pu, Y., Li, J., Chen, X., Zhang, W.Q., 2025。PPGs-BERT：利用音素序列和 BERT 从自发语音中检测阿尔茨海默病，载于：Interspeech 2025会议论文集，第554–558页。Syed, M.S.S., Syed, Z.S., Lech, M., Pirogova, E., 2020。通过自发语音自动筛查阿尔茨海默病性痴呆，载于：Interspeech 2020，第2222–2226页。doi:10.21437/Interspeech.2020-3158。Syed, Z.S., Syed, M.S.S., Lech, M., Pirogova, E., 2021。应对ADReSS挑战赛2021：用于从自发语音中识别阿尔茨海默病性痴呆的 MUET-RMIT系统，载于：Interspeech 2021会议论文集，第3815–3819页。

Tang, Z., Wang, D., Xu, Y., Sun, J., Lei, X., Zhao, S., Wen, C., Tan, X., Xie, C., Zhou, S., 等，2021年。KeSpeech：普通话及其八种次方言的开源语音数据集，载于：第三十五届神经信息处理系统大会数据集与基准赛道（第二轮）。Veliche, I.E., Huang, Z., Ayyat Kochaniyan, V., Peng, F., Kalinli, O., Seltzer, M.L., 2024年。迈向语音识别公平性的衡量：Fair-speech 数据集，载于：Interspeech 2024 论文集，第1385–1389页。Wang, C., Riviere, M., Lee, A., Wu, A., Talnikar, C., Haziza, D., Williamson, M., Pino, J., Dupoux, E., 2021年。VoxPopuli：用于表征学习、半监督学习和解释的大规模多语言语音语料库，载于：第59届计算语言学协会年会暨第11届自然语言处理国际联合会议论文集（第1卷：长篇论文），第993–1003页。Wang, W., Song, Y., Jha, S., 2024年。Usat：一种通用说话人自适应文本转语音方法。IEEE/ACM 音频、语音与语言处理汇刊。Zhao, G., Sonsaat, S., Silpachai, A., Lucic, I., Chukharev-Hudilainen, E., Levis, J., Gutierrez-Osuna, R., 2018年。L2-Arctic：非英语母语者语音语料库，载于：Interspeech 2018 论文集，第2783–2787页。Zhu, Y., Obyat, A., Liang, X., Batsis, J.A., Roth, R. M., 2021年。WavBERT：利用 wav2vec 和 BERT 的语义与非语义语音进行痴呆检测，载于：INTERSPEECH，第3790页。