**国内外研究现状：**

表1 主观评价方法

|  |  |
| --- | --- |
| 客观评价方法 | 描述 |
| 构音障碍检查方法 | 但该量表以主观描述为主，缺乏量化评分, 不利于定量统计分析。 |
| GRBAS分级评估 | GRBAS等级标准主要关注患者发音的清晰度，并未涉及到构音器官和构音运动的评估 |
| Frenchay构音障碍 | Frenchay量表包含反射、呼吸、唇的运动、颌的运动、软腭运动、喉的运动、舌的运动、以及语言的理解和掌握8个检查任务，每个项目中包含 2-6个子项目。 |

表2 客观评价方法

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 工作 | 数据库 | 特征集 | 分类器 | 效果 |
| Milton[1] | UASpeech | MFCCs GNE HNR | LDA | 95% |
| Kim M J[2] | QoLT 2011 | PE-SFCC | i-vector PLDA | 60.78% SID |
| Bhat[3] | UASpeech | Multitapered spectral estimation | ANN | 96.44% SD |
| TORGO | ANN | 98.7% SD |
| Vásquez-Correa[4] | 没有来源 | F0、共振峰、MFCC、i-vector | SVR | 相关系数0.63 |
| Liu[5] | CanPEV | GOP和88eGeMAPs | MLP | 三分类81% |
| Tulics[6] | From the National Institute of Oncology | ASR phoneme posterior features | SVM,DNN对比 | DNN>SVM  95% 85% |
| Tulics[7] | ASR PP是否有用 | SVN\DNN | ASR后验概率用处不大  88% 89% |
| Tripathi[8] | UASpeech | 384维IS09 | SVM | 53.9% SID  97.4% SD |
| H. M. Chandrashekar[9] | TORGO | Spectro-temporal features | CNN | 98.3% SD  49.27% SID |
| Mohammed[10] | SVD | MFCC | 迁移学习ResNet34 | 95.41% |
| Bhat[11] | TORGO | Fbank MFCC i-vectors | BLSTM  迁移学习 | 98.2% SD |
| SiddhantGupta[12] | UASpeech | Wave、偏移 | ResNet | 98.9% |
| Joshy[13] | UASpeech | MFCC\CQCC | CNN\DNN\LSTM\SVM | CNN>DNN>SCM>RF  i-VECTOR>MFCC>CQCC |
| TORGO |
| Vásquez-Correa[14] | 没有来源 | 多模态  Speech\gait\handwriting | CNN | 准确度97.6% |
| Biswajit[15] | (PC-GITA | 时频、Jitter、Shimmer等特征 | SVM | 89% |
| Biswajit[16] | PC-GITA | HCCs和MFCC | MLP | 82% |

1. 病变和

**现有的数据库：**

表 3 公开发布的构音障碍语音库

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 数据库 | 时长（小时） | 人数 | 词汇量 | 模态 |
| 粤语（CUDYS） | 10 | 16（11/5） | - | 音频+视频 |
| 荷兰语（EST） | 6.3 | 16（16/0） | - | 音频 |
| 英语（Nemours） | 2.5-3 | 11（11/0） | - | 音频 |
| 英语（UASpeech） | 102.7 | 29（16/13） | 455 | 音频+视频 |
| 英语（TORGO） | 15 | 14（7/7） | 1573 | 音频+视频 |
| 中文(MSDM) | **44.1** | **50(25/25)** |  | **音频+视频** |

|  |  |  |
| --- | --- | --- |
| **数据特点** | **描述** | **导致的问题** |
| 构音障碍语音数据量不足 | 现有大部分的构音障碍语音库规模较小，一般仅有10小时的语音数据，且被试人数20人以内。 | 构音障碍数据的稀疏性，模型容易过拟合 |
| 构音障碍语音数据多样性 | 不同构音障碍患者，由于说话人、口音、病因和病情程度等，造成语音多样性 | 类内差距比类间差距大得多，模型较难拟合 |
| 数据规范问题 | 现有的数据集数据标注没有统一标准，而且主观性强均没有统计标注正确率 | 无法剔除垃圾数据，无法进行科学的量化 |

**思考**

考虑到数据量少、数据差异大、数据规范等问题。得到准确而通用的分类\回归模型是一个挑战。

目前有几个初步想法：

1. 使用针对性的迁移学习，针对不同的发音使用不同的预训练模型。一方面解决数据量稀缺和数据差异大的问题，一方面有望提升多分类的准确性。
2. 使用多模态融合，针对一些不能使用语音特征解决的问题，比如唇部的运动、面部控制等。
3. 使用个性化自适应技术，解决说话人、口音、病因和病情程度等，带来的语音多样性问题。
4. 尝试更底层和有效的声学特征，比如宽带窄带频谱信息、小波变换等。

**引用**

[1] M. S. Paja and T. H. Falk, “Automated dysarthria severity classification for improved objective intelligibility assessment of spastic dysarthric speech,” 2012.

[2] M. J. Kim, J. Yoo, and H. Kim, “Dysarthric speech recognition using dysarthria-severity-dependent and speaker-adaptive models.,” in *Interspeech*, 2013, pp. 3622–3626.

[3] C. Bhat, B. Vachhani, and S. K. Kopparapu, “Automatic assessment of dysarthria severity level using audio descriptors,” in *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Mar. 2017, pp. 5070–5074. doi: 10.1109/ICASSP.2017.7953122.

[4] J. C. Vásquez-Correa, J. R. Orozco-Arroyave, T. Bocklet, and E. Nöth, “Towards an automatic evaluation of the dysarthria level of patients with Parkinson’s disease,” *J. Commun. Disord.*, vol. 76, pp. 21–36, Dec. 2018, doi: 10.1016/j.jcomdis.2018.08.002.

[5] Y. Liu, T. Lee, T. Law, and K. Y.-S. Lee, “Acoustical Assessment of Voice Disorder With Continuous Speech Using ASR Posterior Features,” *IEEEACM Trans. Audio Speech Lang. Process.*, vol. 27, no. 6, pp. 1047–1059, Jun. 2019, doi: 10.1109/TASLP.2019.2905778.

[6] M. G. Tulics, G. Szaszák, K. Mészáros, and K. Vicsi, “Artificial Neural Network and SVM based Voice Disorder Classification,” in *2019 10th IEEE International Conference on Cognitive Infocommunications (CogInfoCom)*, 2019, pp. 307–312. doi: 10.1109/CogInfoCom47531.2019.9089908.

[7] M. G. Tulics, G. Szaszák, K. Mészáros, and K. Vicsi, “Using ASR Posterior Probability and Acoustic Features for Voice Disorder Classification,” in *2020 11th IEEE International Conference on Cognitive Infocommunications (CogInfoCom)*, Sep. 2020, pp. 000155–000160. doi: 10.1109/CogInfoCom50765.2020.9237866.

[8] A. Tripathi, S. Bhosale, and S. K. Kopparapu, “Improved Speaker Independent Dysarthria Intelligibility Classification Using Deepspeech Posteriors,” in *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2020, pp. 6114–6118. doi: 10.1109/ICASSP40776.2020.9054492.

[9] H. M. Chandrashekar, V. Karjigi, and N. Sreedevi, “Spectro-Temporal Representation of Speech for Intelligibility Assessment of Dysarthria,” *IEEE J. Sel. Top. Signal Process.*, vol. 14, no. 2, pp. 390–399, 2020, doi: 10.1109/JSTSP.2019.2949912.

[10] M. A. Mohammed *et al.*, “Voice Pathology Detection and Classification Using Convolutional Neural Network Model,” *Appl. Sci.*, vol. 10, no. 11, Art. no. 11, Jan. 2020, doi: 10.3390/app10113723.

[11] C. Bhat and H. Strik, “Automatic Assessment of Sentence-Level Dysarthria Intelligibility Using BLSTM,” *IEEE J. Sel. Top. Signal Process.*, vol. 14, no. 2, pp. 322–330, 2020, doi: 10.1109/JSTSP.2020.2967652.

[12] S. Gupta *et al.*, “Residual Neural Network precisely quantifies dysarthria severity-level based on short-duration speech segments,” *Neural Netw.*, vol. 139, pp. 105–117, Jul. 2021, doi: 10.1016/j.neunet.2021.02.008.

[13] A. A. Joshy and R. Rajan, “Automated Dysarthria Severity Classification Using Deep Learning Frameworks,” in *2020 28th European Signal Processing Conference (EUSIPCO)*, 2021, pp. 116–120. doi: 10.23919/Eusipco47968.2020.9287741.

[14] J. C. Vasquez-Correa, T. Arias-Vergara, J. R. Orozco-Arroyave, B. Eskofier, J. Klucken, and E. Noth, “Multimodal Assessment of Parkinson’s Disease: A Deep Learning Approach,” *IEEE J. Biomed. Health Inform.*, vol. 23, no. 4, pp. 1618–1630, Jul. 2019, doi: 10.1109/JBHI.2018.2866873.

[15] B. Karan, S. S. Sahu, J. R. Orozco-Arroyave, and K. Mahto, “Non-negative matrix factorization-based time-frequency feature extraction of voice signal for Parkinson’s disease prediction,” *Comput. Speech Lang.*, vol. 69, p. 101216, Sep. 2021, doi: 10.1016/j.csl.2021.101216.

[16] B. Karan and S. Sekhar Sahu, “An improved framework for Parkinson’s disease prediction using Variational Mode Decomposition-Hilbert spectrum of speech signal,” *Biocybern. Biomed. Eng.*, vol. 41, no. 2, pp. 717–732, Apr. 2021, doi: 10.1016/j.bbe.2021.04.014.