

Overview of Automatic Speech Analysis and Technologies for Neurodegenerative Disorders: Diagnosis and Assistive Applications

Shakeel A. Sheikh, Md. Sahidullah, *Member, IEEE* and Ina Kodrasi, *Senior Member, IEEE*

Abstract—Advancements in spoken language technologies for neurodegenerative speech disorders are crucial for meeting both clinical and technological needs. This overview paper is vital for advancing the field, as it presents a comprehensive review of state-of-the-art methods in pathological speech detection, automatic speech recognition, pathological speech intelligibility enhancement, intelligibility and severity assessment, and data augmentation approaches for pathological speech. It also highlights key challenges, such as ensuring robustness, privacy, and interpretability. The paper concludes by exploring promising future directions, including the adoption of multimodal approaches and the integration of large language models to further advance speech technologies for neurodegenerative speech disorders.

Index Terms—Pathological speech, neurodegenerative speech disorders, speech processing, deep learning.

I. INTRODUCTION

SPEECH production is a complex mechanism that involves cognitive planning, coordinated muscle activity, and sound creation [1]. The process starts in the brain with the conceptualization of a message, followed by the organization of phonetic and prosodic plans, such as rhythm and style. The motor cortex then orchestrates the activation of approximately 100 muscles, allowing articulatory organs such as the tongue, lips, and jaw to shape the vocal tract and produce specific sounds. The initial sound is generated in the larynx, where the air from the lungs causes the vocal folds to vibrate. These phonatory structures adjust voice quality and prosody, while the articulatory organs further refine the sound by altering the shape of the vocal tract. Finally, the resulting speech is emitted through the oral and nasal cavities. Given the intricate coordination required for this process, any disruption of these finely tuned mechanisms can severely alter communication and result in pathological speech.

In this work, we define pathological speech as speech that deviates from neurotypical patterns due to underlying impairments. These deviations can manifest along multiple dimensions, including voice, articulation, prosody, and language. Voice impairments involve abnormalities in vocal fold vibration or in breath control, leading to hoarseness, breathiness, or a strained voice [2, 3]. Articulation impairments involve abnormalities in the coordination or movement of the various articulators, leading to slurred, imprecise, or segmented speech [4, 5]. Furthermore, prosodic impairments involve

abnormalities in the rhythm, stress, or intonation of speech, leading to speech that may sound flat or monotonic [6]. Finally, language impairments involve abnormalities in the formulation or comprehension of linguistic content, leading to difficulties with word retrieval, sentence construction, or understanding spoken or written language [7]. While pathological speech can arise from a wide range of neurological, structural, or functional impairments, our objective is to provide an overview of automated methods and speech-based technologies targeting pathological speech arising due to neurodegenerative disorders.

Neurodegenerative disorders such as Parkinson's disease (PD), Amyotrophic Lateral Sclerosis (ALS), or Alzheimer's disease are leading causes of voice, articulation, prosody, and language disruptions [4, 8, 9]. These disorders impair the brain regions and motor systems responsible for initiating, planning, and controlling the movements needed for speech production, resulting in a variety of speech disorders such as dysarthria, aphasia, apraxia of speech, or dysphonia [2, 4, 5, 7, 10–12]. Dysarthria and apraxia of speech, commonly seen in PD and ALS, are primarily characterized by articulatory and prosodic deficiencies [13], vowel distortions, reduced loudness variation, hypernasality, or syllabification [4, 5]. Dysphonia, also frequent in PD and ALS, is marked by abnormal voice quality such as hoarseness and breathiness [2, 3]. In contrast, aphasia typically presents as difficulties with word-finding and sentence construction, and is most commonly associated with Alzheimer's disease or other forms of dementia [7].

As the population grows and ages, the prevalence of neurological disorders, and consequently of various speech disorders, rapidly increases. In 2019, the World Health Organization estimated that over 8.5 million people worldwide were living with PD [14], up from 6.1 million in 2016 [15] and 2.5 million in 1990 [15]. By 2040, this figure is projected to surpass 17 million [15]. Similarly, dementia affected more than 46 million people globally in 2015 and this figure is expected to rise to 131.5 million by 2050 [16]. The prevalence of ALS is also growing significantly, with cases anticipated to increase by nearly 70% between 2015 and 2040 [17]. This increasing prevalence of neurological disorders, and consequently of the associated speech disorders, underscores the need to prioritize speech disorders both in the context of clinical practice as well as in the context of speech-based technologies.

Accurately diagnosing the presence of speech disorders in clinical practice (i.e., distinguishing between neurotypical and impaired speech) is crucial, as the presence of such disorders may serve as an early indicator of neurodegenerative conditions [18–20]. Further, an accurate differential diagnosis (e.g., discriminating between dysarthria and apraxia of

Shakeel A. Sheikh and Ina Kodrasi are with Idiap Research Institute, Switzerland (e-mail: {shakeel.sheikh, ina.kodrasi}@idiap.ch).

Md Sahidullah is with TCG CREST, Kolkata, India (e-mail: md.sahidullah@tcgcrest.org).

This work was supported by the Swiss National Science Foundation project no CRSII5_202228 on "Characterisation of motor speech disorders and processes".

speech) can provide important clues about the underlying neuropathology [21, 22]. Monitoring speech characteristics such as severity and intelligibility after diagnosis is also essential for tracking disease progression and evaluating the effectiveness of speech therapy interventions over time [6, 8, 12, 23].

Speech assessment in clinical practice relies on established perceptual evaluation scales that serve as gold standards for diagnosing and characterizing various aspects of speech impairments. For example, the GRBAS scale [24, 25] evaluates voice quality through parameters like grade, roughness, breathiness, asthenia, and strain. The Consensus Auditory Perceptual Evaluation-Voice (CAPE-V) [23] offers a similar perceptual framework, excluding asthenia. In addition to these general assessment scales, specific scales are employed for specific conditions. For instance, the Unified Parkinson's Disease Rating Scale (UPDRS) [26] includes components for speech and motor function assessment in PD, while the Bogenhausen Dysarthria Scales (BoDyS) [27] focus on the severity and profile of dysarthric impairments. Intelligibility, a key outcome measure in many disorders, is often assessed using tools such as the Assessment of Intelligibility of Dysarthric Speech [28], which standardizes both single-word and sentence intelligibility evaluations.

Traditionally, clinicians conduct these evaluations through costly and time-consuming auditory-perceptual assessments [12, 23], as illustrated in Fig. 1. Diagnosing speech disorders and distinguishing between various conditions can be particularly challenging, even for experienced clinicians. This difficulty arises from the subtle nature of clinical-perceptual characteristics which are often hard to detect by ear, especially in cases of mild impairments. The overlapping characteristics of certain speech disorders, such as dysarthria and apraxia of speech, further complicate the process. Consequently, inter-rater agreement for (differential) diagnosis of speech disorders among clinicians can be low [29, 30].

These clinical challenges highlight the growing need for complementary, technology-driven approaches to support diagnosis, monitoring, and intervention. In parallel, the increasing prevalence of speech disorders poses significant accessibility challenges in patients' everyday interactions with speech-based technologies. For example, individuals with dysarthria or apraxia of speech often experience difficulty using mainstream virtual assistants such as Cortana, Alexa, and Siri [31]. As speech disorders become more prevalent with neurodegenerative conditions, it is critical to prioritize both the diagnosis and treatment of these disorders in clinical practice, while also ensuring that patients with such impairments have equitable access to speech-based technologies. Addressing these barriers could lessen the burden on the health care system and significantly improve the patients' quality of life and their ability to engage with everyday digital tools.

Aiming to assist the clinical diagnosis and treatment of patients suffering from neurodegenerative disorders, there has been a growing interest in the research community to develop automated and objective methods for pathological speech analysis. A schematic illustration of such analysis is depicted in Fig. 1. These advanced technologies are designed to minimize bias, enhance diagnostic accuracy, and stream-

line the assessment process, ensuring greater efficiency and consistency. Clinicians can then use insights provided by the automatic models to organize therapeutic sessions accordingly, ensuring that the treatment addresses the specific impairments identified. Further, clinicians may perform additional manual acoustic analysis to gain further insights into the patient's speech patterns and impairments, providing complementary information to the automatic model's decision for a comprehensive therapeutic approach. Besides the clinical domain, efforts have been directed towards developing various speech-based technological applications aimed at pathological speakers, such as automatic speech recognition systems (ASRs) [32], speech synthesis systems [33–36] or intelligibility enhancement solutions [37–40]. To our knowledge, there is currently no comprehensive survey paper that discusses the research directions and challenges of this area both from a clinical and a technological perspective. Although, there have been related works, such as [41–43], they are primarily focused on specific characteristics, applications, or disorders, limiting their scope. For instance, [44] provides an overview of acoustic-articulatory characteristics in neurodegenerative disorders. Other studies, such as [45–50], mainly focus their review on the discrimination between neurotypical and impaired speech. Gupta et al. [51] discussed some of the wider challenges faced in the pathological speech domain. However, this work addresses only a limited set of challenges and is now somewhat outdated given the rapid advancements in the field over the past decade. In contrast, as depicted in Fig. 2, our work aims to fill this gap by providing an extensive review of the field encompassing pathological speech from various clinical and technological perspectives, such as automatic discrimination between neurotypical speech and speech disorders, ASR systems for pathological target speakers, enhancement systems aiming to enhance the intelligibility of pathological speech, severity and intelligibility assessment, and data augmentation approaches.

The remainder of the paper is organized as follows. Section II describes various pathological datasets employed in the literature. Section III provides a high level overview of the various speech representations used in pathological speech analysis. Section IV describes the different approaches to automatic pathological speech detection. Section V examines pathological speech in the context of ASR systems. Section VI discusses speech enhancement techniques aimed at improving pathological speech intelligibility. Section VII describes approaches for automatically estimating the intelligibility and severity of pathological speakers. Section VIII summarizes the various data augmentation methods used in automatic pathological speech systems. Finally, Section IX presents challenges and promising future research directions in the field. In summary, our contributions are the following:

- 1) We present the first comprehensive survey of automatic approaches for pathological speech from a clinical and technological perspectives ranging from detection, recognition, enhancement, and assessment.
- 2) We highlight current limitations in the field and propose several promising future directions for research in

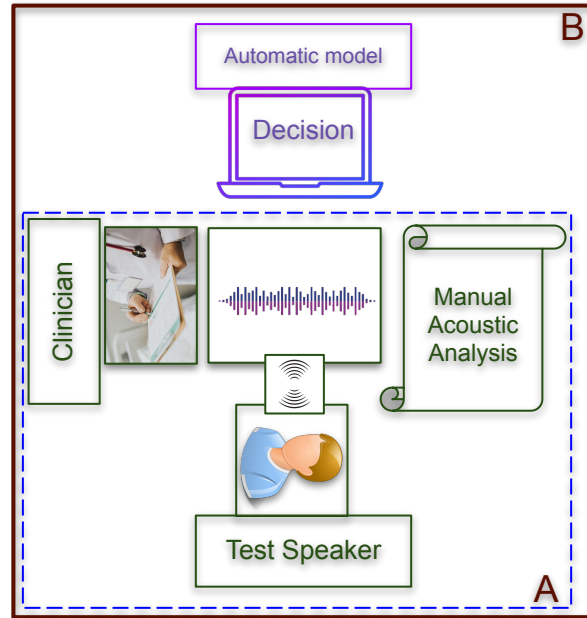


Fig. 1. Traditional auditory-perceptual assessment in clinical practice (bounded by the dashed box A) and automatic pathological speech analysis system (bounded by the solid box B). The clinician listens to the (potential) patient and assesses by ear the various characteristics of the speech. The automatic model is trained to detect and analyze speech impairments. Clinicians may use the insights provided by the automatic model to organize therapeutic sessions accordingly. Additionally, they may perform manual acoustic analysis to gain further insights into the patient's speech patterns and impairments, providing complementary information to the automatic model's decision.

pathological speech processing.

II. PATHOLOGICAL SPEECH DATASETS

Datasets for pathological speech research are scarce due to several inherent challenges in data collection. One of the primary difficulties is the sensitive nature of the population involved. Recruiting participants suffering from neurological disorders requires careful consideration of ethical and privacy concerns, as well as navigating the potential stigma associated with such disorders. Additionally, the physical and cognitive challenges faced by these individuals can complicate the process of obtaining high-quality speech recordings. Another key challenge is the variability in pathological speech patterns. Speech impairments can manifest in numerous ways and may vary widely across individuals, even within the same diagnostic category. This variability makes it difficult to create standardized protocols for data collection that ensure consistency and relevance across samples. Moreover, speech impairments may fluctuate over time, further complicating the process of capturing representative speech samples. When collecting datasets for pathological speech research, it is essential to prioritize inclusivity, i.e., ensuring a diverse representation of different speech disorders, age groups, and demographics. Additionally, data collection protocols should consider the comfort and cooperation of participants. Ethical considerations must also be at the forefront, ensuring informed consent and the protection of participant privacy. Lastly, it is important to design flexible and scalable collection methods that can capture a range of speech characteristics while maintaining consistent quality across diverse individuals and conditions. In the remainder of this section, we briefly review pathological

datasets¹ commonly used in the literature. The summary of these datasets and their characteristics is presented in Table I.

- *TORG* [52]. The TORG dataset contains English (spontaneous and read) speech recordings from control speakers and patients and the corresponding three-dimensional electromagnetic articulography (EMA). The patients suffer from ALS or Cerebral Palsy (CP). The dataset consists of recordings from 7 (3 female, 4 male) control speakers and 8 (3 female, 5 male) patients.
- *PC-GITA* [53]. The PC-GITA dataset contains Spanish (spontaneous and read) speech recordings from 50 control speakers and 50 patients suffering from PD. The two groups of speakers are age- and sex-matched, with 25 male and 25 female speakers in each group.
- *MoSpeDi* [54]. The MoSpeDi dataset contains French (spontaneous and read) speech recordings from 466 control speakers and 138 patients suffering from various types of motor speech disorders such as dysarthria or apraxia of speech. While subgroups of age- and sex-matched controls and patients can be found within the database, the overall dataset is not age- and sex-matched.
- *Nemours* [55]. The Nemours dataset is an English dataset containing (read) speech recordings from 11 male speakers with varying degrees of dysarthria severity.
- *CUDYS* [56]. The Cantonese Dysarthric Speech Corpus contains Cantonese (read) speech recordings from 5 control speakers and 11 patients suffering from cerebellar degeneration. The control group consists of 2 female and

¹Due to restricted access to many datasets, we are unable to provide complete metadata, including details like the number of male and female speakers

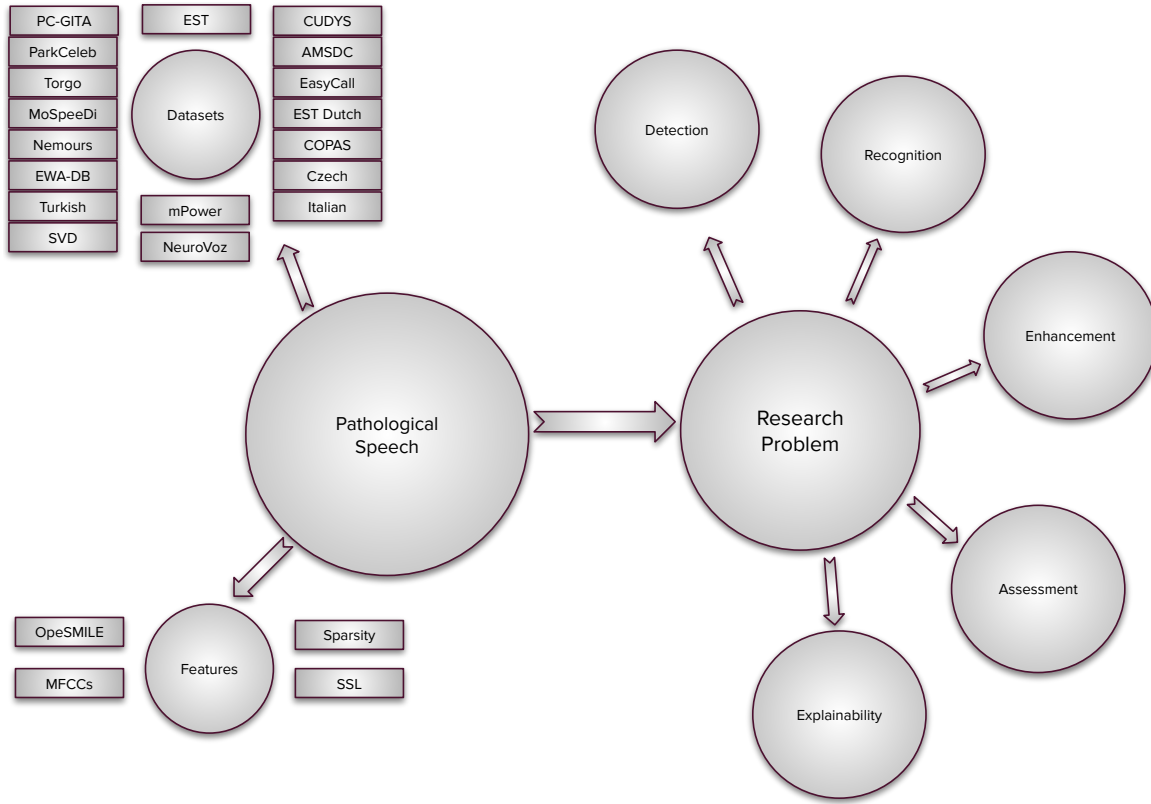


Fig. 2. Overview of key components discussed in this manuscript, including datasets, features, and research directions for pathological speech.

3 male speakers, whereas the dysarthric group consists of 5 female and 6 male speakers.

- *AMSDC* [57]. The Atlanta Motor Speech Disorders Corpus is an English dataset containing (spontaneous and read) speech recordings from 99 (62 male, 37 female) patients diagnosed with various disorders such as PD, ALS, or dementia.
- *EST* [58]. The EST dataset is a Dutch dataset containing (read) speech recordings from 16 male dysarthric patients due to PD, traumatic brain injuries, or cerebrovascular accident.
- *EasyCall* [59]. The EasyCall dataset is an Italian dataset containing (read) speech recordings from 24 controls and 31 patients diagnosed with PD, Huntington's disease, ALS, and peripheral neuropathy. The control group consists of 10 female and 14 male speakers, whereas the patient group consists of 11 female and 20 male speakers.
- *COPAS* [60]. The Corpus of Pathological and Normal Speech dataset is a Dutch dataset containing (spontaneous and read) speech recordings from 197 pathological speakers and 122 control speakers. The database comprises 8 distinct pathological categories such as dysarthria, hearing impairment, cleft, etc.
- *ParkCeleb* [61]. The previously reviewed datasets are not longitudinal and do not allow tracking the progression of the speech disorder within the same patient along time. To address this gap, the English ParkCeleb dataset was recently introduced in [61]. This dataset contains (sponta-

neous) audio-visual recordings (such as studio interviews or press conferences) from 40 (2 female, 38 male) control speakers and 40 (2 female, 38 male) patients suffering from PD.

- *Italian Parkinson's Database* [62]. The dataset contains Italian (read) speech recordings from a total of 65 speakers, including 28 patients with PD and 37 control speakers. Sex distribution is non-uniform across groups, with the PD group consisting of 19 male and 9 female speakers and the control group consisting of 23 male and 14 female speakers.
- *NeuroVoz* [63]. The NeuroVoz dataset contains (spontaneous, read, and listen and repeat) speech recordings from 112 Castilian Spanish speakers, including 54 patients diagnosed with PD and 58 control speakers. The control group consists of 28 male speakers, 26 female speakers, and 1 speaker whose sex information is not provided, whereas the patient group consists of 33 male and 20 female speakers.
- *Saarbrücken Voice Database* [64]². The Saarbrücken Voice Database contains (read) speech recordings and electroglottography data from 2,255 German speakers, including 1,356 patients and 869 control speakers [65]. Patients are diagnosed with various voice disorders. The control group consists of 433 male and 436 female speakers, whereas the patient group consists of 629 male

²<https://stimmdb.coli.uni-saarland.de/> (accessed July 07, 2025)

TABLE I

OVERVIEW OF PATHOLOGICAL SPEECH DATASETS (PD: PARKINSON'S DISEASE, CVA: CEREBROVASCULAR ACCIDENT, TBI: TRAUMATIC BRAIN INJURIES, ALS: AMYOTROPHIC LATERAL SCLEROSIS, HUNTINGTON'S DISEASE, AD: ALZHEIMER'S DISEASE, MCI: MILD COGNITIVE IMPAIRMENT, EMA: ELECTROMAGNETIC ARTICULOGRAPHY, EGG: ELECTROGLOTTOGRAPHY, L: LONGITUDINAL DATA). PUBLIC DATASETS ARE OPENLY AVAILABLE ON THE WEB AND CAN BE DOWNLOADED WITHOUT THE NEED FOR EXPLICIT APPROVAL OR SIGNING AN AGREEMENT. ACCESSIBLE DATASETS REQUIRE AN APPLICATION PROCESS OR SIGNING AN AGREEMENT, BUT CAN BE OBTAINED BY THE RESEARCH COMMUNITY UNDER DEFINED CONDITIONS. NON-ACCESSIBLE DATASETS ARE PRIVATE AND NOT AVAILABLE TO THE WIDER RESEARCH COMMUNITY.

Database	Number of Speakers		Type of Impairment	Language	Modality	Accessibility
	Control	Pathological				
PC-GITA	50	50	PD	Spanish	Speech	Accessible
TORG	07	08	ALS/CP	English	Speech + EMA	Public
MoSpeedi	466	138	Dysarthria/Apraxia	French	Speech	Not Accessible
Nemours	-	11	Dysarthria	English	Speech	Not Accessible
CUDYS	05	11	Spino-cerebellar ataxia	Cantonese	Speech	Not Accessible
AMSDC	62	37	CVA, PD	English	Speech	Not Accessible
Dutch EST	-	16	Dysarthria, TBI	Dutch	Speech	Not Accessible
EasyCall	24	31	PD, HD, ALS	Italian	Speech	Public
Saarbrücken Voice Database	869	1,356	Pathological	German	Speech + EGG	Public
Turkish Parkinson Dataset	20	20	PD	Turkish	Speech	Not Accessible
Czech Parkinson's Dataset	22	61	PD	Czech	Speech	Public
Italian Parkinson's Database	28	37	PD	Italian	Speech	Public
NeuroVoz	58	54	PD	Spanish	Speech	Public
COPAS	197	122	Dysarthria and others	Dutch	Speech	Accessible
EWA-DB	896	226	PD, AD, MCI	Slovak	Speech	Accessible
ParkCeleb	40	40	PD	English	Speech + Video + L	Accessible
mPower	5,581	1,087	PD	English	Multimodal	Accessible

and 727 female speakers.

- *Turkish Parkinson Speech Dataset* [66]. This dataset contains (read) speech recordings from 40 Turkish speakers, including 20 controls and 20 patients with PD. The PD group contains 6 female and 14 male speakers, whereas the control group contains 10 female and 10 male speakers.
- *Czech Parkinson's Dataset* [67]. This dataset contains (read) speech recordings from 83 Czech speakers, including 22 control speakers and 61 patients diagnosed with PD or atypical parkinsonian syndromes. The patient group consists of 30 female and 31 male speakers, whereas the control group consists of 11 female and 11 male speakers.
- *EWA-DB* [68]. The EWA-DB dataset consists of (spontaneous and read) speech recordings of 1,122 Slovak speakers, including 896 control speakers and 226 patients diagnosed with PD, mild cognitive impairment, or Alzheimer's disease. The patient group consists of 121 male and 105 female speakers, whereas the control group consists of 248 male and 648 female speakers.
- *mPower Parkinson's Dataset* [69]. The mPower study collected multimodal smartphone sensor data from 6805 participants, including 1,087 patients diagnosed with PD and 5,581 control speakers. Patient and control status were self-reported. The dataset was collected through the mPower mobile application and includes four distinct sensor-based assessment modalities, i.e., spatial memory evaluation, gait analysis through walking tasks, manual dexterity measurement via finger tapping, and vocal function assessment using sustained phonation recordings.

As presented in Table I, a considerable number of the datasets used in the literature are private and not openly accessible. This limits their availability for broader research and

replication efforts. Although some datasets such as TORG are publicly available, they suffer from detrimental recording artifacts [70], which compromise their usefulness for studying pathological speech characteristics and for developing reliable automatic detection methods [70, 71]. The lack of accessible, high-quality data poses a major challenge to the development and validation of effective algorithms, slowing progress in understanding and addressing pathological speech conditions. Ensuring the availability of clean, reliable, and comprehensive datasets is therefore essential for advancing research and practical applications.

Another critical limitation is the lack of linguistic diversity. Existing datasets are available only in a few languages, which restricts their applicability to non-represented populations and limits the potential for cross-linguistic comparisons. This linguistic bias poses a challenge for developing universal models that can generalize across languages.

In addition to these issues, most datasets are relatively small, often including only a limited number of participants. This scarcity hinders the development of robust and generalizable models and makes it difficult to capture the wide variability in speech disorders across different individuals and conditions. Furthermore, demographic imbalances are common. Many datasets exhibit skewed gender representation, which may introduce biases into trained models. In some cases, age groups or types of pathological conditions are also underrepresented, further limiting generalizability.

These limitations underscore the pressing need for larger, more diverse, and better-balanced pathological speech datasets that reflect the complexity of real-world populations. Addressing these gaps is essential for building reliable and inclusive tools for pathological speech analysis.

III. SPEECH REPRESENTATIONS FOR

AUTOMATIC APPROACHES

Pathological speech exhibits a range of acoustic anomalies, including deviations in pitch, loudness, vowel space reduction, and articulation [4, 72]. Additionally, it can lead to asymmetrical tension in the vocal folds, resulting in irregular vibrations and, consequently, an abnormal fundamental frequency. Patients can also show inconsistent rhythmic structures in comparison to control groups [73]. Excessively high or low fundamental frequency, combined with excessive vocal intensity, can exacerbate the severity of pathological voice conditions, producing characteristics such as shrillness, screechiness, hoarseness, or huskiness [74].

To capture such abnormalities as biomarkers for automatic pathological speech processing, researchers have employed various handcrafted acoustic features such as OpenSMILE [70], Mel-frequency cepstral coefficients (MFCCs) [75], or spectro-temporal sparsity features [76]. Raw speech signals and various time-frequency representations such as the short-time Fourier transform (STFT) have also been directly exploited in combination with deep learning approaches to directly learn pathology-discriminant cues [77, 78]. More recently, latent embeddings derived from self-supervised models have been used as more powerful representations of speech patterns and the various impairments [70, 79, 80]. It is important to note that a considerably larger set of features and representations have been employed to characterize disorders across various pathologies than the ones outlined above. Here, we briefly discuss the representations that are most commonly used in the literature.

A. Handcrafted Acoustic Features

OpenSMILE: Among the various features used in pathological speech detection, OpenSMILE features have been widely explored in the literature in conjunction with traditional machine learning algorithms [40, 70, 81–92]. The OpenSMILE feature set includes a 6552-dimensional feature vector that primarily consists of low-level audio features such as CHROMA, CENS, loudness, MFCCs, and other spectral features. In the context of pathological speech, these features can capture the subtle anomalies in speech patterns that are indicative of disorders. OpenSMILE’s comprehensive feature set allows for the analysis of prosodic elements like pitch, jitter, shimmer, and formant frequencies, which are crucial for identifying and differentiating various speech pathologies. However, OpenSMILE features are general features that have been used for a variety of speech applications such as emotion recognition [93] or speech recognition [94] and they are not specifically handcrafted to capture pathological cues..

Spectral and cepstral coefficients: Due to their ability to characterize articulation deficiencies, various spectral and cepstral coefficients such as Linear Predictive Coding (LPC), MFCCs, and Perceptual Linear Prediction (PLP) features, have been successfully used for pathological speech detection [13, 95–104]. LPC features describe the distribution of energy across frequency bands and are directly related to the resonant properties of the vocal tract, making them suitable for analyzing articulatory features like formant frequencies

and tongue positioning. In contrast, MFCCs and PLP features are obtained by transforming the spectral envelope into the cepstral domain, which effectively separates source and filter components while capturing smoothed representations of the vocal tract shape. When extended with temporal dynamics, these features can track transitions and instability in articulation patterns. To enhance robustness against channel variability and noise, the Relative Spectral Transform PLP features have also been considered in [105, 106]. Furthermore, the fusion of modulation spectra features or glottal features with MFCCs has additionally been explored to further improve performance [107, 108].

Spectro-temporal sparsity: Since pathological speech can be breathy, semi-whispery, and is characterized by abnormal pauses and imprecise articulation, it can be expected that its spectro-temporal sparsity differs from the spectro-temporal sparsity of neurotypical speech. To characterize spectro-temporal sparsity, various sparsity-based features have been introduced in [76, 109]. Although such features have been shown to be discriminative of various speech disorders [76, 86, 92, 109], they are highly sensitive to environmental artefacts such as noise and reverberation.

B. Time-Frequency Representations

Input representations such as the STFT and its variants allow for the analysis of speech signals in both time and frequency domains, providing valuable insights into speech characteristics such as pitch, formant shifts, and spectral irregularities. The STFT is often employed due to its ability to capture dynamic changes in the signal over time, which is critical for identifying variations in speech patterns linked to speech disorders [34, 110–121]. Furthermore, wavelet transforms and the Continuous Wavelet Transform offer better time-frequency localization, which is especially beneficial for analyzing transient and non-stationary features of pathological speech [122].

C. Raw Waveform Representation

While handcrafted acoustic features and time-frequency representations have demonstrated strong performance, researchers have also explored end-to-end pathological speech detection using raw input representations, eliminating the need for handcrafted features or time-frequency representations. Studies such as [78, 123, 124] have shown that raw waveform-based methods can capture discriminative pathological patterns and outperform traditional feature-based approaches. However, these methods may require more training data compared to other approaches for a robust performance.

D. Self Supervised Embeddings

Even though handcrafted acoustic features, time-frequency, and raw waveform input representations have achieved promising results in the analysis of pathological speech, their performance remains limited. The current state-of-the-art input representations for pathological speech primarily consist of latent embeddings derived from self-supervised models such as wav2vec2 [125], HuBERT [126], or WaveLM [127]. Research

in e.g., [79, 92, 128–151] has shown that these advanced models are effective in capturing complex speech patterns and nuances, leading to improved performance in tasks like detection and recognition. However, ongoing research continues to refine these embeddings and assess their robustness and generalizability across diverse pathological speech conditions, tasks, and applications.

IV. AUTOMATIC PATHOLOGICAL SPEECH DETECTION

As discussed in Section I, clinicians traditionally rely on laborious and time-consuming auditory-perceptual measures to diagnose speech impairments accurately. To address the various challenges associated with auditory-perceptual assessments, there has been a growing interest in the research community to develop automatic approaches for diagnosing pathological speech. The goal of these approaches is to enhance the detection accuracy to match or surpass human accuracy, thereby ensuring more consistent and reliable diagnosis. Automated systems analyze speech patterns and identify anomalies indicative of these neurodegenerative conditions by leveraging classical machine learning and deep learning models with input representations such as the ones discussed in Section III. In the following, we briefly summarize various classical machine learning-based and deep learning-based approaches.

A. Classical Machine Learning-based Approaches

In recent years, numerous studies have investigated the use of classical machine learning classifiers combined with hand-crafted acoustic features for detecting pathological speech, yielding promising results. For instance, [152] and [153] employed Random Forests and Support Vector Machines (SVMs) on a range of dysphonia measures, achieving strong classification performance on relatively small speaker datasets. Similarly, [154] applied Gaussian Mixture Models and Hidden Markov Models to cepstral features for detecting dysarthric speech in an Indian Tamil language dataset. Orozco-Arroyave et al. [155] further demonstrated that cepstral coefficients are effective and robust when used with SVMs for detecting dysarthric speech in a Spanish dataset. While these findings are encouraging, their scalability to larger, more diverse datasets remains uncertain due to the variability of speech pathologies across populations. Additionally, despite the effectiveness of various spectral and cepstral features, the field still lacks consensus on which features are the most discriminative and generalizable across different conditions and datasets.

Recognizing the importance of developing discriminative feature representations, the research community has devoted significant efforts to this area. For example, [156] investigated the use of articulatory features for pathological speech detection and achieved promising results on a small dataset of 24 Czech speakers. However, the limited sample size restricts the generalizability of the findings, and extracting articulatory features at scale remains technically challenging. In a more comprehensive effort, [157] utilized an extensive feature set including 6,373 acoustic, 3,600 articulatory, and 4 sensory features to detect ALS, reporting state-of-the-art performance

on a dataset of 22 speakers. Norel et al. [158] employed SVMs with openSMILE features to detect pathological speech in a larger cohort of 123 Hebrew speakers. Prabhakera and Alku [82] found that incorporating glottal features alongside openSMILE features improved dysarthric speech detection performance. Further advancements were made in [76] by introducing spectro-temporal sparsity features, which outperformed temporal sparsity features in classifying dysarthric speech using SVMs. Additionally, [159] applied Grassmann discriminant analysis to spectro-temporal subspaces, leveraging singular value decomposition informed by clinical insights into pathological distortions.

A common limitation across these studies is their reliance on relatively small speaker sets and their susceptibility to biases related to sex, age, recording conditions, or language. To evaluate generalization across different domains, [160] conducted a cross-database study, demonstrating that an SVM trained on one dataset suffered a significant performance drop when evaluated on another dataset. This highlights the critical need for models that are robust to distributional shifts, including those stemming from differences in dataset characteristics and recording environments. Moreover, classical machine learning approaches heavily depend on hand-crafted features, which may not fully capture the nuanced and abstract cues associated with pathological speech. These methods also tend to overlook important metadata such as speaker identity, sex, and language.

B. Deep Learning-based Approaches

Several studies have explored CNN-based architectures for pathological speech detection, leveraging their ability to extract local patterns from time-frequency representations. For instance, [39, 161] applied CNNs to dysarthric speech, achieving promising results on small datasets of control speakers and ALS patients. However, these models exhibited significant inter-speaker variability in performance, suggesting a lack of robustness across individuals. This limitation is particularly critical in clinical contexts where model reliability must generalize across diverse patient populations. Building on CNN-based models, [77] incorporated phase-based features, i.e., the modified group delay and instantaneous frequency spectra, as complementary inputs to magnitude spectra. This multi-representational approach improved the CNN performance for pathological speech detection. Similarly, [88] introduced temporal envelope features, reinforcing that temporal dynamics encode critical pathological cues. However, these representation augmentations increase model complexity and demand careful preprocessing, which may hinder real-world deployment. While most approaches focus solely on binary classification of pathological versus neurotypical speech, multi-task learning has shown potential for capturing broader articulatory patterns. Vásquez Correa et al. [162] introduced a multi-task system based on CNNs that simultaneously learned to classify pathological speech and predict 11 articulatory deficit attributes. This approach yielded improved generalization across speakers by constraining the learned representations with related speech tasks, demonstrating the utility of auxiliary objectives in enhancing model robustness.

While CNNs excel at spatial pattern recognition, they lack mechanisms to capture temporal dependencies inherent in speech. Recurrent architectures, particularly LSTMs, have been employed to address this gap [83, 123, 163, 164]. These models demonstrate improved performance but often require larger datasets to train effectively. Moreover, although the LSTM methodology in [163] showed promising results, the model has been evaluated only on syllables and its generalization to other speech modes remains unexplored.

More recently, there has been a shift toward end-to-end models using self-supervised learning (SSL) to bypass manual feature engineering. SSL models like wav2vec 2.0 leverage raw waveform inputs to learn contextual embeddings, achieving state-of-the-art results across several pathological speech datasets [70, 91, 165]. These models outperform classical approaches in both accuracy and scalability. However, despite their empirical success, SSL-based models function as black boxes, raising concerns about clinical interpretability and decision transparency.

V. PATHOLOGICAL SPEECH RECOGNITION

ASR systems have significantly advanced over the years, demonstrating impressive results in converting raw speech signals into their corresponding textual form. This has resulted in the widespread use of various interactive devices, including smartphones and voice assistants. However, they often fail to recognize low resource pathological speech, including speech from patients suffering from various neurodegenerative conditions such as PD or ALS [31].

To alleviate this problem, several attempts have been made in improving the performance of ASR systems for pathological speakers. For example, [166] proposed a two-step speaker adaptation method. In the first step, a model trained on extensive control speech data is fine-tuned for dysarthric ASR. In the second step, this fine-tuned model undergoes further adaptation to a specific dysarthric speaker. Additionally, [167] proposed the additive angular margin loss to address intra-class variation among dysarthric speakers, demonstrating promising results on Japanese speakers. Green et al. [168] also showed that personalized ASR systems fine-tuned on pathological speech exhibit better recognition performance compared to speaker-independent ASR systems. However, the reliance on specific speakers in personalized ASR systems poses challenges for generalization settings. Hermann and Magimai-Doss [169] utilize lattice-free maximum mutual information to mitigate insertion errors, which are otherwise prevalent due to the slow speaking rates of individuals with dysarthria. In a different approach, Xiong et al. [170] utilized transfer learning and found that speaker-based data selection leads to negative transfer. They recommended using utterance-based data selection with an entropy distribution to enhance recognition. Yue et al. [171] further implemented a multi-task learning (MTL) framework through auto-encoder joint learning, utilizing bottleneck features on out-of-domain data. Employing MTL in pathological ASR showed a lower word error rate compared to its single-task counterpart. Shahamiri [172] demonstrated that state-of-the-art ASR systems for

pathological speech are significantly impacted by phoneme inaccuracies. To address this, they proposed Speech Vision, a transfer learning paradigm that converts word utterances into visual feature representations, aiming to recognize the shape of the word rather than relying on phonemes.

A comprehensive overview of ASR systems for pathological speakers in terms of progress and challenges is provided in [32], where it is shown that pathological speech recognition performance can be significantly improved through a combination of neural architecture search, data augmentation, speaker adaptation, and multi-modal learning. These techniques address challenges such as limited training data, high inter-speaker variability, and reduced speech intelligibility. It should be noted that due to the large number of parameters in state-of-the-art ASR models, fine-tuning them on low-resource pathological speech data can be highly expensive. However, leveraging fine-tuning techniques such as Low-Rank Adaptation (LoRA) offers a more efficient alternative, reducing computational costs and making training more feasible [173, 174]. Additionally, LoRA's modular approach might make it easier to adapt the fine-tuned models to new speakers, providing flexibility in real-world applications or clinical scenarios.

VI. INTELLIGIBILITY ENHANCEMENT OF PATHOLOGICAL SPEECH

Pathological speech enhancement refers to improving the intelligibility and quality of speech affected by impairments such as dysarthria. The benefits of such enhancement are twofold. First, enhanced speech can ease human-human and human-machine communication for pathological speakers, promoting their social and digital inclusion. Second, enhancement methods can be leveraged for data augmentation (cf. Section VIII), aiding in the development of robust models for pathological speech processing.

Early efforts in pathological speech enhancement focused on explicitly correcting articulatory and acoustic deficiencies. For instance, [33] improved intelligibility by restructuring formant trajectories to better match intended speech targets. Rudzicz [175] applied a range of techniques including pronunciation correction, phoneme insertion, tempo adjustment, and removal of disfluencies. Hosom et al. [176] modified short-term spectral features to enhance word-level intelligibility, though their approach was speaker-dependent. Lalitha et al. [177] proposed a speaker-specific Cepstrum-based method, while [178] introduced a two-stage framework that combined ASR with speech synthesis to correct mispronunciations.

Several mapping-based approaches aiming to transform dysarthric speech into more intelligible forms have also been explored. For example, [179] proposed a speech enhancement method where a CNN is trained to directly map dysarthric utterances to their control counterparts. At the feature level, techniques such as LPC mapping and frequency warping of LPC poles have been explored in [180] and [181]. Additionally, [111] investigated feature-level mapping using time-delay neural networks.

More recent advances leverage voice conversion models for pathological intelligibility enhancement. For instance, [37]

framed the task as a style transfer problem and employed GANs to convert dysarthric to typical speech. Similarly, [38] utilized CycleGAN for dysarthric-to-typical speech conversion, and [182] applied DiscoGAN to map pathological and typical speech features at the acoustic level. Among these, [183] showed that time-stretching combined with MaskCycleGAN outperformed other GAN-based models in intelligibility enhancement, although the technique's practicality remains limited due to computational complexity. In a more comprehensive system, [184] integrated Transformer-TTS, CycleVAE-VC, and LPCNet to generate highly intelligible dysarthric speech, albeit at the expense of naturalness. Additionally, [185] proposed an end-to-end voice conversion system using knowledge distillation for enhanced dysarthric speech synthesis. To avoid adversarial training instability in GAN-based voice conversion models, neural encoder-decoder architectures have also been investigated for improving pathological speech intelligibility. For instance, [186] introduced Unit-DSR, which converts dysarthric speech into discrete linguistic units and then reconstructs speech from these normalized units using a neural vocoder.

In summary, the evolution of pathological speech enhancement methods reflects a progression from early signal processing and mapping approaches to advanced generative models and neural architectures. While GAN-based models have significantly improved intelligibility, challenges remain in balancing enhancement quality, naturalness, and computational efficiency. Neural encoder-decoder frameworks show promise in addressing these challenges, marking an important direction for future research.

VII. INTELLIGIBILITY AND SEVERITY ASSESSMENT OF PATHOLOGICAL SPEECH

Intelligibility. Intelligibility of pathological speech is a critical indicator for evaluating the effectiveness of speech therapy and tracking the progression of various disorders. To reduce the burden of evaluating pathological speech intelligibility in clinical practice, automatic approaches have been proposed in the literature. Automatic pathological speech intelligibility assessment methods are typically categorized into two main approaches, i.e., blind and non-blind approaches [110]. In blind approaches, the objective is to assess the intelligibility of impaired speech without exploiting reference neurotypical speech data [187–193]. These approaches primarily focus on extracting acoustic features such as jitter, fundamental frequency, shimmer, formant frequencies, etc., that are believed to be closely correlated with speech intelligibility. These features are then used in regression models to estimate the intelligibility of pathological speech. Non-blind approaches, by contrast, rely on intelligible speech from neurotypical speakers as a basis for estimating the intelligibility of pathological speech [194–203]. Such approaches typically use features extracted from ASR systems, which have been trained on large amounts of control speech, to train regression models to estimate the intelligibility of pathological speech. To avoid the burden of collecting and transcribing a large amount of neurotypical speech data required for such systems, [204] proposed the pathological short-time objective intelligibility measure (P-STOI) adapted from

the speech enhancement domain. The P-STOI measure first calculates an utterance-dependent fully intelligible representation from a small set of control speakers. The intelligibility of the pathological utterance is then evaluated by quantifying its divergence from this reference representation in terms of the short-time spectral correlation. While advantageous, P-STOI requires recordings of the same utterance from intelligible control speakers, which may not always be available. To mitigate this issue, [205] developed a method to generate synthetic reference speech for assessing pathological speech intelligibility. In a different approach, [206] introduced subspace based intelligibility measures based on the premise that dominant spectral patterns in pathological speech deviate significantly from those of intelligible speech. Although such measures result in a lower performance than measures exploiting neurotypical intelligible speech, they can be directly used in practical scenarios where such speech material is not available or easy to generate.

Severity. Besides intelligibility assessment, severity assessment is another important research area where developing tools for this purpose could greatly assist in automatizing the tedious process of screening patients and categorizing them into different subgroups based on the severity level. The methods developed in the literature for this purpose can be categorized into two categories, i.e., traditional machine learning-based approaches using the Mahalanobis distance classifier [187], SVMs [207], GMMs [208], or decision trees [209], and deep learning-based approaches [90, 210–215]. While deep learning approaches aim to automatically extract acoustic cues correlated with severity from raw or minimally processed speech signals, traditional machine learning approaches for severity assessment rely on (clinically informed) handcrafted acoustic features. For example, motivated by auditory processing knowledge, [216] introduced perceptually enhanced single frequency cepstral coefficients for assessing the severity of pathological speech. Vásquez-Correa et al. [217] showed that articulation features extracted from continuous speech signals to create i-vectors were advantageous in quantifying the dysarthria severity level. Based on the knowledge that pathological speakers often exhibit irregular rhythm patterns in their speech, [218, 219] explored rhythm-based features for severity assessment.

Besides speech impairments, patients with various pathological conditions often display distinct facial expressions. To harness these visual features, [220] introduced the first audio-visual pathological severity classification system using CNNs. To leverage metadata information such as age, sex, and type of pathological condition, [221] proposed a multi-head attention-based MTL framework. This approach jointly optimizes severity, type, sex, and age classifications for pathological speech, thereby enhancing the robustness of latent features across these additional factors. Recently, there has been growing interest in using SSL embeddings to measure pathological speech severity. This approach is promising due to the scarcity of labeled pathological data and the ability to leverage unlabeled data and metadata from other datasets, making it well-suited for resource-constrained settings [79, 112].

VIII. DATA AUGMENTATION FOR PATHOLOGICAL SPEECH APPLICATIONS

To mitigate overfitting in deep learning models dealing with the low resource pathological speech data, data augmentation techniques have been used for various tasks. Existing approaches to data augmentation in pathological speech are broadly based on traditional strategies (such as incorporating noise, reverberation, or multiple datasets), perturbation strategies, voice conversion, and text-to-speech (TTS) synthesis.

Takashima et al. [222] employed a strategy that involved combining diverse pathological speech data from multiple languages to increase the number of data samples. Their results indicate that such a traditional data augmentation approach can be advantageous for pathological ASR. Based on the temporal and speed differences between pathological and control speech, [119, 223, 224] have explored vocal tract length perturbation, tempo perturbation, and speed perturbation as data augmentation approaches for pathological ASR. Similarly, [225] introduced a data augmentation technique that adjusts the phonetic-level tempo of healthy speech to resemble atypical speech, and vice versa. Their findings demonstrated that the former approach is more effective for pathological ASR. More recently, voice conversion and TTS systems are also commonly employed to generate pathological speech from healthy speech [37, 226, 227]. These synthetic samples serve a dual purpose, i.e., they can be used for speech enhancement, improving the clarity and quality of pathological speech; and they can be used for expanding the dataset, thereby enhancing the diversity of training data. Increased data diversity is beneficial to avoid overfitting and improve the performance of deep learning models in tasks such as pathological speech detection and pathological ASR [37, 226, 228, 229]. Since pathological speech datasets typically have a limited vocabulary, using a voice conversion or TTS model can also be used to expand the set of out-of-vocabulary words [113]. Recently, [147] investigated different data augmentation strategies in pathological ASR, demonstrating that GAN-based conversion methods are more effective than perturbation-based augmentation approaches. However, a comprehensive investigation of the advantages of all data augmentation strategies in various pathological speech processing tasks using multiple datasets is still lacking.

IX. CHALLENGES AND RESEARCH DIRECTIONS IN PATHOLOGICAL SPEECH PROCESSING

In clinical settings, the integration of automated speech processing systems for pathological speech analysis is crucial for advancing the diagnosis, therapy, and monitoring of various disorders. While standardized clinical scales are invaluable in perceptually evaluating pathological speech, the incorporation of automated systems can offer additional benefits in terms of objectivity, efficiency, and real-time feedback. Tools such as VoxTester [230], for instance, provide clinicians with quantifiable speech metrics, including articulation precision and speech rate, which help monitor disease progression and assess the effectiveness of interventions. However, to fully realize the potential of these tools, further research

is needed to ensure that they can seamlessly integrate into clinical workflows, offering clinicians user-friendly, reliable, and interpretable outputs. Addressing issues such as clinician training, data security, patient consent, and adaptability to various clinical settings is crucial for the real-world adoption of these systems. Moreover, creating systems that can operate longitudinally, i.e., tracking speech performance over time to capture subtle changes in speech function, would significantly enhance therapeutic outcomes. We believe that addressing the challenges and research directions outlined in the remainder of this section will be key to enabling the seamless integration of these technologies into clinical practice in the future.

A. Impact of Speech Mode

The large majority of previously reviewed deep learning-based pathological speech approaches have been proposed and validated on controlled speech tasks. Controlled speech, also known as non-spontaneous speech, involves utterances produced within a structured context, typically requiring participants to repeat phonetically balanced, carefully crafted texts. This mode is designed to elicit specific pathological biomarkers by standardizing the motor planning and execution demands. Tasks may include reading aloud or repeating scripted phrases. In contrast, spontaneous speech consists of unplanned utterances, such as storytelling or casual conversation, which reflect real-world communicative behavior. It places greater demands on cognitive planning, articulation, and natural language generation, and may therefore reveal more authentic or varied pathological cues. While this distinction has received some attention in the context of detection, like in [92], its implications extend across other subfields such as severity assessment, intelligibility prediction, enhancement, and recognition. For instance, severity assessment models trained only on controlled speech may generalize poorly to real-world settings. Similarly, enhancement models may be tuned to the acoustic patterns of controlled tasks, failing to capture the variability in spontaneous speech. Incorporating spontaneous speech more broadly into pathological speech research could enhance practical utility and performance across these subfields. Given its ease of collection and alignment with natural communication, spontaneous speech provides a more suitable and informative basis for training and evaluating models [92, 231, 232]. Future work should consider systematically analyzing the effect of speech mode across multiple application areas to ensure generalizability and real-world effectiveness.

B. Robustness in Pathological Speech Detection

The methods developed so far for automatic pathological speech detection have typically been designed and tested under specific environmental conditions and for a particular language. Each dataset used for pathological speech detection has its own unique variabilities in terms of both inter-speaker and intra-speaker differences. Additionally, different age groups exhibit distinct speaker attributes, adding complexity to pathological speech analysis. Language is another important factor; separate models have been independently

developed for different languages. However, there is a need for a more robust model that is language-agnostic (to a possible degree given that pathological cues might be different in different languages), age-agnostic, accent-agnostic, and sex-agnostic to enhance overall effectiveness [151]. Additionally, datasets such as TORGO are contaminated with noise. Researchers have shown that models trained on these datasets tend to learn environmental factors rather than focusing on genuine pathological cues [70]. The development of pathological speech detection models robust to environmental distortions has been very limited, with only a few small studies addressing this area [71, 233–235]. For example, [234] employed a test-time adaptation method to fine-tune pre-trained models on a validation set augmented with the test noise extracted from the test utterance. This method improves the robustness of state-of-the-art pathological speech detection methods, offering a promising solution to deploying such applications in realistic clinical settings. Similarly, [236] proposed an approach to resolve the noise disparity between the two groups of speakers in the TORGO database, such that models developed on this database learn pathology-discriminant cues instead of noise-discriminant ones. Besides robustness to environmental distortions, adversarial robustness of pathological speech detection models is another important topic and research direction. The impact of acoustically imperceptible adversarial perturbations on deep learning-based pathological speech detection models has been explored in [71]. Results revealed a high vulnerability of such models to adversarial perturbations, with adversarial training ineffective in enhancing robustness.

C. Improving the Performance of Automatic Pathological Speech Analysis

While many automatic pathological speech detection methods have shown remarkable performance, the exploitation of common attributes such as pathological cues, speaker characteristics, age, and sex across different speakers remains limited. A promising direction is to approach the pathological speech detection problem as a semi-supervised node graph classification task using graph neural networks (GNNs), as demonstrated in [231]. This approach could involve constructing an inter-speaker graph based on utterances from various speakers, where the graph's connectivity would help form speaker clusters based on the presence or absence of pathological cues. In this context, domain knowledge can be easily integrated by establishing edges based on factors such as sex, age, and the severity scale of the patients. Additionally, the recent availability of longitudinal multimodal data [61] enables novel applications of GNNs in disease monitoring. By representing the patient history as a temporal graph where nodes capture multimodal features (e.g., vocal biomarkers, facial expressivity) and edges encode their dynamic interactions, GNNs can model disease progression through evolving graph topologies.

An additional area that remains under-explored in current research, potentially due to the lack of large datasets, is the adoption of Bayesian frameworks and generative approaches for pathological speech processing. Bayesian frameworks are

particularly relevant in healthcare, as they allow the model to estimate uncertainties in its predictions, which is essential for high-risk settings where incorrect predictions can have serious consequences. These models are ideal for situations where speech data may vary considerably due to speaker differences or environmental factors. By providing probabilistic models, Bayesian methods can account for such variability, making them highly suitable for clinical applications where individual differences are pronounced.

D. Privacy

In pathological speech-based applications such as detection models or ASR systems, privacy-preserving solutions are crucial for safeguarding sensitive patient data. Since these systems process speech that reveals sensitive medical information, ensuring data security and confidentiality is of paramount importance. By employing privacy-preserving techniques such as federated learning, differential privacy, and encryption, providers can ensure that individuals' speech data is anonymized and never exposed to unauthorized parties. This not only fosters trust among patients but also complies with stringent data protection regulations, thereby mitigating the risk of breaches. Developing privacy-preserving pathological speech processing systems where one must balance two conflicting goals, i.e., increasing the utility of the models while preserving the privacy of the users, remains an important challenging research directions.

E. Multimodal Pathological Speech Analysis

Most existing systems primarily focus on leveraging speech cues for detecting pathological speech. However, complementary cues may also exist in visual forms, such as facial expressions or lip movements, which can provide additional insights. To effectively utilize these visual cues, multimodal self-supervised methods such as AV-HuBERT [237] could prove advantageous, as they facilitate the integration of multimodal audio and visual information, potentially enhancing the accuracy and robustness of pathological speech detection models. Furthermore, combining other modalities, such as medical imaging, health records, brain signals, and textual data, could offer a more complete characterization of disorders. These multimodal systems may help in capturing underlying neural, visual or linguistic patterns associated with speech pathology, improving pathological speech processing across diverse patient groups.

F. Explainability and Interpretability

In the domain of pathological speech detection, explainability and interpretability are crucial due to their clinical significance. Although these terms are often used interchangeably, it is important to distinguish between them. As defined in [238], interpretable models are designed to be inherently understandable, whereas explainable models provide post-hoc explanations for the decisions made by existing black box systems that are otherwise incomprehensible to humans.

Despite their importance, relatively little attention has been paid to explainability and interpretability in pathological

speech detection, with most existing research focusing primarily on improving model accuracy. Nonetheless, a few notable efforts have begun to address this gap [239–243]. For example, [239] mapped high-dimensional acoustic features to binary phonological representations, recognizing that raw acoustic features are difficult to interpret in clinical practice. Xu et al. [243] applied the SHAP algorithm to identify the most influential features in their model, highlighting the role of consonant-vowel transitions in reversing classification decisions. Similarly, [232] used canonical correlation analysis to show that the 0–210 Hz frequency range strongly influences model outputs.

Recently, [151] found that non-interpretable SSL embeddings outperform interpretable features (e.g., prosodic, linguistic, and cognitive descriptors) in both multilingual and cross-lingual contexts. This highlights a growing challenge, i.e., as models increasingly rely on high-performing but opaque representations like SSL embeddings, there is a pressing need to develop methods that make these models explainable and clinically trustworthy.

G. Large Language Models for Pathological Speech

Given the recent advancements in multimodal large language models (LLMs), which have demonstrated significant progress across a wide range of applications [244, 245], exploring their potential for pathological speech analysis appears to be an inevitable and promising direction. Multimodal LLMs could be leveraged for tasks such as detecting speech impairments, ASR, and enhancing the reconstruction of unintelligible or difficult speech. Their ability to capture complex patterns may lead to more accurate models for personalized therapy, rehabilitation, and assistive communication tools for individuals with speech impairments. Additionally, multimodal LLMs should be explored for their ability to explain their decisions. This exploration could bridge the gap between traditional speech processing techniques and state-of-the-art language models, opening up new avenues for more effective and adaptive speech rehabilitation systems.

X. CONCLUSION

This paper provides a comprehensive overview of speech analysis and technologies for pathological speech arising due to neurological disorders, encompassing detection, recognition, intelligibility assessment, and enhancement. Additionally, it compiles a thorough list of both accessible and non-accessible pathological speech datasets, which will serve as valuable resources for future research and accelerate progress in the field. Finally, it outlines potential future research directions, particularly in the context of robust and interpretable models that can be deployed in clinical practice.

ACKNOWLEDGMENT

This work was supported by the Swiss National Science Foundation project CRSII5_202228 on “Characterisation of motor speech disorders and processes”.

REFERENCES

- [1] K. Simonyan, H. Ackermann, E. F. Chang, and J. D. Greenlee, “New developments in understanding the complexity of human speech production,” *Journal of Neuroscience*, vol. 36, no. 45, Nov. 2016.
- [2] G. K. Sewall, J. Jiang, and C. N. Ford, “Clinical evaluation of Parkinson’s-related dysphonia,” *Laryngoscope*, vol. 116, no. 10, pp. 1740–1744, Oct. 2024.
- [3] N. Isshiki, H. Okamura, M. Tanabe, and M. Morimoto, “Differential diagnosis of hoarseness,” *Folia Phoniatrica*, vol. 21, no. 1, pp. 9–19, 1969.
- [4] J. R. Duffy, *Motor Speech Disorders: Substrates, Differential Diagnosis, and Management*, 4th ed. Elsevier Health Sciences, 2019.
- [5] J. Ogar, H. Slama, N. Dronkers, S. Amici, and M. L. Gorno-Tempini, “Apraxia of speech: An overview,” *Neurocase*, vol. 11, no. 6, pp. 427–459, Dec. 2005.
- [6] J. Tröger, F. Dörr, L. Schwed, N. Linz, A. König, T. Thies, J. R. Orozco-Arroyave, and J. Rusz, “An automatic measure for speech intelligibility in dysarthrias—validation across multiple languages and neurological disorders,” *Frontiers in Digital Health*, vol. 6, p. 1440986, 2024.
- [7] J. L. Cummings, F. Benson, M. A. Hill, and S. Read, “Aphasia in dementia of the Alzheimer type,” *Neurology*, vol. 35, no. 3, pp. 394–401, Mar. 1985.
- [8] J. S. Damico, N. Müller, and M. J. Ball, *The Handbook of Language and Speech Disorders*. Wiley Online Library, 2010.
- [9] X. Huang, A. Acero, H.-W. Hon, and R. Reddy, *Spoken Language Processing: A Guide to Theory, Algorithm, and System Development*, 1st ed. USA: Prentice Hall PTR, 2001.
- [10] E. A. Strand, J. R. Duffy, H. M. Clark, and K. Josephs, “The apraxia of speech rating scale : A tool for diagnosis and description of apraxia of speech,” *Journal of Communication Disorders*, vol. 51, pp. 43–50, Sep. 2014.
- [11] F. S. Juste, F. C. Sassi, J. B. Costa, and C. R. F. de Andrade, “Frequency of speech disruptions in Parkinson’s disease and developmental stuttering: A comparison among speech tasks,” *PLoS One*, vol. 13, no. 6, June 2018.
- [12] R. T. Sataloff, *Clinical Assessment of Voice, Second Edition*. Plural Publishing, Sep 2017.
- [13] L. Moro-Velazquez, J. A. Gomez-Garcia, J. I. Godino-Llorente, J. Villalba, J. Rusz, S. Shattuck-Hufnagel, and N. Dehak, “A forced Gaussians based methodology for the differential evaluation of Parkinson’s disease by means of speech processing,” *Biomedical Signal Processing and Control*, vol. 48, pp. 205–220, Feb 2019.
- [14] W. H. Organization, <https://www.who.int/news-room/fact-sheets/detail/parkinson-disease#:~:text=Global%20estimates%20in%202019%20showed,of%20over%20100%25%20since%202000.,> 2023, [Online; accessed 23.12.2024].
- [15] GBD 2016 Parkinson’s Disease Collaborators, “Global, regional, and national burden of Parkinson’s disease, 1990–2016: a systematic analysis for the global burden of disease study 2016,” *The Lancet Neurology*, vol. 17, no. 11, Nov. 2016.
- [16] A. Wimo, G.-C. Ali, M. Guerchet, M. Prince, M. Prina, and Y.-T. Wu, “World Alzheimer report 2015,” *Alzheimer’s Disease International*, Tech. Rep., 2015.
- [17] K. C. Arthur, A. Calvo, T. R. Price, J. Geiger, A. Chip, and B. J. Traynor, “Projected increase in Amyotrophic Lateral Sclerosis from 2015 to 2040,” *Nature Communications*, vol. 7, no. 12408, Aug. 2016.
- [18] J. Rusz, R. Cmejla, T. Tykalova, H. Ruzickova, J. Klempir, V. Majerova, J. Picmausova, J. Roth, and E. Ruzicka, “Imprecise vowel articulation as a potential early marker of parkinson’s disease: Effect of speaking task,” *The Journal of the Acoustical Society of America*, vol. 134, no. 3, pp. 2171–2181, Sep 2013.
- [19] P. Rong, Y. Yunusova, J. Wang, and J. Green, “Predicting early bulbar decline in Amyotrophic Lateral Sclerosis: A speech

- subsystem approach,” *Behavioural Neurology*, vol. 2015, pp. 1–11, Jul. 2015.
- [20] J. M. Tracy, Y. Özkanca, D. C. Atkins, and R. H. Ghomi, “Investigating voice as a biomarker: Deep phenotyping methods for early detection of Parkinson’s disease,” *Journal of Biomedical Informatics*, vol. 104, p. 103362, April 2020.
 - [21] J. R. Duffy, R. K. Peach, and E. A. Strand, “Progressive Apraxia of Speech as a sign of motor neuron disease,” *American Journal of Speech-Language Pathology*, vol. 16, no. 3, pp. 198–208, Aug. 2007.
 - [22] J. R. Duffy, *Parkinson’s disease and movement disorders: diagnosis and treatment guidelines for the practicing physician*. Humana Press, 2000, ch. Motor speech disorders: Clues to neurologic diagnosis, pp. 35–53.
 - [23] G. B. Kempster, B. R. Gerratt, K. V. Abbott, J. Barkmeier-Kraemer, and R. E. Hillman, “Consensus auditory-perceptual evaluation of voice: Development of a standardized clinical protocol,” *American Journal of Speech Language Pathology*, vol. 18, no. 2, pp. 124–132, Oct. 2009.
 - [24] K. Omori, “Diagnosis of voice disorders,” *Japan Medical Association Journal*, vol. 54, no. 4, pp. 248–253, July 2011.
 - [25] N. Sáenz-Lechón, J. I. Godino-Llorente, V. Osma-Ruiz, M. Blanco-Velasco, and F. Cruz-Roldán, “Automatic assessment of voice quality according to the GRBAS scale,” in *Proc. Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, vol. 2006, New York, USA, Aug 2006, pp. 2478–2481.
 - [26] G. T. Stebbins and C. G. Goetz, “Factor structure of the unified Parkinson’s disease rating scale: Motor examination section,” *Movement Disorders: Journal of the Movement Disorder Society*, vol. 13, no. 4, pp. 633–636, July 1998.
 - [27] W. Ziegler, A. Staiger, T. Schölderle, and M. Vogel, “Gauging the auditory dimensions of dysarthric impairment: Reliability and construct validity of the Bogenhausen dysarthria scales (BoDyS),” *Journal of Speech, Language, and Hearing Research*, vol. 60, no. 6, pp. 1516–1534, June 2017.
 - [28] K. M. Yorkston and D. R. Beukelman, *Assessment of Intelligibility of Dysarthric Speech*. C.C. Publications, 1981.
 - [29] K. Bunton, R. Kent, J. R. Duffy, J. Rosenbek, and J. Kent, “Listener agreement for auditory-perceptual ratings of dysarthria,” *Journal of Speech, Language, and Hearing Research*, vol. 50, no. 6, pp. 1481–1495, Jan. 2008.
 - [30] S. Fonville, H. B. van der Worp, P. Maat, M. Aldenhoven, A. Algra, and J. van Gijn, “Accuracy and inter-observer variation in the classification of dysarthria from speech recordings,” *IEEE Transactions on Speech and Audio Processing*, vol. 255, no. 10, pp. 1545–1548, Oct. 2008.
 - [31] L. De Russis and F. Corno, “On the impact of dysarthric speech on contemporary ASR cloud platforms,” *Journal of Reliable Intelligent Environments*, vol. 5, pp. 163–172, 2019.
 - [32] S. Liu, M. Geng, S. Hu, X. Xie, M. Cui, J. Yu, X. Liu, and H. Meng, “Recent progress in the CUHK dysarthric speech recognition system,” *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 29, pp. 2267–2281, 2021.
 - [33] A. Kain, X. Niu, J.-P. Hosom, Q. Miao, and J. P. H. van Santen, “Formant re-synthesis of dysarthric speech,” in *Proc. 5th ISCA Workshop on Speech Synthesis (SSW 5)*, Pittsburgh, PA, USA, July 2004, pp. 25–30.
 - [34] M. Soleymanpour, M. T. Johnson, R. Soleymanpour, and J. Berry, “Synthesizing dysarthric speech using multi-speaker TTS for dysarthric speech recognition,” in *Proc. of International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Singapore, May 2022, pp. 7382–7386.
 - [35] C. Veaux, J. Yamagishi, and S. King, “Using HMM-based speech synthesis to reconstruct the voice of individuals with degenerative speech disorders,” in *Proc. of Annual Conference of the International Speech Communication*, Portland, OR, USA, Sept. 2012, pp. 967–970.
 - [36] M. Dhanalakshmi, T. A. Mariya Celin, T. Nagarajan, and P. Vijayalakshmi, “Speech-input speech-output communication for dysarthric speakers using HMM-based speech recognition and adaptive synthesis system,” *Circuits, Systems, and Signal Processing*, vol. 37, no. 2, pp. 674–703, Feb. 2018.
 - [37] S. H. Yang and M. Chung, “Improving dysarthric speech intelligibility using cycle-consistent adversarial training,” in *Proc. of International Conference on Bio-inspired Systems and Signal Processing*, Valletta, Malta, Feb. 2020.
 - [38] B. M. Halpern, J. Fritsch, E. Hermann, R. van Son, O. Scharenborg, and M. Magimai-Doss, “An objective evaluation framework for pathological speech synthesis,” in *Proc. of Speech Communication; 14th ITG Conference*, Kiel, Germany, Sep. 2021, pp. 1–5.
 - [39] H. M. Chandrashekar, V. Karjigi, and N. Sreedevi, “Spectro-temporal representation of speech for intelligibility assessment of dysarthria,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 14, no. 2, pp. 390–399, Feb. 2020.
 - [40] A. Tripathi, S. Bhosale, and S. K. Kopparapu, “Improved speaker independent dysarthria intelligibility classification using deepspeech posteriors,” in *Proc. of International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Virtual Barcelona, Spain, May 2020, pp. 6114–6118.
 - [41] J. A. Gómez-García, L. Moro-Velázquez, and J. I. Godino-Llorente, “On the design of automatic voice condition analysis systems. part ii: Review of speaker recognition techniques and study on the effects of different variability factors,” *Biomedical Signal Processing and Control*, vol. 48, pp. 128–143, Feb. 2019.
 - [42] L. Moro-Velazquez, J. A. Gomez-Garcia, J. D. Arias-Londoño, N. Dehak, and J. I. Godino-Llorente, “Advances in Parkinson’s disease detection and assessment using voice and speech: A review of the articulatory and phonatory aspects,” *Biomedical Signal Processing and Control*, vol. 66, p. 102418, April.
 - [43] K. Ding, M. Chetty, A. Noori Hoshyar, T. Bhattacharya, and B. Klein, “Speech based detection of Alzheimer’s disease: A survey of AI techniques, datasets and challenges,” *Artificial Intelligence Review*, vol. 57, no. 12, p. 325, Oct. 2024.
 - [44] H. P. Rowe, S. Shellikeri, Y. Yunusova, K. V. Chenausky, and J. R. Green, “Quantifying articulatory impairments in neurodegenerative motor diseases: A scoping review and meta-analysis of interpretable acoustic features,” *International Journal of Speech-Language Pathology*, vol. 25, no. 4, pp. 486–499, Aug. 2023.
 - [45] L. van Gelderen and C. Tejedor-García, “Innovative speech-based deep learning approaches for Parkinson’s disease classification: A systematic review,” *arXiv preprint arXiv:2407.17844*, 2024.
 - [46] C. Cordella, M. J. Marte, H. Liu, and S. Kiran, “An introduction to machine learning for speech-language pathologists: Concepts, terminology, and emerging applications,” *Perspectives of the ASHA Special Interest Groups*, pp. 1–19, 2024.
 - [47] Z. Brahmi, M. Mahyoob, M. Al-Sarem, J. Algaraady, K. Bouselmi, and A. Alblwi, “Exploring the role of machine learning in diagnosing and treating speech disorders: A systematic literature review,” *Psychology Research and Behavior Management*, pp. 2205–2232, May 2024.
 - [48] R. Gupta, D. R. Gunjawate, D. D. Nguyen, C. Jin, and C. Madill, “Voice disorder recognition using machine learning: A scoping review protocol,” *BMJ Open*, vol. 14, no. 2, Feb. 2024.
 - [49] M. U. Rehman, A. Shafique, S. S. Jamal, Y. Gheraibia, A. B. Usman *et al.*, “Voice disorder detection using machine learning algorithms: An application in speech and language pathology,” *Engineering Applications of Artificial Intelligence*, vol. 133, p. 108047, July 2024.
 - [50] F. Amato, G. Saggio, V. Cesarini, G. Olmo, and G. Costantini, “Machine learning and statistical-based voice analysis of Parkinson’s disease patients: A survey,” *Expert Systems with*

- Applications*, vol. 219, p. 119651, June 2023.
- [51] R. Gupta, T. Chaspari, J. Kim, N. Kumar, D. Bone, and S. Narayanan, "Pathological speech processing: State-of-the-art, current challenges, and future directions," in *Proc. of International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Shanghai, China, March 2016, pp. 6470–6474.
 - [52] F. Rudzicz, A. K. Namasivayam, and T. Wolff, "The TORGO database of acoustic and articulatory speech from speakers with dysarthria," *Language Resources and Evaluation*, vol. 46, no. 4, pp. 523–541, Dec. 2012.
 - [53] J. R. Orozco-Arroyave, J. D. Arias-Londoño, J. F. Vargas-Bonilla, M. C. González-Rátiva, and E. Nöth, "New Spanish speech corpus database for the analysis of people suffering from Parkinson's disease," in *Proc. of International Conference on Language Resources and Evaluation (LREC'14)*, Reykjavik, Iceland, May 2014, pp. 342–347.
 - [54] "MoSpeedi-ChaSpeePro dataset," <https://www.unige.ch/fapse/mospeedi/mospeedi-dataset>, database available within the MoSpeedi-ChaSpeePro project consortium.
 - [55] X. Menendez-Pidal, J. Polikoff, S. Peters, J. Leonzio, and H. Bunnell, "The Nemours database of dysarthric speech," in *Proc. of International Conference on Spoken Language Processing, ICSLP '96*, vol. 3, Philadelphia, USA, Oct. 1996, pp. 1962–1965 vol.3.
 - [56] K. H. Wong, Y. T. Yeung, E. H. Y. Chan, P. C. M. Wong, G.-A. Levow, and H. Meng, "Development of a Cantonese dysarthric speech corpus," in *Proc. of Annual Conference of the International Speech Communication Association*, Sept. 2015, pp. 329–333.
 - [57] J. Laures-Gore, S. Russell, R. Patel, and M. Frankel, "The Atlanta Motor Speech Disorders Corpus: Motivation, Development, and Utility," *Folia Phoniatrica et Logopaedica: International Association of Logopedics and Phoniatrics (IALP)*, vol. 68, no. 2, pp. 99–105, 2016.
 - [58] E. Yilmaz, M. Ganzeboom, L. Beijer, C. Cucchiari, and H. Strik, "A Dutch dysarthric speech database for individualized speech therapy research," in *Proc. of International Conference on Language Resources and Evaluation (LREC'16)*, Portorož, Slovenia, May 2016, pp. 792–795.
 - [59] R. Turrisi, A. Braccia, M. Emanuele, S. Giulietti, M. Pugliatti, M. Sensi, L. Fadiga, and L. Badino, "EasyCall corpus: A dysarthric speech dataset," in *Proc. of Annual Conference of the International Speech Communication*, Brno, Czech Republic, Sept. 2021, pp. 41–45.
 - [60] G. Van Nuffelen, M. De Bodt, C. Middag, and J.-P. Martens, "Dutch corpus of pathological and normal speech (COPAS)," Antwerp University Hospital and Ghent University, Tech. Rep., 2009.
 - [61] A. Favaro, A. Butala, T. Thebaud, J. Villalba, N. Dehak, and L. Moro-Velázquez, "Unveiling early signs of Parkinson's disease via a longitudinal analysis of celebrity speech recordings," *NPJ Parkinson's Disease*, vol. 10, no. 1, p. 207, 2024.
 - [62] G. Dimauro, V. Di Nicola, V. Bevilacqua, D. Caivano, and F. Girardi, "Assessment of Speech Intelligibility in Parkinson's Disease Using a Speech-To-Text System," *IEEE Access*, vol. 5, pp. 22 199–22 208, 2017.
 - [63] J. Mendes-Laureano, J. A. Gómez-García, A. Guerrero-López, E. Luque-Buzo, J. D. Arias-Londoño, F. J. Grandas-Pérez, and J. I. Godino-Llorente, "NeuroVoz: A Castilian Spanish corpus of Parkinsonian speech," *Scientific Data*, vol. 11, no. 1, p. 1367, Dec. 2024.
 - [64] M. Putzer and W. Barry, "Saarbrücken Voice Database," Institute of Phonetics, University of Saarland, 2021, accessed July 2025. [Online]. Available: <https://stimddb.coli.uni-saarland.de/>
 - [65] D. Martínez, E. Lleida, A. Ortega, A. Miguel, and J. Villalba, "Voice pathology detection on the Saarbrücken voice database with calibration and fusion of scores using multifocal toolkit," in *Proc. of Advances in Speech and Language Technologies for Iberian Languages: IberSPEECH*, Madrid, Spain, Nov. 2012, pp. 99–109.
 - [66] B. E. Sakar, M. E. Isenkul, C. O. Sakar, A. Sertbas, F. Gungen, S. Delil, H. Apaydin, and O. Kursun, "Collection and analysis of a Parkinson speech dataset with multiple types of sound recordings," *IEEE Journal of Biomedical and Health Informatics*, vol. 17, no. 4, pp. 828–834, July 2013.
 - [67] J. Hlavnička, R. Čmejla, J. Klempfř, E. Růžicka, and J. Rusz, "Acoustic tracking of pitch, modal, and subharmonic vibrations of vocal folds in Parkinson's disease and Parkinsonism," *IEEE Access*, vol. 7, pp. 150 339–150 354, Oct. 2019.
 - [68] M. Rusko, R. Sabo, M. Trnka, A. Zimmermann, R. Malaschitz, E. Ružický, P. Brandoburová, V. Kevická, and M. Škorvánek, "Slovak database of speech affected by neurodegenerative diseases," *Scientific Data*, vol. 11, no. 1, pp. 1–16, Dec. 2024.
 - [69] B. M. Bot, C. Suver, E. C. Neto, M. Kellen, A. Klein, C. Bare, M. Doerr, A. Pratap, J. Wilbanks, E. Dorsey *et al.*, "The mPower study, Parkinson disease mobile data collected using researchkit," *Scientific data*, vol. 3, no. 1, pp. 1–9, March 2016.
 - [70] G. Schu, P. Janbakhshi, and I. Kodrasi, "On using the UA-Speech and Torgo databases to validate automatic dysarthric speech classification approaches," in *Proc. of International Conference on Acoustics, Speech, and Signal Processing*, Rhodes Island, Greece, June 2023, pp. 1–5.
 - [71] M. Amiri and I. Kodrasi, "Suppressing noise disparity in training data for automatic pathological speech detection," in *Proc. of International Workshop on Acoustic Signal Enhancement*, Aalborg, Denmark, Sept. 2024, pp. 110–114.
 - [72] K. L. Lansford and J. M. Liss, "Vowel acoustics in dysarthria: Speech disorder diagnosis and classification," *Journal of Speech, Language, and Hearing Research*, vol. 57, pp. 57–67, Feb. 2014.
 - [73] Á. Piña Méndez, A. Taitz, O. Palacios Rodríguez, I. Rodríguez Leyva, and M. F. Assaneo, "Speech's syllabic rhythm and articulatory features produced under different auditory feedback conditions identify Parkinsonism," *Scientific Reports*, vol. 14, no. 1, p. 15787, July 2024.
 - [74] S. B. Davis, "Acoustic characteristics of normal and pathological voices," in *Speech and Language*. Elsevier, June 1979, vol. 1, pp. 271–335.
 - [75] J. R. Orozco-Arroyave, F. Hönl, J. D. Arias-Londoño, J. F. Vargas-Bonilla, S. Skodda, J. Rusz, and E. Nöth, "Voiced/unvoiced transitions in speech as a potential biomarker to detect Parkinson's disease," in *Proc. of Annual Conference of the International Speech Communication Association*, Dresden, Germany, Sept. 2015.
 - [76] I. Kodrasi and H. Bourlard, "Spectro-temporal sparsity characterization for dysarthric speech detection," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 28, pp. 1210–1222, April 2020.
 - [77] P. Janbakhshi and I. Kodrasi, "Experimental investigation on STFT phase representations for deep learning-based dysarthric speech detection," in *Proc. of International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Singapore, May 2022, pp. 6477–6481.
 - [78] J. Mallela, A. Illa, Y. Belur, N. Atchayaram, R. Yadav, P. Reddy, D. Gope, and P. K. Ghosh, "Raw speech waveform based classification of patients with ALS, Parkinson's disease and healthy controls using CNN-BLSTM," in *Proc. of Annual Conference of the International Speech Communication*, Shanghai, China, Oct. 2020, pp. 4586–4590.
 - [79] F. Javanmardi, S. Tirronen, M. Kodali, S. R. Kadiri, and P. Alku, "Wav2vec-based detection and severity level classification of dysarthria from speech," in *Proc. of International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Rhodes Island, Greece, June 2023, pp. 1–5.
 - [80] L. P. Violeta, W. C. Huang, and T. Toda, "Investigating self-supervised pretraining frameworks for pathological speech

- recognition,” in *Proc. of Annual Conference of the International Speech Communication*, Incheon, Korea, Sept. 2022, pp. 41–45.
- [81] F. Eyben, M. Wöllmer, and B. Schuller, “OpenSmile: The Munich versatile and fast open-source audio feature extractor,” in *Proc. of ACM International Conference on Multimedia*, Feirenze, Italy, Oct. 2010, pp. 1459–1462.
- [82] N. N. Prabhakera and P. Alku, “Dysarthric speech classification using glottal features computed from non-words, words and sentences,” in *Proc. of Annual Conference of the International Speech Communication*, Hyderabad, India, Sept. 2018, pp. 3403–3407.
- [83] J. Millet and N. Zeghidour, “Learning to detect dysarthria from raw speech,” in *Proc. of International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Brighton, UK, May 2019, pp. 5831–5835.
- [84] N. P. Narendra and P. Alku, “Dysarthric speech classification from coded telephone speech using glottal features,” *Speech Communication*, vol. 110, pp. 47–55, Jul. 2019.
- [85] L. Alhinti, S. Cunningham, and H. Christensen, “Recognising emotions in dysarthric speech using typical speech data,” in *Proc. of Annual Conference of the International Speech Communication*, Shanghai, China, Oct. 2020.
- [86] M. L. Ina Kodrasi, Michaela Pernon and H. Bourlard, “Automatic discrimination of apraxia of speech and dysarthria using a minimalistic set of handcrafted features,” in *Proc. of Annual Conference of the International Speech Communication*, Shanghai, China, Oct. 2020, pp. 4991–4995.
- [87] N. P. Narendra and P. Alku, “Automatic assessment of intelligibility in speakers with dysarthria from coded telephone speech using glottal features,” *Computer Speech & Language*, vol. 65, p. 101117, Jan. 2021.
- [88] I. Kodrasi, M. Pernon, M. Laganaro, and H. Bourlard, “Automatic and perceptual discrimination between dysarthria, apraxia of speech, and neurotypical speech,” in *Proc. of International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Toronto, Canada, June 2021, pp. 7308–7312.
- [89] A. Tripathi, S. Bhosale, and S. K. Kopparapu, “Automatic speaker independent dysarthric speech intelligibility assessment system,” *Computer Speech & Language*, vol. 69, p. 101213, Sep. 2021.
- [90] A. A. Joshy and R. Rajan, “Automated dysarthria severity classification: A study on acoustic features and deep learning techniques,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 30, pp. 1147–1157, May 2022.
- [91] F. Javanmardi, S. R. Kadiri, and P. Alku, “Pre-trained models for detection and severity level classification of dysarthria from speech,” *Speech Communication*, vol. 158, p. 103047, Mar. 2024.
- [92] S. Shakeel, A. and K. Ina, “Impact of speech mode in automatic pathological speech detection,” in *Proc. of European Signal Processing Conference*, Lyon, France, Aug. 2024.
- [93] T. Liu and X. Yuan, “Paralinguistic and spectral feature extraction for speech emotion classification using machine learning techniques,” *EURASIP Journal on Audio, Speech, and Music Processing*, vol. 2023, no. 1, p. 23, May 2023.
- [94] S. Toyama, D. Saito, and N. Minematsu, “Use of global and acoustic features associated with contextual factors to adapt language models for spontaneous speech recognition,” in *Proc. of Annual Conference of the International Speech Communication*, Stockholm, Sweden, Aug. 2017, pp. 543–547.
- [95] S. R. Shahamiri and S. S. Binti Salim, “Artificial neural networks as speech recognisers for dysarthric speech: Identifying the best-performing set of MFCC parameters and studying a speaker-independent approach,” *Advanced Engineering Informatics*, vol. 28, no. 1, pp. 102–110, Jan. 2014.
- [96] C. Bhat, B. Vachhani, and S. Kopparapu, “Recognition of dysarthric speech using voice parameters for speaker adaptation and multi-taper spectral estimation,” in *Proc. of Annual Conference of the International Speech Communication*, San Francisco, USA, Sept. 2016, pp. 228–232.
- [97] G. Vyas, M. K. Dutta, J. Prinosil, and P. Harár, “An automatic diagnosis and assessment of dysarthric speech using speech disorder specific prosodic features,” in *Proc. of International Conference on Telecommunications and Signal Processing (TSP)*, Vienna, Austria, June 2016, pp. 515–518.
- [98] B.-F. Zaidi, M. Boudraa, S.-A. Selouani, D. Addou, and M. S. Yakoub, “Automatic recognition system for dysarthric speech based on MFCC’s, PNCC’s, jitter and shimmer coefficients,” in *Proc. of Advances in Computer Vision, Advances in Intelligent Systems and Computing*, vol. 944, April 2020, pp. 500–510.
- [99] B. A. Al-Qatab and M. B. Mustafa, “Classification of dysarthric speech according to the severity of impairment: An analysis of acoustic features,” *IEEE Access*, vol. 9, pp. 18 183–18 194, Feb. 2021.
- [100] B. F. Zaidi, S. A. Selouani, M. Boudraa, and M. Sidi Yakoub, “Deep neural network architectures for dysarthric speech analysis and recognition,” *Neural Computing and Applications*, vol. 33, no. 15, pp. 9089–9108, Jan. 2021.
- [101] P. Sahane, S. Pangaonkar, and S. Khandekar, “Dysarthric speech recognition using multi-taper Mel frequency cepstrum coefficients,” in *Proc. of International Conference on Computing, Communication and Green Engineering (CCGE)*, Pune, India, Sept. 2021, pp. 1–4.
- [102] L. P. Sahu and G. Pradhan, “Analysis of short-time magnitude spectra for improving intelligibility assessment of dysarthric speech,” *Circuits, Systems, and Signal Processing*, vol. 41, no. 10, pp. 5676–5698, Oct. 2022.
- [103] J. Jothieswari, T. Manicka Sundara Valli, and S. Suguna, “Enhancing dysarthria detection: Harnessing ensemble models and MFCC,” in *Proc. of Smart Trends in Computing and Communications*, Pune, India, 2024, pp. 135–147.
- [104] J. Godino-Llorente and P. Gomez-Vilda, “Automatic detection of voice impairments by means of short-term cepstral parameters and neural network based detectors,” *IEEE Transactions on Biomedical Engineering*, vol. 51, no. 2, pp. 380–384, Feb 2004.
- [105] L. Moro-Velázquez, J. A. Gómez-García, J. I. Godino-Llorente, J. Villalba, J. R. Orozco-Aroyave, and N. Dehak, “Analysis of speaker recognition methodologies and the influence of kinetic changes to automatically detect Parkinson’s disease,” *Applied Soft Computing*, vol. 62, pp. 649–666, Jan 2018.
- [106] L. Moro-Velázquez, J. Villalba, and N. Dehak, “Using X-Vectors to automatically detect Parkinson’s disease from speech,” in *Proc. of International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Virtual Barcelona, Spain, May 2020, pp. 1155–1159.
- [107] J. D. Arias-Londoño, J. I. Godino-Llorente, M. Markaki, and Y. Stylianou, “On combining information from modulation spectra and Mel-frequency cepstral coefficients for automatic detection of pathological voices,” *Logopedics Phoniatrics Vocology*, vol. 36, no. 2, pp. 60–69, July 2011.
- [108] S. R. Kadiri and P. Alku, “Analysis and detection of pathological voice using glottal source features,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 14, no. 2, pp. 367–379, Dec. 2019.
- [109] I. Kodrasi and H. Bourlard, “Statistical modeling of speech spectral coefficients in patients with Parkinson’s disease,” in *Proc. of Speech Communication; 13th ITG-Symposium*, Oldenburg, Germany, Oct. 2018, pp. 1–5.
- [110] P. Janbakhshi, “Automatic pathological speech assessment,” Ph.D. dissertation, EPFL, June 2022.
- [111] C. Bhat, B. Das, B. Vachhani, and S. K. Kopparapu, “Dysarthric speech recognition using time-delay neural network based denoising autoencoder,” in *Proc. of Annual Conference of the International Speech Communication*, Hyderabad,

- India, 2018, pp. 451–455.
- [112] S. Rathod, M. Charola, A. Vora, Y. Jogi, and H. A. Patil, “Whisper features for dysarthry severity level classification,” in *Proc. of Annual Conference of the International Speech Communication*, Dublin, Ireland, Aug. 2023, pp. 1523–1527.
 - [113] J. Harvill, D. Issa, M. Hasegawa-Johnson, and C. Yoo, “Synthesis of new words for improved dysarthric speech recognition on an expanded vocabulary,” in *Proc. of International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Toronto, Canada, June 2021, pp. 6428–6432.
 - [114] S. A. Naeini, L. Simmatis, D. Jafari, Y. Yunusova, and B. Taati, “Improving dysarthric speech segmentation with emulated and synthetic augmentation,” *IEEE Journal of Translational Engineering in Health and Medicine*, vol. 12, pp. 382–389, March 2024.
 - [115] K. Vattis, B. Oubre, A. C. Luddy, J. S. Ouillon, N. M. Eklund, C. D. Stephen, J. D. Schmahmann, A. S. Nunes, and A. S. Gupta, “Sensitive quantification of cerebellar speech abnormalities using deep learning models,” *IEEE Access*, vol. 12, pp. 62 328–62 340, April 2024.
 - [116] Z. Yue, E. Loweimi, H. Christensen, J. Barker, and Z. Cvetkovic, “Acoustic modelling from raw source and filter components for dysarthric speech recognition,” *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 30, pp. 2968–2980, Sept. 2022.
 - [117] Y. Zhao, M. Song, Y. Yue, and M. Kuruvilla-Dugdale, “Personalizing TTS voices for progressive dysarthria,” in *Proc. of IEEE EMBS International Conference on Biomedical and Health Informatics (BHI)*, Virtual Athens, Greece, July 2021, pp. 1–4.
 - [118] L. Wu, D. Zong, S. Sun, and J. Zhao, “A sequential contrastive learning framework for robust dysarthric speech recognition,” in *Proc. of International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Toronto, Canada, June 2021, pp. 7303–7307.
 - [119] Z. Yue, E. Loweimi, and Z. Cvetkovic, “Raw source and filter modelling for dysarthric speech recognition,” in *Proc. of International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Singapore, May 2022, pp. 7377–7381.
 - [120] M. Geng, X. Xie, Z. Ye, T. Wang, G. Li, S. Hu, X. Liu, and H. Meng, “Speaker adaptation using spectro-temporal deep features for dysarthric and elderly speech recognition,” *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 30, pp. 2597–2611, July 2022.
 - [121] K. Vora, D. Padalia, D. Mehta, and D. Sharma, “Hybrid CNN-LSTM network to detect dysarthria using Mel-frequency cepstral coefficients,” in *Proc. of International Conference on Advances in Science and Technology (ICAST)*, Mumbai, India, Dec. 2022, pp. 615–621.
 - [122] J. C. Vázquez-Correa *et al.*, “Convolutional neural network to model articulation impairments in patients with Parkinson’s disease,” in *Proc. of Annual Conference of the International Speech Communication*, Stockholm, Sweden, Aug. 2017, pp. 314–318.
 - [123] C. D. Rios-Urrego, S. A. Moreno-Acevedo, E. Nöth, and J. R. Orozco-Arroyave, “End-to-end Parkinson’s disease detection using a deep convolutional recurrent network,” in *Proc. of International Conference on Text, Speech, and Dialogue*, Brno, Czech Republic, Sept. 2022, pp. 326–338.
 - [124] N. Narendra, B. Schuller, and P. Alku, “The detection of Parkinson’s disease from speech using voice source information,” *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 29, pp. 1925–1936, May 2021.
 - [125] A. Baevski, Y. Zhou, A. Mohamed, and M. Auli, “Wav2vec 2.0: A framework for self-supervised learning of speech representations,” in *Proc. of Annual Conference on Neural Information Processing Systems*, vol. 33, Virtual Online, Dec. 2020, pp. 12 449–12 460.
 - [126] W.-N. Hsu, B. Bolte, Y.-H. H. Tsai, K. Lakhotia, R. Salakhutdinov, and A. Mohamed, “HuBERT: Self-supervised speech representation learning by masked prediction of hidden units,” *IEEE/ACM Trans. Audio, Speech and Lang. Proc.*, vol. 29, p. 3451–3460, Oct. 2021.
 - [127] S. Chen, C. Wang, Z. Chen, Y. Wu, S. Liu, Z. Chen, J. Li, N. Kanda, T. Yoshioka, X. Xiao, J. Wu, L. Zhou, S. Ren, Y. Qian, Y. Qian, J. Wu, M. Zeng, X. Yu, and F. Wei, “WavLM: Large-scale self-supervised pre-training for full stack speech processing,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 16, no. 6, pp. 1505–1518, July 2022.
 - [128] Y. Jiang, T. Wang, X. Xie, J. Liu, W. Sun, N. Yan, H. Chen, L. Wang, X. Liu, and F. Tian, “Perceiver-prompt: Flexible speaker adaptation in Whisper for Chinese disordered speech recognition,” in *Proc. of Annual Conference of the International Speech Communication*, Kos, Greece, Sept. 2024, pp. 2025–2029.
 - [129] S. R. Kadiri, F. Javanmardi, and P. Alku, “Investigation of self-supervised pre-trained models for classification of voice quality from speech and neck surface accelerometer signals,” *Computer Speech & Language*, vol. 83, p. 101550, Jan. 2024.
 - [130] S. Tirronen, S. R. Kadiri, and P. Alku, “Hierarchical multi-class classification of voice disorders using self-supervised models and glottal features,” *IEEE Open Journal of Signal Processing*, vol. 4, pp. 80–88, Feb. 2023.
 - [131] H. Kheddar, Y. Himeur, S. Al-Maadeed, A. Amira, and F. Bensaali, “Deep transfer learning for automatic speech recognition: Towards better generalization,” *Knowledge-Based Systems*, vol. 277, p. 110851, Oct. 2023.
 - [132] J. D. Fritsch, “Novel methods for detection and analysis of atypical aspects in speech,” Ph.D. dissertation, EPFL, Lausanne, May 2023.
 - [133] D. Ribas, M. A. Pastor, A. Miguel, D. Martínez, A. Ortega, and E. Lleida, “Automatic voice disorder detection using self-supervised representations,” *IEEE Access*, vol. 11, pp. 14 915–14 927, Feb. 2023.
 - [134] M. Geng, X. Xie, R. Su, J. Yu, Z. Jin, T. Wang, S. Hu, Z. Ye, H. Meng, and X. Liu, “On-the-fly feature based rapid speaker adaptation for dysarthric and elderly speech recognition,” in *Proc. of Annual Conference of the International Speech Communication*, Dublin, Ireland, Aug. 2022.
 - [135] D. Mujtaba, N. R. Mahapatra, M. Arney, J. S. Yaruss, C. Herring, and J. Bin, “Inclusive ASR for disfluent speech: Cascaded large-scale self-supervised learning with targeted fine-tuning and data augmentation,” in *Proc. of Annual Conference of the International Speech Communication*, Kos, Greece, Sept. 2024, pp. 1275–1279.
 - [136] K. Soky, S. Li, C. Chu, and T. Kawahara, “Domain and language adaptation using heterogeneous datasets for Wav2vec2.0-based speech recognition of low-resource language,” in *Proc. of International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Rhodes Island, Greece, June 2023, pp. 1–5.
 - [137] M. Maisonneuve, C. Fredouille, M. Lalain, A. Ghio, and V. Woisard, “Towards objective and interpretable speech disorder assessment: A comparative analysis of CNN and transformer-based models,” in *Proc. of Annual Conference of the International Speech Communication*, Kos, Greece, Sept. 2024, pp. 1970–1974.
 - [138] T. Weise, A. Maier, K. C. Demir, P. A. Pérez-Toro, T. Arias-Vergara, B. Heismann, E. Nöth, M. Schuster, and S. H. Yang, “Impact of including pathological speech in pre-training on pathology detection,” in *Proc. of Text, Speech, and Dialogue*, Pilsen, Czech Republic, Sept. 2023, pp. 141–153.
 - [139] X. Liu, X. Du, J. Liu, R. Su, M. L. Ng, Y. Zhang, Y. Yang, S. Zhao, L. Wang, and N. Yan, “Automatic assessment of dysarthria using audio-visual vowel graph attention network,” *arXiv preprint arXiv:2405.03254*, 2024.
 - [140] M. Kheirkhazadeh, “Speech classification using acoustic embedding and large language models applied on Alzheimer’s

- disease prediction task,” Ph.D. dissertation, KTH, Sweden, Aug. 2023.
- [141] M. K. Baskar, T. Herzig, D. Nguyen, M. Diez, T. Polzehl, L. Burget, and J. H. Cernocký, “Speaker adaptation for Wav2vec2 based dysarthric ASR,” in *Proc. of Annual Conference of the International Speech Communication*, Incheon, South Korea, Sept. 2022.
- [142] S. Cullen, “Improving dysarthric speech recognition by enriching training datasets,” Ph.D. dissertation, TU Dublin, Ireland, March 2022.
- [143] T. Nguyen, C. Fredouille, A. Ghio, M. Balaguer, and V. Woisard, “Exploring pathological speech quality assessment with ASR-powered Wav2Vec2 in data-scarce context,” in *Proc. of Joint International Conference on Computational Linguistics, Language Resources and Evaluation*, Torino, Italy, May 2024, pp. 6935–6944.
- [144] Y. Lin, L. Wang, Y. Yang, and J. Dang, “CFDRN: A cognition-inspired feature decomposition and recombination network for dysarthric speech recognition,” *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 31, pp. 3824–3836, Sept. 2023.
- [145] P. Wang and H. Van hamme, “Benefits of pre-trained mono- and cross-lingual speech representations for spoken language understanding of Dutch dysarthric speech,” *EURASIP Journal on Audio, Speech, and Music Processing*, vol. 2023, no. 1, p. 15, Apr. 2023.
- [146] S. Hu, X. Xie, Z. Jin, M. Geng, Y. Wang, M. Cui, J. Deng, X. Liu, and H. Meng, “Exploring self-supervised pre-trained ASR models for dysarthric and elderly speech recognition,” in *Proc. of International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Rhodes Island, Greece, June 2023, pp. 1–5.
- [147] H. Wang, Z. Jin, M. Geng, S. Hu, G. Li, T. Wang, H. Xu, and X. Liu, “Enhancing pre-trained ASR system fine-tuning for dysarthric speech recognition using adversarial data augmentation,” in *Proc. of International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Seoul, Korea, April 2024, pp. 12311–12315.
- [148] A. Hernandez, P. A. Pérez-Toro, E. Nöth, J. R. Orozco-Arroyave, A. Maier, and S. H. Yang, “Cross-lingual self-supervised speech representations for improved dysarthric speech recognition,” in *Proc. of Annual Conference of the International Speech Communication*, Incheon, Korea, Sept. 2022, pp. 51–55.
- [149] M. Spijkerman, “Using voice conversion and time-stretching to enhance the quality of dysarthric speech for automatic speech recognition,” Ph.D. dissertation, University of Groningen, Netherlands, July 2022.
- [150] J. Shor and S. Venugopalan, “TRILLsson: Distilled Universal Paralinguistic Speech Representations,” in *Proc. of Annual Conference of the International Speech Communication*, Incheon, Korea, Sep. 2022, pp. 356–360.
- [151] A. Favaro, Y.-T. Tsai, A. Butala, T. Thebaud, J. Villalba, N. Dehak, and L. Moro-Velázquez, “Interpretable speech features vs. DNN embeddings: What to use in the automatic assessment of Parkinson’s disease in multi-lingual scenarios,” *Computers in Biology and Medicine*, vol. 166, p. 107559, Nov. 2023.
- [152] A. Tsanas, M. A. Little, P. E. McSharry, J. Spielman, and L. O. Ramig, “Novel speech signal processing algorithms for high-accuracy classification of Parkinson’s disease,” *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 5, pp. 1264–1271, May 2012.
- [153] E. Vaiciukynas, A. Verikas, A. Gelzinis, and M. Baciauskienė, “Detecting Parkinson’s disease from sustained phonation and speech signals,” *PLoS One*, vol. 12, no. 10, pp. 1–16, Oct. 2017.
- [154] S. Jothilakshmi, “Automatic system to detect the type of voice pathology,” *Applied Soft Computing*, vol. 21, pp. 244–249, Aug. 2014.
- [155] J. R. Orozco-Arroyave, F. Hönl, J. D. Arias-Londoño, J. F. Vargas-Bonilla, and E. Nöth, “Spectral and cepstral analyses for Parkinson’s disease detection in Spanish vowels and words,” *Expert Systems*, vol. 32, no. 6, pp. 688–697, Dec. 2015.
- [156] M. Novotný, J. Rusz, R. Čmejla, and E. Růžička, “Automatic evaluation of articulatory disorders in Parkinson’s disease,” *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 22, no. 9, pp. 1366–1378, Sept. 2014.
- [157] J. Wang, P. V. Kothalkar, B. Cao, and D. Heitzman, “Towards automatic detection of Amyotrophic Lateral Sclerosis from speech acoustic and articulatory samples,” in *Proc. of Annual Conference of the International Speech Communication*, San Francisco, USA, Sept. 2016, pp. 1195–1199.
- [158] R. Norel, M. Pietrowicz, C. Agurto, S. Rishoni, and G. Cecchi, “Detection of Amyotrophic Lateral Sclerosis (ALS) via acoustic analysis,” in *Proc. of Annual Conference of the International Speech Communication*, Hyderabad, India, Sept. 2018, pp. 377–381.
- [159] P. Janbakhshi, I. Kodrasi, and H. Bourlard, “Subspace-based learning for automatic dysarthric speech detection,” *IEEE Signal Processing Letters*, vol. 28, pp. 96–100, Dec. 2021.
- [160] S. Gillespie, Y.-Y. Logan, E. Moore, J. Laures-Gore, S. Russell, and R. Patel, “Cross-database models for the classification of dysarthria presence,” in *Proc. of Annual Conference of the International Speech Communication*, Stockholm, Sweden, Aug. 2017, pp. 3127–3131.
- [161] S. R. Mani Sekhar, G. Kashyap, A. Bhansali, A. A. Andrew, and K. Singh, “Dysarthric-speech detection using transfer learning with convolutional neural networks,” *Information & Communications Technology Express*, vol. 8, no. 1, pp. 61–64, March 2022.
- [162] J. C. Vázquez Correa, T. Arias, J. R. Orozco-Arroyave, and E. Nöth, “A multitask learning approach to assess the dysarthria severity in patients with Parkinson’s disease,” in *Proc. of Annual Conference of the International Speech Communication*, Hyderabad, India, Sept. 2018, pp. 456–460.
- [163] A. Mayle *et al.*, “Diagnosing dysarthria with long short-term memory networks,” in *Proc. of Annual Conference of the International Speech Communication*, Graz, Austria, Sept. 2019, pp. 4514–4518.
- [164] C. Bhat and H. Strik, “Automatic assessment of sentence-level dysarthria intelligibility using BLSTM,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 14, no. 2, pp. 322–330, Feb. 2020.
- [165] F. Javanmardi, S. R. Kadiri, and P. Alku, “Exploring the impact of fine-tuning the wav2vec2 model in database-independent detection of dysarthric speech,” *IEEE Journal of Biomedical and Health Informatics*, vol. 28, no. 8, pp. 4951–4962, April 2024.
- [166] R. Takashima, T. Takiguchi, and Y. Ariki, “Two-step acoustic model adaptation for dysarthric speech recognition,” in *Proc. of International Conference on Acoustics, Speech and Signal Processing*, Virtual Barcelona, Spain, May 2020, pp. 6104–6108.
- [167] Y. Takashima, R. Takashima, T. Takiguchi, and Y. Ariki, “Dysarthric speech recognition based on deep metric learning,” in *Proc. of Annual Conference of the International Speech Communication*, Shanghai, China, Oct. 2020, pp. 4796–4800.
- [168] J. R. Green, R. L. MacDonald *et al.*, “Automatic speech recognition of disordered speech: Personalized models outperforming human listeners on short phrases,” in *Proc. of Annual Conference of the International Speech Communication*, Brno, Czech Republic, Sept. 2021, pp. 4778–4782.
- [169] E. Hermann and M. Magimai.-Doss, “Dysarthric speech recognition with lattice-free MMI,” in *Proc. of International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Virtual Barcelona, Spain, May 2020, pp. 6109–6113.

- [170] F. Xiong, J. Barker, Z. Yue, and H. Christensen, "Source domain data selection for improved transfer learning targeting dysarthric speech recognition," in *Proc. of International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Virtual Barcelona, Spain, May 2020, pp. 7424–7428.
- [171] Z. Yue, H. Christensen, and J. Barker, "Autoencoder bottleneck features with multi-task optimisation for improved continuous dysarthric speech recognition," in *Proc. of Annual Conference of the International Speech Communication*, Shanghai, China, Oct. 2020, pp. 4581–4585.
- [172] S. R. Shahamiri, "Speech vision: An end-to-end deep learning-based dysarthric automatic speech recognition system," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 29, pp. 852–861, May 2021.
- [173] E. J. Hu, yelong shen, P. Wallis, Z. Allen-Zhu, Y. Li, S. Wang, L. Wang, and W. Chen, "LoRA: Low-rank adaptation of large language models," in *Proc. of International Conference on Learning Representations*, Virtual, Apr. 2022.
- [174] Z. Song, J. Zhuo, Y. Yang, Z. Ma, S. Zhang, and X. Chen, "LoRA-whisper: Parameter-efficient and extensible multilingual ASR," in *Proc. of Annual Conference of the International Speech Communication*, Kos, Greece, Sept. 2024, pp. 3934–3938.
- [175] F. Rudzicz, "Adjusting dysarthric speech signals to be more intelligible," *Computer Speech & Language*, vol. 27, no. 6, pp. 1163–1177, Sep. 2013.
- [176] J.-P. Hosom, A. Kain, T. Mishra, J. van Santen, M. Fried-Oken, and J. Staehely, "Intelligibility of modifications to dysarthric speech," in *Proc. of International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 1, Hong Kong, China, Apr. 2003, pp. I–I.
- [177] V. Lalitha, P. Prema, and L. Mathew, "A Kepstrum based approach for enhancement of dysarthric speech," in *Proc. of International Congress on Image and Signal Processing*, vol. 7, Yantai, China, Oct. 2010, pp. 3474–3478.
- [178] M. Dhanalakshmi and P. Vijayalakshmi, "Intelligibility modification of dysarthric speech using HMM-based adaptive synthesis system," in *Proc. of International Conference on Biomedical Engineering (ICoBE)*, Penang, Malaysia, Mar. 2015, pp. 1–5.
- [179] S. Wang, Y. Tsao, W. Zheng, H. Yeh, P. Li, S. Fang, and Y. Lai, "Dysarthric speech enhancement based on convolution neural network," in *Proc. of Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, Glasgow, Scotland, UK, July 2022, pp. 60–64.
- [180] S. A. Kumar and C. S. Kumar, "Improving the intelligibility of dysarthric speech towards enhancing the effectiveness of speech therapy," in *Proc. of International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, Jaipur, India, Sep. 2016, pp. 1000–1005.
- [181] A. Roy, L. Thakur, G. Vyas, and G. Raj, "Towards improving the intelligibility of dysarthric speech," in *Proc. of International Conference on Soft Computing and Signal Processing*, vol. 898, Hyderabad, India, Feb. 2019, pp. 547–559.
- [182] M. Purohit, M. Patel, H. Malaviya *et al.*, "Intelligibility improvement of dysarthric speech using MMSE DiscoGAN," in *Proc. of International Conference on Signal Processing and Communications (SPCOM)*, Bangalore, India, July 2020, pp. 1–5.
- [183] L. Prananta, B. M. Halpern, S. Feng, and O. Scharenborg, "The effectiveness of time stretching for enhancing dysarthric speech for improved dysarthric speech recognition," in *Proc. of Annual Conference of the International Speech Communication*, Incheon, Korea, Sept. 2022, pp. 36–40.
- [184] K. Matsubara, T. Okamoto, R. Takashima, T. Takiguchi, T. Toda, Y. Shiga, and H. Kawai, "High-intelligibility speech synthesis for dysarthric speakers with LPCNet-based TTS and CycleVAE-based VC," in *Proc. of International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Toronto, Canada, June 2021, pp. 7058–7062.
- [185] D. Wang, J. Yu *et al.*, "End-to-end voice conversion via cross-modal knowledge distillation for dysarthric speech reconstruction," in *Proc. of International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Virtual, Barcelona, Spain, May. 2020, pp. 7744–7748.
- [186] Y. Wang, X. Wu, D. Wang, L. Meng, and H. Meng, "UNIT-DSR: Dysarthric speech reconstruction system using speech unit normalization," in *Proc. of International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Seoul, Korea, Apr. 2024, pp. 12 306–12 310.
- [187] M. S. Paja and T. H. Falk, "Automated dysarthria severity classification for improved objective intelligibility assessment of spastic dysarthric speech," in *Proc. of Annual Conference of the International Speech Communication*, Portland, OR, USA, Sept. 2012, pp. 62–65.
- [188] D. Martínez, P. Green, and H. Christensen, "Dysarthria intelligibility assessment in a factor analysis total variability space," in *Proc. of Annual Conference of the International Speech Communication Association*, Lyon, France, Aug. 2013, pp. 2133–2137.
- [189] J. C. Kim, H. Rao, and M. A. Clements, "Speech intelligibility estimation using multi-resolution spectral features for speakers undergoing cancer treatment," *The Journal of the Acoustical Society of America*, vol. 136, no. 4, pp. 315–321, Oct. 2014.
- [190] R. Hummel, W.-Y. Chan, and T. H. Falk, "Spectral features for automatic blind intelligibility estimation of spastic dysarthric speech," in *Proc. of Annual Conference of the International Speech Communication Association*, Florence, Italy, Aug. 2011, pp. 3017–3020.
- [191] T. H. Falk, W.-Y. Chan, and F. Shein, "Characterization of atypical vocal source excitation, temporal dynamics and prosody for objective measurement of dysarthric word intelligibility," *Speech Communication*, vol. 54, no. 5, pp. 622–631, June 2012.
- [192] T. Haderlein, A. Schützenberger *et al.*, "Robust automatic evaluation of intelligibility in voice rehabilitation using prosodic analysis," in *Proc. of International Conference on Text, Speech, and Dialogue*, Prague, Czech Republic, Aug. 2017, pp. 11–19.
- [193] A. R. Fletcher, A. A. Wisler, M. J. McAuliffe, K. L. Lansford, and J. M. Liss, "Predicting intelligibility gains in dysarthria through automated speech feature analysis," *Journal of Speech, Language, and Hearing Research*, vol. 60, no. 11, pp. 3058–3068, Nov. 2017.
- [194] T. Haderlein, S. Steidl, E. Nöth, F. Rosanowski, and M. Schuster, "Automatic recognition and evaluation of tracheoesophageal speech," in *Proc. of International Conference on Text, Speech and Dialogue*, Brno, Czech Republic, Sept. 2004, pp. 331–338.
- [195] C. Middag, G. Van Nuffelen, J. P. Martens, and M. De Bodt, "Objective intelligibility assessment of pathological speakers," in *Proc. of Annual Conference of the International Speech Communication Association*, Brisbane, Australia, Sept. 2008, pp. 1745–1748.
- [196] C. Middag, J.-P. Martens, G. V. Nuffelen, and M. De Bodt, "Automated intelligibility assessment of pathological speech using phonological features," *EURASIP Journal on Advances in Signal Processing*, vol. 2009, no. 1, pp. 1–9, May 2009.
- [197] C. Middag, Y. Saeys, and J.-P. Martens, "Towards an ASR-free objective analysis of pathological speech," in *Proc. of Annual Conference of the International Speech Communication Association*, Makuhari, Chiba, Japan, Sept. 2010, pp. 294–297.
- [198] A. Maier, T. Haderlein, U. Eysholdt, F. Rosanowski, A. Batliner, M. Schuster, and E. Nöth, "PEAKS - A system for the automatic evaluation of voice and speech disorders," *Speech Communication*, vol. 51, no. 5, pp. 425–437, May 2009.
- [199] G. V. Nuffelen, C. Middag, M. De Bodt, and J.-P. Martens, "Speech technology-based assessment of phoneme intelligibility in dysarthria," *International Journal of Language &*

- Communication Disorders*, vol. 44, no. 5, pp. 716–730, Sept. 2009.
- [200] T. Bocklet, K. Riedhammer, U. Eysholdt, and T. Haderlein, “Automatic intelligibility assessment of speakers after laryngeal cancer by means of acoustic modeling,” *Journal of Voice*, vol. 26, no. 3, pp. 390–397, May 2012.
- [201] D. Martínez, E. Lleida, P. Green, H. Christensen, A. Ortega, and A. Miguel, “Intelligibility assessment and speech recognizer word accuracy rate prediction for dysarthric speakers in a factor analysis subspace,” *ACM Transactions on Accessible Computing*, vol. 6, no. 3, pp. 1–21, May 2015.
- [202] L. Ined, B. K. Waad, F. Corinne, and M. Christine, “Automatic prediction of speech evaluation metrics for dysarthric speech,” in *Proc. of Annual Conference of the International Speech Communication Association*, Stockholm, Sweden, Aug. 2017, pp. 1834–1838.
- [203] S. Kalita, S. R. Mahadeva Prasanna, and S. Dandapat, “Intelligibility assessment of cleft lip and palate speech using Gaussian posteriors based on joint spectro-temporal features,” *Journal of the Acoustical Society of America*, vol. 144, no. 4, pp. 2413–2423, Oct. 2018.
- [204] P. Janbakhshi, I. Kodrasi, and H. Bourlard, “Pathological speech intelligibility assessment based on the short-time objective intelligibility measure,” in *Proc. of International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Brighton, UK, May 2019, pp. 6405–6409.
- [205] —, “Synthetic speech references for automatic pathological speech intelligibility assessment,” in *Proc. of International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Virtual Barcelona, Spain, 2020, pp. 6099–6103.
- [206] —, “Automatic pathological speech intelligibility assessment exploiting subspace-based analyses,” *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 28, pp. 1717–1728, 2020.
- [207] E. J. Yeo, S. Kim, and M. Chung, “Automatic severity classification of Korean dysarthric speech using phoneme-level pronunciation features,” in *Proc. of Annual Conference of the International Speech Communication*, Brno, Czech Republic, Sept. 2021, pp. 4838–4842.
- [208] K. L. Kadi, S. A. Selouani, B. Boudraa, and M. Boudraa, “Fully automated speaker identification and intelligibility assessment in dysarthria disease using auditory knowledge,” *Biocybernetics and Biomedical Engineering*, vol. 36, no. 1, pp. 233–247, Jan. 2016.
- [209] R. Dubbio, M. Spisto, L. Verde, V. V. Iuzzolino, G. Senerchia, G. De Pietro, I. De Falco, and G. Sannino, “Precision medicine in ALS: Identification of new acoustic markers for dysarthria severity assessment,” *Biomedical Signal Processing and Control*, vol. 89, p. 105706, Mar. 2024.
- [210] M. Soleymannpour, M. T. Johnson, and J. Berry, “Increasing the precision of dysarthric speech intelligibility and severity level estimate,” *Lecture Notes in Computer Science*, vol. 12997, pp. 670–679, 2021.
- [211] A. A. Joshy and R. Rajan, “Automated dysarthria severity classification using deep learning frameworks,” in *Proc. of European Signal Processing Conference (EUSIPCO)*, Amsterdam, Netherlands, Aug. 2021, pp. 116–120.
- [212] S. Gupta, A. T. Patil, M. Purohit, M. Parmar, M. Patel, H. A. Patil, and R. C. Guido, “Residual neural network precisely quantifies dysarthria severity-level based on short-duration speech segments,” *Neural Networks*, vol. 139, pp. 105–117, Jul. 2021.
- [213] A. A. Joshy and R. Rajan, “Dysarthria severity assessment using squeeze-and-excitation networks,” *Biomedical Signal Processing and Control*, vol. 82, p. 104606, Apr. 2023.
- [214] K. Radha, M. Bansal, and V. R. Dulipalla, “Variable STFT layered CNN model for automated dysarthria detection and severity assessment using raw speech,” *Circuits, Systems, and Signal Processing*, vol. 43, no. 5, pp. 3261–3278, May 2024.
- [215] S. Sajiha, K. Radha, D. Venkata Rao, N. Sneha, S. Gunnam, and D. P. Bavirisetti, “Automatic dysarthria detection and severity level assessment using CWT-layered CNN model,” *EURASIP Journal on Audio, Speech, and Music Processing*, vol. 2024, no. 1, p. 33, June 2024.
- [216] K. Gurugubelli and A. K. Vuppala, “Perceptually enhanced single frequency filtering for dysarthric speech detection and intelligibility assessment,” in *Proc. of International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Brighton, UK, May 2019, pp. 6410–6414.
- [217] J. C. Vázquez-Correa, J. R. Orozco-Arroyave, T. Bocklet, and E. Nöth, “Towards an automatic evaluation of the dysarthria level of patients with parkinson’s disease,” *Journal of Communication Disorders*, vol. 76, pp. 21–36, Nov. 2018.
- [218] A. Hernandez, E. J. Yeo, S. Kim, and M. Chung, “Dysarthria detection and severity assessment using rhythm-based metrics,” in *Proc. of Annual Conference of the International Speech Communication*, Shanghai, China, Oct. 2020, pp. 2897–2901.
- [219] A. Hernandez, S. Kim, and M. Chung, “Prosody-based measures for automatic severity assessment of dysarthric speech,” *Applied Sciences*, vol. 10, no. 19, p. 6999, Jan. 2020.
- [220] H. Tong, H. Sharifzadeh, and I. McLoughlin, “Automatic assessment of dysarthric severity level using audio-video cross-modal approach in deep learning,” in *Proc. of Annual Conference of the International Speech Communication*, Shanghai, China, Oct. 2020, pp. 4786–4790.
- [221] A. A. Joshy and R. Rajan, “Dysarthria severity classification using multi-head attention and multi-task learning,” *Speech Communication*, vol. 147, pp. 1–11, Feb. 2023.
- [222] Y. Takashima, T. Takiguchi, and Y. Ariki, “End-to-end dysarthric speech recognition using multiple databases,” in *Proc. of International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Brighton, UK, May 2019, pp. 6395–6399.
- [223] B. Vachhani, C. Bhat, and S. K. Kopparapu, “Data augmentation using healthy speech for dysarthric speech recognition,” in *Proc. of Annual Conference of the International Speech Communication*, Hyderabad, India, Sept. 2018, pp. 471–475.
- [224] M. Geng, X. Xie, S. Liu, J. Yu, S. Hu, X. Liu, and H. Meng, “Investigation of data augmentation techniques for disordered speech recognition,” in *Proc. of Annual Conference of the International Speech Communication*, Shanghai, China, Oct. 2020, pp. 696–700.
- [225] F. Xiong, J. Barker, and H. Christensen, “Phonetic analysis of dysarthric speech tempo and applications to robust personalised dysarthric speech recognition,” in *Proc. of International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Brighton, UK, May. 2019, pp. 5836–5840.
- [226] W.-Z. Leung, M. Cross, A. Ragni, and S. Goetze, “Training data augmentation for dysarthric automatic speech recognition by text-to-dysarthric-speech synthesis,” in *Proc. of Annual Conference of the International Speech Communication*, Kos, Greece, June 2024, pp. 2494–2498.
- [227] E. Hermann and M. Magimai. Doss, “Few-shot dysarthric speech recognition with text-to-speech data augmentation,” in *Proc. of Annual Conference of the International Speech Communication*, Dublin, Ireland, Aug. 2023, pp. 156–160.
- [228] Z. Jin, M. Geng, X. Xie, J. Yu, S. Liu, X. Liu, and H. Meng, “Adversarial data augmentation for disordered speech recognition,” in *Proc. of Annual Conference of the International Speech Communication*, Brno, Czech Republic, Sept. 2021, pp. 4803–4807.
- [229] Z. Jin, M. Geng, J. Deng, T. Wang, S. Hu, G. Li, and X. Liu, “Personalized adversarial data augmentation for dysarthric and elderly speech recognition,” *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 32, pp. 413–429, Oct. 2024.
- [230] G. Dimauro, D. Caivano, V. Bevilacqua, F. Girardi, and

- V. Napolitano, "VoxTester, software for digital evaluation of speech changes in Parkinson disease," in *Proc. of International Symposium on Medical Measurements and Applications (MeMeA)*, Benevento, Italy, 2016, pp. 1–6.
- [231] S. A. Sheikh, Y. Kaloga, and I. Kodrasi, "Graph neural networks for Parkinsons disease detection," in *Proc. of International Conference on Acoustics, Speech and Signal Processing*, Hyderabad, India, April 2024, pp. 1–5.
- [232] Y. Kaloga, S. A. Sheikh, and I. Kodrasi, "Multiview canonical correlation analysis for automatic pathological speech detection," in *Proc. of International Conference on Acoustics, Speech and Signal Processing*, Hyderabad, India, April 2024, pp. 1–5.
- [233] A. Wisler, V. Berisha, A. Spanias, and J. Liss, "Noise robust dysarthric speech classification using domain adaptation," in *Proc. of Digital Media Industry & Academic Forum (DMI AF)*, Santorini, Greece, July 2016, pp. 135–138.
- [234] M. Amiri and I. Kodrasi, "Test-time adaptation for automatic pathological speech detection in noisy environments," in *Proc. of European Signal Processing Conference*, Lyon, France, Aug. 2024, pp. 86–90.
- [235] E. J. Ibarra, J. D. Arias-Londoño, M. Zañartu, and J. I. Godino-Llorente, "Towards a corpus (and language)-independent screening of Parkinson's disease from voice and speech through domain adaptation," *Bioengineering*, vol. 10, no. 11, p. 1316, 2023.
- [236] M. Amiri and I. Kodrasi, "Adversarial robustness analysis in automatic pathological speech detection approaches," in *Proc. of Annual Conference of the International Speech Communication*, Rhodes Islands, Greece, Sept. 2024, pp. 1415–1419.
- [237] B. Shi, W.-N. Hsu, K. Lakhotia, and A. Mohamed, "Learning audio-visual speech representation by masked multimodal cluster prediction," in *Proc. of International Conference on Learning Representations*, Virtual, April 2022.
- [238] C. Rudin, "Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead," *Nature Machine Intelligence*, vol. 1, no. 5, pp. 206–215, May 2019.
- [239] Y. Jiao, V. Berisha, and J. Liss, "Interpretable phonological features for clinical applications," in *Proc. of International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, New Orleans, LA, USA, Mar. 2017, pp. 5045–5049.
- [240] H. P. Rowe, S. E. Gutz, M. F. Maffei, and J. R. Green, "Acoustic-based articulatory phenotypes of Amyotrophic Lateral Sclerosis and Parkinson's disease: Towards an interpretable, hypothesis-driven framework of motor control," in *Proc. of Annual Conference of the International Speech Communication*, Shanghai, China, Oct. 2020, pp. 4816–4820.
- [241] R. Turrisi and L. Badino, "Interpretable dysarthric speaker adaptation based on optimal-transport," in *Proc. of Annual Conference of the International Speech Communication*, Incheon, Korea, Sep. 2022, pp. 26–30.
- [242] D. Gimeno-Gómez, C. Botelho, A. Pompili, A. Abad, and C.-D. Martínez-Hinarejos, "Unveiling interpretability in self-supervised speech representations for Parkinson's diagnosis," *IEEE Journal of Selected Topics in Signal Processing*, vol. 1, pp. 1–14, Feb. 2025.
- [243] L. Xu, J. Liss, and V. Berisha, "Dysarthria detection based on a deep learning model with a clinically-interpretable layer," *JASA Express Letters*, vol. 3, no. 1, p. 015201, Jan. 2023.
- [244] M. U. Hadi, Q. A. Tashi, A. Shah, R. Qureshi, A. Muneer, M. Irfan, A. Zafar, M. B. Shaikh, N. Akhtar, J. Wu, S. Mirjalili, and M. Shah, "Large language models: A comprehensive survey of its applications, challenges, limitations, and future prospects," *Authorea Preprints*, Aug. 2024.
- [245] D. Wagner, S. P. Bayerl, I. Baumann, K. Riedhammer, E. Nöth, and T. Bocklet, "Large language models for dysfluency detection in stuttered speech," in *Proc. of Annual Conference of the International Speech Communication*, Kos, Greece, Sept.

2024, pp. 5118–5122.



Shakeel A. Sheikh Shakeel A. Sheikh is currently working as a Data Science Innovation Research Scientist in the Oncology Data Science section at Novartis AG, Switzerland. His research focuses on the development and application of advanced data science and machine learning methodologies in oncology. Prior to this, he worked as a postdoctoral research scientist on the ChaSpeePro project at the IDIAP Research Institute, affiliated with EPFL (École Polytechnique Fédérale de Lausanne), Switzerland. He also held a postdoctoral position at the Cluster of Excellence Cognitive Interaction Technology (CITEC), Bielefeld University, Germany, in 2023. He earned his Ph.D. in 2023 from the MULTISPEECH team at LORIA-INRIA, Department of Informatics and Mathematics, Faculty of Sciences, Université de Lorraine, Nancy, France. His doctoral research focused on deep learning techniques for pathological speech. He obtained his M.S. in Informatics from Istanbul University, Turkey, in 2019, and holds a B.Tech in Computer Science and Engineering from the University of Kashmir, Jammu and Kashmir, India in 2015.



Md Sahidullah is an Assistant Professor in the Artificial Intelligence and Machine Learning group at the Institute for Advancing Intelligence, TCG CREST. His research interests lie in machine learning and speech/audio processing, with a focus on speech privacy and security, audio analytics for healthcare, and the development of voice-enabled technologies. He has over nine years of post-PhD experience in the field. He received his Ph.D. in Speech Processing from the Department of Electronics and Electrical Communication Engineering at the Indian Institute of Technology Kharagpur in 2015. He holds a B.E. in Electronics and Communication Engineering from Vidyasagar University (2004) and an M.E. in Computer Science and Engineering from the West Bengal University of Technology (2006). He was a postdoctoral researcher at the School of Computing, University of Eastern Finland (2014–2017), and later held a Starting Researcher position with the MULTISPEECH team at Inria Nancy – Grand Est, France (2018–2021). He has contributed to several national and international projects in Finland, France, and the EU. Since 2017, he has co-organized the ASVspoof Challenge, the leading international competition on audio deepfake detection. He has served on technical program committees for top conferences such as ICASSP and INTERSPEECH and is currently a member of the editorial boards of IEEE/ACM Transactions on Audio, Speech and Language Processing, IEEE Journal of Biomedical and Health Informatics, and Computer Speech & Language.



Ina Kodrasi (Senior Member, IEEE) received the M.Sc. degree in Communications, Systems, and Electronics from Jacobs University Bremen, Germany, in 2010, and the Ph.D. degree from the University of Oldenburg, Germany, in 2015. From 2015 to 2017, she was a Postdoctoral Researcher at the University of Oldenburg, working on multi-microphone speech dereverberation and noise reduction. Since December 2018, she has been with the Idiap Research Institute, Switzerland, where she leads the Signal Processing for Communication Group. Her research interests include signal processing, multichannel processing, pathological speech processing, and machine learning. Dr. Kodrasi received the ITG Best Paper Award in 2019 and the EURASIP Best Ph.D. Award in 2020. She has served as a member of the IEEE Signal Processing Society Technical Committee on Audio and Acoustic Signal Processing and the EURASIP Technical Area Committee on Acoustic, Speech and Music Signal Processing. Since 2023, she has been serving as an Editor for the IEEE/ACM Transactions on Audio, Speech, and Language Processing and the EURASIP Journal on Audio, Speech, and Music Processing.