Psychology Research and Behavior Management

Dovepress
open access to scientific and medical research

Open Access Full Text Article

REVIEW

Exploring the Role of Machine Learning in Diagnosing and Treating Speech Disorders: A Systematic Literature Review

Zaki Brahmi[1], Mohammad Mahyoob[2], Mohammed Al-Sarem[1], Jeehaan Algaraady[3], Khadija Bousselmi[4], Abdulaziz Alblwi[1]

[1]Department of Computer Science, Taibah University, Madina, Kingdom of Saudi Arabia; [2]Department of Languages and Translation, Taibah University, Madina, Kingdom of Saudi Arabia; [3]Department of English, Taiz University, Taiz, Yemen; [4]Department of Computer Science, LISTIC, University of Savoie Mont Blanc, Chambéry, France

Correspondence: Mohammad Mahyoob, Email mqassem@taibahu.edu.sa

Purpose: Speech disorders profoundly impact the overall quality of life by impeding social operations and hindering effective communication. This study addresses the gap in systematic reviews concerning machine learning-based assistive technology for individuals with speech disorders. The overarching purpose is to offer a comprehensive overview of the field through a Systematic Literature Review (SLR) and provide valuable insights into the landscape of ML-based solutions and related studies.
Methods: The research employs a systematic approach, utilizing a Systematic Literature Review (SLR) methodology. The study extensively examines the existing literature on machine learning-based assistive technology for speech disorders. Specific attention is given to ML techniques, characteristics of exploited datasets in the training phase, speaker languages, feature extraction techniques, and the features employed by ML algorithms.
Originality: This study contributes to the existing literature by systematically exploring the machine learning landscape in assistive technology for speech disorders. The originality lies in the focused investigation of ML-speech recognition for impaired speech disorder users over ten years (2014–2023). The emphasis on systematic research questions related to ML techniques, dataset characteristics, languages, feature extraction techniques, and feature sets adds a unique and comprehensive perspective to the current discourse.
Findings: The systematic literature review identifies significant trends and critical studies published between 2014 and 2023. In the analysis of the 65 papers from prestigious journals, support vector machines and neural networks (CNN, DNN) were the most utilized ML technique (20%, 16.92%), with the most studied disease being Dysarthria (35/65, 54% studies). Furthermore, an upsurge in using neural network-based architectures, mainly CNN and DNN, was observed after 2018. Almost half of the included studies were published between 2021 and 2022).
Keywords: speech disorder, speech recognition, dysarthria, machine learning, assistive technologies

Introduction

Humans are inherently social creatures, with an innate inclination towards engagement and interaction. In this context, speech as verbal messaging is a unique characteristic of humans, and it plays a leading role in humans' capacity to convey their thoughts, concerns, and perspectives to others[1]. However, individuals with speech impairments encounter significant academic, psychological, and social challenges while engaging with their communities.[2–4] The number of individuals with disabilities is continuously increasing. The World Health Organization states that about 1.3 billion people with a disability worldwide need assistive technology (AT)[5]. This number could increase by 2030 to about 2 billion people. The UN Convention on the Rights of Persons with Disabilities (UNCRPD) has confirmed AT provision as a fundamental human right.[6]

Received: 21 January 2024
Accepted: 7 May 2024
Published: 31 May 2024

Many interventions in the context of speech disorders are detected. Based on the causes of underlying speech disorders, some studies have provided treatment or assistance interventions for individuals with speech impairments, such as.[7–10] While others apply machine learning and deep learning methods to detect, classify, predict, and assess speech disorders; among them are[11–15] Machine Learning (ML) is a dominant branch of artificial intelligence (AI), covering remarkable advancements in research and industry. Machine learning showed notable impacts on improving communication tools for individuals with speech impairments as they enhance the accuracy and accessibility of speech recognition and word predictability, such as AI-driven speech-to-text and text-to-speech applications. Moreover, ML provides a host of powerful, automated algorithms designed to handle vast amounts of data across various disciplines like speech recognition[16,17] Natural Language Processing[18,19] human-computer interaction,[20] computer vision[21] health informatics,[22] recommender systems,[23] vocabulary context-aware prediction[24] and more.

Recent research demonstrated that deep signal analysis of voice using ML techniques to recognize speech with disorders showed promising results by extracting significant features from these signals, such as Mel-frequency Cepstral Coefficients (MFCCs) and Spectro Temporal utterances. Combining these two features shows more reliable results than others.[25–27]

Notably applying ML techniques in speech recognition and augmented communication, enhancing accessibility and user experience, predictive and contextual communication, voice synthesis, and personalized language models.[25,28] ML models can learn and adapt from ML-powered ATS users. Collecting, annotating, and analyzing large and diverse datasets of disordered speech samples would enable ML algorithms to identify specific users' speech patterns and nuances. That helps develop personalized models integrated into assistive devices, such as speech-generating or voice-recognition systems.[29,30] Moreover, ML approaches as data-driven approaches can play a valuable role in diagnosing and treating speech disorders.[31]

Despite the advantages of the SLR mentioned above, it is imperative to acknowledge the subsequent drawbacks. Most proposed SLRs focused on only one type of speech disorder, such as,[32,33] where only aphasia and Dysarthria are studied, respectively. Other SLRs[34,35] pay attention to one patient's age: children's age. In,[36] the focus is on assistive technologies used.

The present systematic literature review aimed to identify, categorize, and compare effective speech disorder detection for analyzing multiple speech disorders suitable for all age categories instead of choosing only a particular disorder or speech analysis tool as observed in the existing reviews. The proposed inclusive systematic review seeks to study the role of ML approaches in identifying, classifying, and evaluating these disorders. In addition, the study focuses on the ML role in treating these disorders, considering their potential causes, either biological, psychological, or environmental, regardless of the presence of cognitive impairments such as Down's syndrome[37] or Alzheimer's disease.[38,39] This work aims to comprehensively analyze the ML techniques for speech impairment recognition, focusing on the challenges and limitations. The primary contribution of our work implies:

- Providing a review highlighting the existing ML methods, algorithms, features extraction techniques, models' performance metrics, and the characteristics of the obtained datasets, focusing on discovering the state-of-The-art of all these techniques utilized by scholars in the field.
- Identifying the existing categories of speech disorders and clarifying how different ML approaches address these disorders.
- Shedding light on the limitations and challenges in the existing ML-based speech disorder detection, classification, and evaluation.
- Identifying gaps and potential opportunities for further research and improvements.

To achieve the review aims, we conducted the present systematic literature review following the guidelines outlined in the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) statement for a systematic review.[40] This protocol allowed us to carefully select the related studies, extract pertinent information, and present the results focusing on addressing the research questions. The rest of this paper is organized as follows: Background presents the background, containing the essential concepts related to our study. Research Methodology describes the methodology used to collect and select the articles studied and the research question. Discussion analyzes existing ML solutions for speech disorder recognition by answering the research questions. Section 5 presents new research directions to improve assistive technology for speech disorder users. In the final section, we conclude the paper by summarizing the work and highlighting future directions.

Background

This section will introduce the significant concepts pertinent to our research.

Speech Disorder

Speech is the central procedure of communicating thoughts, emotions, and ideas to others. It involves the sophisticated coordination of various body parts, such as the head, neck, and chest. This coordination is necessary for effective interaction. A speech disorder is a health condition that impairs a person's capability to utter words due to damage to muscles, nerves, or vocal structure. Speech disorders are complicated and varied conditions that can be shown in several forms, such as stuttering, Dysarthria, aphasia, Parkinson's Disease, Apraxia of speech, stammering, phonological disorders, and ataxia.

In the broader literature, the term "speech and language disorders" is categorized under communication disorders disability, alongside hearing disorders, deafness, and physical disabilities that impact speech, as depicted in Figure 1 [41] which presents the most known speech disorder:

- Dysarthria: considered as a physical disorder. According to the American Speech-Language-Hearing Association (ASHA),[42] Dysarthria is a motor speech disorder caused by muscle problems. It can make it hard to talk. As a sign of dysarthria, we can find speech that is too soft or too loud, the sound that is hoarse or breathy, etc.
- Aphasia is a type of language disorder affecting the ability to make or understand speech and read or write, as defined by the National Aphasia Association,[43] The cause of aphasia is always a brain injury, most often from a stroke, especially in older people.
- Dysophania: Dysphonia International describes Dysphonia as occurring when there is a change in the normal vocal tone, which could result from a structural or functional issue.[44] It is characterized by altered vocal quality, pitch, loudness, or vocal effort (shaky voice; rhythmic pitch and loudness undulations).

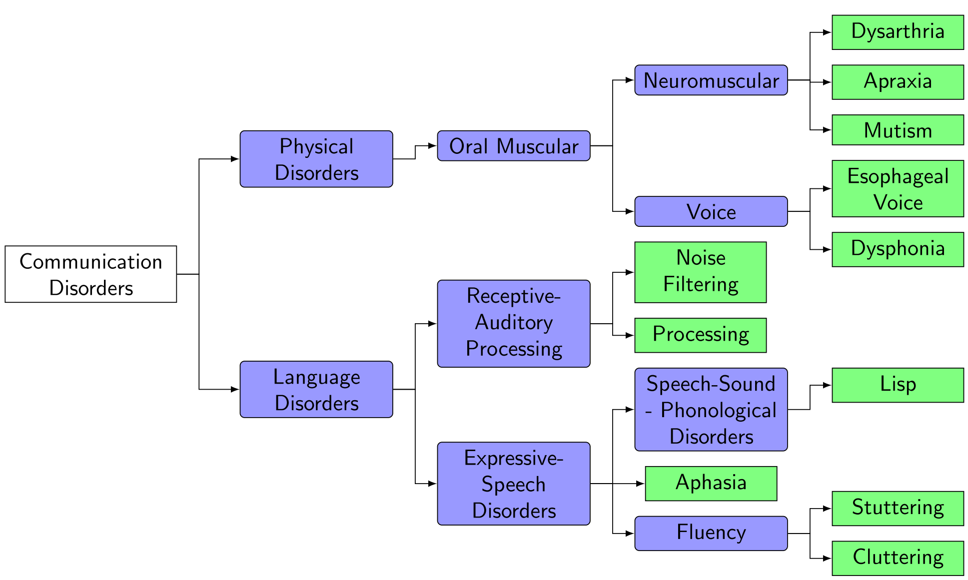

Figure 1 Speech disorder taxonomy.
Note: Adapted from Defining Speech and Language Disorders; 2023. Available from: https://speechandlanguagedisabilities.weebly.com/. Accessed December 11, 2023.[41]

- Parkinson's Disease: Parkinson's Foundation defines Parkinson's Disease as: "A neurodegenerative disorder that affects predominately the dopamine-producing ('dopaminergic') neurons in a specific area of the brain called substantia nigra". Parkinson's disease is characterized by hypokinetic dysarthria, which is featured by abnormalities like the inability to maintain loudness, monotonous and harsh voice, articulation errors, and reduced fluency.[45,46]
- Apraxia of speech: Apraxia is a neurological disorder in which people cannot do learned movements on command, even though they know what they are supposed to do and are willing to do it.[2,47] A patient with Apraxia of speech has difficulty moving their mouth in the way needed to produce sounds and words.
- Stammering/Stuttering: As defined by the British Stammering Association, it is a speech disorder that involves frequent and significant problems with normal fluency and flow of speech.[48] A symptom of stammering is a person can repeat sounds or words, stretch or prolong sound (eg "Hello fffffffreind"), and there is a silenced spot where a sound gets stuck.
- Phonological disorders: Also called speech-sound, it happens when people have trouble making certain sounds, even if there is no physical reason for the problem. Lisp is an example of this type of speech disorder. As an example of phonological disorder signs, the patient leaves off sounds from words, like saying "coo" instead of "school".

These disorders can substantially affect a person's transmission abilities and overall superiority of life. Seeking qualified help and treatment alternatives is crucial for handling and improving speech disorders. Unfortunately, the number of individuals with speech disorders continuously increased, as declared by the World Health Organization. Table 1 shows statistics on speech and language disorders around the world.

Machine Learning

Machine Learning (ML) models have emerged as valuable tools in speech disorders that have significantly empowered people with these disabilities through cutting-edge assistive technology solutions. AI and ML can analyze big data, identify patterns, make predictions, and imitate human cognitive functions.[54]

Machine learning, often abbreviated as ML, is the subfield of Artificial intelligence that intends to enable computers to learn from data and make predictions without being explicitly programmed. It is gaining more and more attention due to its significant role in many fields, including healthcare, manufacturing, finance, speech disorders, and more. It powers many technological advancements, like speech recognition, recommendation systems, self-driving cars, and predictive analytics. The main goal of machine learning is to build a model that performs well on both the training and test datasets. Data, comprising features and labels, is used for model training. During training, the model learns patterns and relationships between features and labels. The trained model is then tested on a separate dataset and used for inference. Machine learning algorithms can be classified into several categories, as illustrated in Figure 2:

Table 1 Statistics of Speech and Language Disorders

Country	Number	Type
General Statistics	10%[49] f the World's Population	Communication disorders
	Approximately 2 in 1000 were born with dysarthric adulthood	Dysarthric
	Approximately 5% of children, 3.5% of preschool-age children	Speech sound disorder
UK	180,000[49]	Aphasia
United States	2 millions[50]	Aphasia
Australia	1.2 million, according to Speech Pathology Australia[51]	Communication disability
South India	4.29% school-aged children[52]	Communication disorders
Spain	1.05% school-aged children[53]	Communication disorders

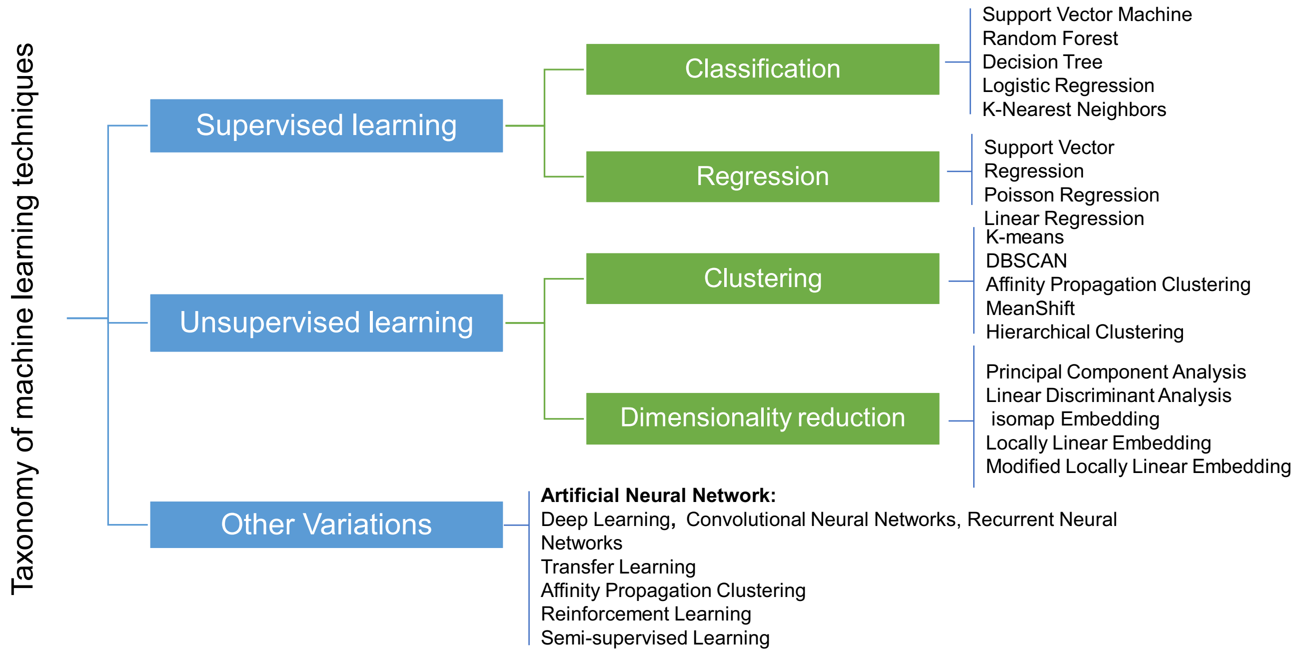

Figure 2 Machine learning taxonomy.

- Supervised Learning: In this type of learning, the algorithm is trained on a labelled dataset denoted as (X,y), where X represents the input features, and y represents the corresponding output labels. In supervised learning, the primary objective is to learn how to make predictions or classifications based on this labelled dataset. The regression and classification techniques are the primary techniques in this category.
- Unsupervised Learning: In this setting, the algorithms are trained on unlabeled datasets X. It aims to find patterns, groups, and structures within the datasets. Clustering and dimensionality reduction are commonly utilized techniques in this category.
- Other variations: This type of ML includes, but is not limited to, semi-supervised Learning and Reinforcement Learning. For instance, semi-supervised Learning is a hybrid method that combines supervised and unsupervised learning. Artificial neural networks are the most used in speech recognition.

Automatic Speech Recognition (ASR) for Speech Disorder

The main component of assistive technologies for people with speech disorder disabilities is Automatic speech recognition, which is the process by which a computer can recognize and act upon spoken language or utterance.[12] An ASR, as illustrated in Figure 3, can produce a text from a speech by analyzing and processing speech signals using different ML

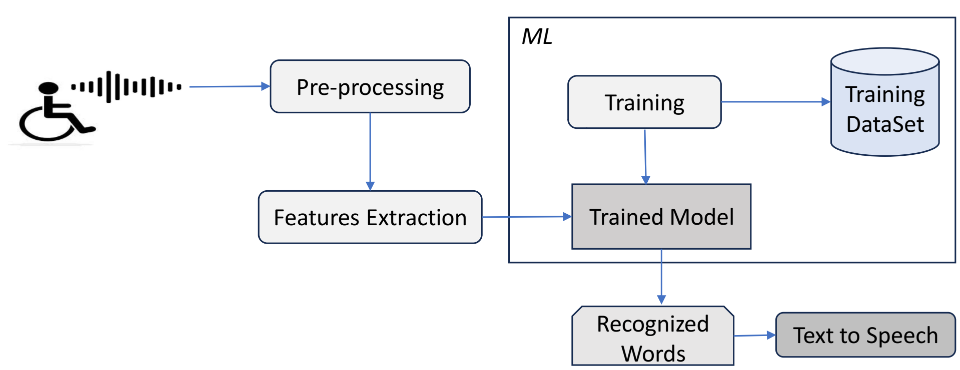

Figure 3 ASR Architecture.

techniques, such as Convolutional neural networks[29] or Deep Learning.[30] Indeed, the primary objective of the ASR system is to evaluate the various speech signals concerning various phonemes, syllables, and words/sentences. In the context of speech disorder, the patient can use an ASR to detect voice disorder and the voice pathologist to make an intelligent assessment of the patient.[55]

The performance of the ASR mainly depends on the training dataset, which is sorted into training and testing sets by randomly selecting the observations from healthy and sick voices.[56] The learning set is used to build the machine learning model, while the testing set is used to evaluate the final model's performance and generalization.[35] We need to process and turn the user's speech into a set of features to use ML algorithms.

Before extracting the features, the original speech signal has to go through preprocessing, which is the initial and most crucial step in the automatic speech recognition process. This step consists of cleaning the speech signal from ambient and undesirable noises, detecting speech activity, and normalizing the length of the vocal tract. The purpose of preprocessing a speech signal is to enhance the computational efficiency of speech recognition systems[57] by utilizing various preprocessing techniques, such as speech pre-emphasis, vocal tract length normalization, voice activity detection, noise removal, framing, and windowing.

The feature extraction procedure involves identifying the audio signal components that can be used to identify linguistic content while removing background noise and irrelevant information. In general, feature extraction is the process of generating the speech signal in digital form. Features can be mainly categorized into four categories: linguistic, contextual, acoustic, and hybrid. Various feature extraction techniques can be used in this first step, such as:

- Acoustic analysis measures the sound information in a speech to extract features related to phonation, articulation, Prosody, voice quality, etc.[13] For instance, articulation features can be vowel quality, coordination of laryngeal and supralaryngeal activity, precision of consonant articulation, tongue movement, occlusion weakening, and speech timing. Prosodic features can be pitch, loudness, and duration[58] In contrast, voice quality involves jitter, shimmer, first three formants, and harmonic-to-noise ratio.
- Mel-frequency cepstral coefficient (MFCC): used to represent the audio signal power spectrum and to record the timbral information of sounds.[59,60] The MFCCs are a set of coefficients that together form a Melfrequency cepstrum. MFCCs provide a suitable number of frequency channels to analyze audio, with only 12 parameters related to the amplitude of frequencies.
- Glottal Flow Signal: The glottal flow refers to the airflow that originates from the lungs and proceeds through the vocal folds in the larynx. The vocal folds vibrate, causing them to open and close periodically. An inverse filtering of the voice signal can obtain the glottal flow signal. Many parameters can be obtained from the glottal flow signal, but they are unsuitable for speech disorders.[61] Time-domain features are made by measuring how strong the speech signal gets over time. Time-domain features include energy, zero-crossing rate, pitch, and Linear predictive coding (LPC). Frequency-domain parameters where features are made up of the signals' frequency domain, also called its spectrum.[62]
- Spectro-temporal sparsity is mainly related to the diversity of disordered speech.[63] The main goal of the spectral features is to learn characteristics such as volume reduction, changes in format position, imprecise articulation, and hoarse voice. At the same time, the temporal features aim to capture patterns such as increased disfluencies and pauses.
- Discrete Wavelet Transform (DWT): Aiming to analyze non-stationary signals with multi-resolution potential, the Wavelet transform can be used as a time-frequency transform. DWT can do both the pathological voices' time and frequency domain analyses.[64] Thus, it is incredibly useful for detecting vocal issues.

One of the critical challenges in any ASR system is the number of features that can increase the cost of computation time and the system's performance. As a solution, feature selection can reduce the number of features by removing redundant and irrelevant features and boosting system performance. Many techniques of feature selection exist, such as Support Vector Machine-Recursive Feature Elimination (SVM-RFE)[65], minimum Redundancy Maximum Relevance (mRMR),[66] Chi-square,[67] and Principal Component Analysis (PCA), Local Learning-Based Feature Selection (LLBFS),[68] and Least

Absolute Shrinkage and Selection Operator (LASSO).[69] For instance, LASSO modifies the absolute value of feature coefficients; a feature with a coefficient that becomes zero will be removed from the set of features. Authors[70] use LASSO, LLBFS, Relief, and mRMR feature selection methods.

Related Work

Although many works have been proposed in the literature,[11–15,71,72] to our knowledge, few reviews have explored using AI and ML in identifying, predicting, and assessing different speech disorders. For instance,[32] surveyed existing works on automatic assessment systems designed to evaluate patients' aphasia and the severity level of patients. In another study,[33] a review focused on the characteristics of dysarthric speech and introduced assistive solutions like robust automatic speech recognition (ASR) systems. Meanwhile,[73] another study comprehensively analyzed the different voice disorders. They explored the existing machine learning (ML) approaches leveraged to develop automatic detection systems for voice disorders.

Additionally,[36] a systematic review delved into applying various ML models on the Internet of Things (IoT)-based Assistive Technology research. The study focused on the context of these models' applications and examined the IoT devices that cater to people's cognitive, hearing, visual, and degenerative diseases. In another systematic review,[34] studies involved in speech assessment methods for children and adolescents with different speech impairments were presented. A few ML-based approaches are presented in this study.

A systematic literature review of online speech therapy systems for childhood speech communication disorders is presented.[31] To compare these systems, authors used the following criteria: features of the proposed system, end user, used ML algorithm or not, and evaluation metrics. A SLR dedicated to an automatic Speech Recognition System for Tonal Languages is proposed in.[74]

Despite the advantages of the SLR, as mentioned above, it is imperative to acknowledge the subsequent drawbacks. It is clear to notice that the majority of proposed SLR focused on only one type of speech disorder, such as[32] and,[33] where only aphasia and Dysarthria are studied, or on a specific type of language, such as Tonal Languages in.[74] Other SLRs[34,35] pay attention to one patient's age. In,[36] the focus is on assistive technologies used as assessment technology for speech disorder patients.

Compared to the SLR, as mentioned earlier, the present systematic literature review aimed to identify, categorize, and compare effective speech assessment methods for analyzing multiple speech disorders suitable for all age categories with speech disorder disability instead of choosing only a particular disorder or speech analysis tool as observed in the existing reviews.

Research Methodology

This section outlines the methodology that was employed to conduct the subsequent investigation. Our research strategy was conducted in three phases. The first and second phases consisted of article selection processes by defining the paper selection strategy and research questions, and the third phase was data synthesis, as presented in the discussion section.

Research Strategy

This study will discuss and present a detailed exploration of machine learning (ML) and deep learning (DL) techniques in categorizing, recognizing, and predicting speech disorders. The study seeks to provide a detailed overview of the progressions in this discipline, shedding light on the various engaged feature extraction techniques, algorithms, datasets, and methodologies.

Based on our research questions, we defined principal keywords to research existing approaches in the literature dealing with this topic and obtained results from various databases. After gathering all the articles from the sources, we applied our filtering rules. First, we kept only recently published papers from the last ten years. Then, we filtered the results based on title, abstract, and Keywords. A second filter is applied to keep only papers published in reputable journals or international conferences classified as A or B. The overall process is shown in Figure 4.

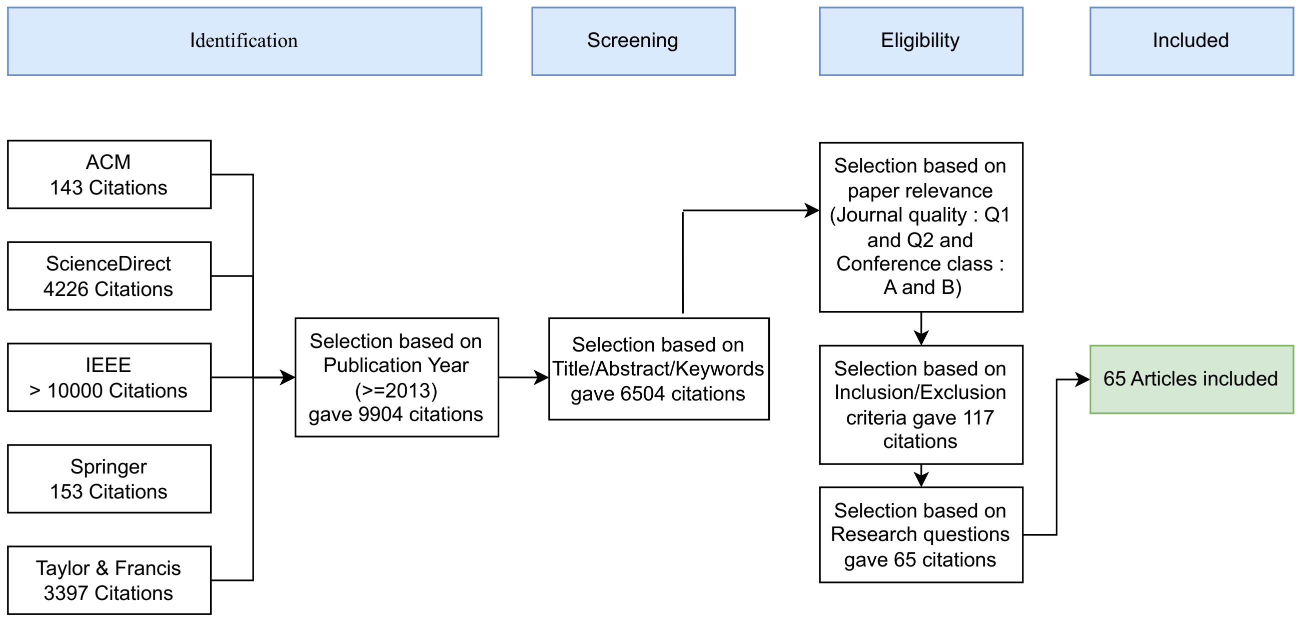

Figure 4 PRISMA protocol-based paper selection process.

Search Databases Selection

The databases used for our search are ACM,[75] IEEE,[75] ScienceDirect,[76] Springer,[77] and Taylor & Francis. Papers were selected based on their titles, abstracts, and keywords. We considered journal papers ranking Q1 or Q2 and conference papers of classes A or B. The number of papers obtained from each database at each step of our selection process according to the PRISMA protocol is discussed in the coming sections. We finally obtained 65 papers that were highly relevant from the different databases.

Search Keywords

After the final step of our paper selection process, we proceeded to full-text screening, which gave us the exact theme related to this survey. After filtering by title, abstract, and keywords, a final set of 65 papers was chosen to conduct this survey based on their relevant content. We used an automatic search method and used the following search strings using keywords set from Table 2: one instance of "keyword set 1" AND one instance of "keyword set 2". We also used combinations of search strings from the different lines of the keywords table using the "OR" operator. For example: "Speech disorders" OR "Assistive speech disorders" AND "ML".

Table 2 Search Keywords Definition

Keywords Set 1	Keywords Set 2
Speech disorders Assistive technology for speech disorders Speech impairment Voice disorder Dysarthria Parkinson's Disease, Aphasia Parkinson's Disease, Apraxia	Tool ML (Machine Learning) Software Expert system AI (Artificial Intelligence) Neural network, Deep learning, Contrastive learning, Support vector machine, Detection system, Convolutional Neural Network (CNN) Recurrent neural networks (RNNs)

Inclusive/Exclusive Criteria

The inclusion and exclusion criteria determine the systematic literature review's scope. They are first defined after deciding on the research issue and before accomplishing the search, but scoping searches could be necessary to choose pertinent inclusion and exclusion criteria. Various rules may be used to define these criteria, as shown in Table 3. A set of 65 papers was obtained after applying the criteria in Table 3.

Research Questions

This study will present a detailed exploration of the function of machine learning (ML) techniques in addressing speech disorders. The study seeks to provide a detailed overview of the progressions in this discipline, shedding light on the various engaged feature extraction methods, preprocessing techniques, datasets, and performance metrics. The underlying empirical question of this review is: What are the current ML algorithms? The study investigates the extent to which the models are comprehensive and inclusive for detecting and classifying speech disorders. All the scientific studies are synthesized to provide evidence for the following specific questions:

Q1: What details of the bibliographic profile are within the realm of existing studies?
Q1.1: What types of speech disorders are included in the existing studies?
Q1.2: How has the number of studies on this topic changed over the years?
Q1.3: What platforms (eg, journals, conferences, workshops) were selected by studies authors for dissemination?
Q2: What datasets and languages were used in the studies?
Q3: What preprocessing procedures are employed in constructing machine learning models?
Q4: What feature extraction and classification are prevalent in the studies?
Q5: What are the existing ML algorithms for speech disorder recognition?
Q6: What performance metrics have been used to gauge the efficacy of the proposed ML models?

These questions aim to thoroughly explore the ML field in speech disorders, focusing on various aspects like types of disorders studied, the evolution of the research over time, methodologies used, and the effectiveness of different approaches.

Discussion

In this section, we synthesize the analysis of the research papers proposing an ML-based solution for patients with speech disorders and provide the answers to the identified research questions. In total, the selected papers discussed in this paper are solutions.

Q1: What details of the bibliographic profile are within the realm of existing studies?
Q1.1: What types of speech disorders are included in the existing studies?

Of all the papers studied, we have distinguished several types of speech impairment problems: impaired vowel articulation, Dysarthria, Aphasia, Dysphonia, Apraxia of speech, stuttering, stammer, and Phonological disorders. We could classify these problems according to the number of papers dealing with them. Figure 5 summarizes the papers

Table 3 the Inclusive and Exclusive Search Criteria

Inclusion Criteria	Exclusion Criteria
Articles written in English	The full text is unavailable
The article is not relevant or related to speech disorders.	Studies related to other neurological problems
Journals with a good reputation or A/B conference papers	Studies about ASR systems not related to speech impairment

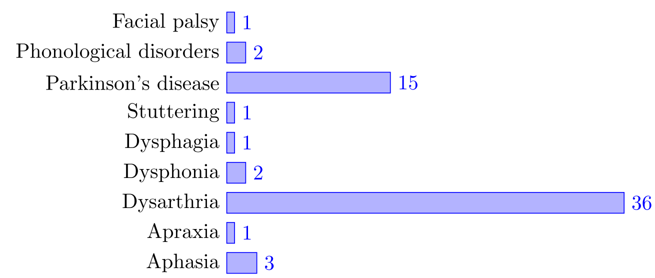

Figure 5 Number of papers per speech impairment type.

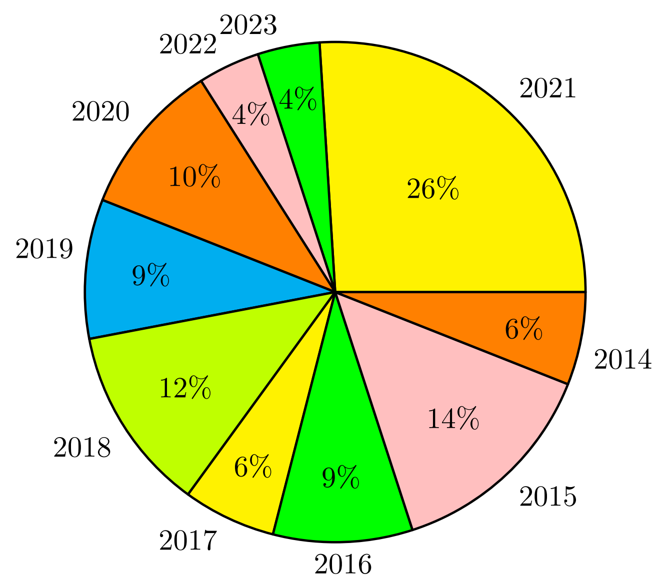

Figure 6 Classification of studied papers by year.

dealing with the same problem. We can conclude that most of the papers have focused on the problem of Dysarthia, as it is the most common speech impairment disorder. Some papers in this category were only focused on dysarthric people with Parkinson's disease. The other speech impairments were treated with fewer papers like Aphasia, Apraxia, Dysphonia, and Dysphagia. We classified the rest of the papers under "Speech impairment" as dealing with specific problems like imprecise vowel articulation or severe speech impairment problems.

Q1.2: How has the number of studies on this topic changed over the years?

According to Figure 6, we can conclude that the problem of speech disorders has recently gained increasing attention from researchers due to the significant technological progression in the development of automatic speech detection systems (ASR). Approximately a quarter of the papers were published in the most recent year, with a noticeable percentage in the previous two years. ASR systems have extended the horizons for new methods in dealing with persons with different speech impairments.

Q1.3: What platforms (eg, journals, conferences, workshops) were selected by the studies' authors for dissemination?

Table 4 Showing the Classification of Studied Papers by Year and Publication Source

Source	ACM	IEEE	Science Direct	Springer	Total
2014	1	1	0	2	4
2015	9	0	0	0	9
2016	3	1	1	1	6
2017	2	0	1	0	3
2018	3	0	3	1	7
2019	2	2	1	1	6
2020	2	0	1	3	6
2021	5	5	6	0	16
2022	4	0	0	0	4
2023	4	0	0	0	4
Total	35	9	13	8	65

We analyzed the publications' source by year and publication source, as shown in Table 4. According to this table, we can conclude that most of the publications were made at ACM, which publishes a well-known journal related to this issue, namely the "IEEE/ACM Transactions on Audio, Speech, and Language Processing".

Q2: What datasets and languages were used in the studies?

This section will provide insight into the datasets used in the literature, significantly impacting the investigation's precision and progress. In studied papers for speech disorder impairments, all datasets are mainly real-world datasets built from patients and healthy speakers. We can categorize these datasets into two categories: public and private. Public datasets are offered at publicly available sources such as TOROGO,[78] and UASpeech.[79] The second category contains datasets that, to be used, we need to contact authors such as,[80,81] and.[82] Table 5 shows a comparison of datasets used by papers. This comparison is mainly based on features of each dataset, such as speech disorders type, languages supported, and instance.

Table 5 Summary of Datasets Used by Retained Papers. M = Male, and F= Female

Dataset	Repository	Speech Disorders Type	Instances	Languages	References
AphasiaBank	Aphasia[83]	Aphasic	180 with and 140 without	English, Spanish, German, Italian, Hungarian, Mandarin, Chinese	[84]
PC-GITA[85]	-	Parkinson's disease	50 with (25 M, 25 F) and 50 without	Spanish	[26,55,86–90]
TORGO[78]	TORGO[91]	Dysarthria	7 With (4 M, 3 F): 2762 utterances, 5980 from healthy speakers	English	[14,55,86,89,92–95]

(*Continued*)

Table 5 (Continued).

Dataset	Repository	Speech Disorders Type	Instances	Languages	References
Finnish[55] PD		Parkinson's disease	35 with (14 male, 21 fe- male), and 32 without (12 male, 20 female)	Finnish	[55]
UA Speech[79]	UASpeech[96]	Dysarthria: cerebral palsy	15 with (4 M, 11 F), and 13 without. 765 isolated words per speaker	English	[26,89,97–103]
PDSTU[104]	-	Dysarthria	35 (14 M, 21 F) with, and 32 (12 M, 20 F) without	Finnish	[87]
BREF[105]	BREF[106]	Dysphonia	120 speakers (55 M, 65 F)	French	[15]
C2SI[107]	-	Dysarthria, Stuttering	94 with (51 M, 43 F), and 41 without (9 M, 32 F)	French	[15]
[72]	-	Dysphonia	23 with (13 F, 10 M), and 24 without. 224 voice phonation samples	English	[72]
Nemours[108]	-	Dysarthria	10 with (10 M, 0 F). 740 short nonsense	English	[100,109,110]
[80]	-	Dysarthria	8 with (4 M, 4 F), 1 without	French	[80]
DesPhoAPaDy[111]	-	Dysarthria, ataxia	89 with (39 F, 50 M), 29 without (14 F, 15 M)	French	[80]
[112]	-	Dysarthria	144 with (92 M, 52 F), and 30 (20 M, 10 F) without	Korean	[112]
[12]	-	Dysarthria	9 M and 1 F	Native Italian	[12]
[81]	-	Phonological disorders	100,000 audio files, collected from 1000 pronunciation assessments	Portuguese	[81]
[82]	-	Dysarthria	5 with, and 7 without	Central Thai (Siamese)	[82]
CCP	-	Phonological disorders	86 children (43 F, 43 M)	English	[113]
TYPALOC[114]	-	Dysarthria, Apaxia	28 with, and 12 (6 M, 6 F) without	French	[80]
SSNCE[115]	SSNCE[116]	Dysarthria	20 with (7 F, 13 M), and 10 without (5 F, 5 M). Each speaker recorded 365 utterances	Tamil	[117]
MoSpeeDi	MoSpeeDi[118]	Dysarthria	20 With (14 M, 6 F) and 20 Without (10 M, 10 F). 17 speakers, 5229 words	French	[26]
ATR[119]	-	Dysarthria	17 speakers, 5229 words	Japanese	[96]
TIMIT[120]	TIMIT[121]	Dysarthria	630 speakers. 3310 sentences	English	[111]
CMUArctic	Festvox[122]	Dysarthria	5 M, and 2 F. 1150 utterances	English	[123]
EMA[97]	-	Dysarthria	3 (1 M, 2 F), 680 utterances	English	[124]
IEMOCAP[125]	-	Dysarthria	1 M and 1 F. 3900 utterances	English: USA	[124]

(*Continued*)

Table 5 (Continued).

Dataset	Repository	Speech Disorders Type	Instances	Languages	References
Parkinsons[28]	Parkinsons[126]	Parkinson's disease	23 with, and 8 without	English	[127]
[70]	-	Parkinson's disease	20 with and 20 without. 1040 speech signals	Turkish	[70]
Spanish datasets[13]	-	Dysphagia	46 with (23 M, 23 F), and 46 without (23 M, 23 F)	Spanish	[13]
[11]	-	Aphasia	65 with (- M, - F), and 15 without (- M, - F)	English	[11]
Aphasic Speech Corpus[128]	-	Aphasia	17 with (11 M, 6 F), and 14 without (7 M, 7 F)	English	[128]
PC-GITA extended[129]		Parkinson's Disease	68 with (35 M, 33 F), and 50 without (25 M, 25 F)	Spanish	[129]
French Speech Corpus[130]		Facial palsy	24 with (16 M, 16 F), 8 without (4 M, 4 F)	French	[131]
Czech Speech Corpus[58]		Parkinson's Disease	46 without, 24 with (20 M, 4 F)	Czech	[58]

From Table 5, we can derive the subsequent observations:

- As we have examined various databases and classified them based on the languages of each database, it is evident that English is the predominant language, constituting over half of the data, followed by French at around 20%. Other languages like Spanish, Japanese, Italian, and Korean are represented to a lesser extent.
- Public datasets, including TOROGO and UASpeech, are frequently used in multiple research papers.
- The datasets primarily consist of recorded sentences from speakers with a balanced gender distribution. These speakers are typically divided into roughly equal groups of patients and healthy individuals, although there are exceptions in some datasets, such as ATR,[113] TIMIT,[129] and EMA.[132]
- Not all datasets are exclusively focused on dysarthria patients; some, like the EMA dataset,[132] are applicable in other areas.
- Regrettably, the Arabic language receives less attention, suggesting a scarcity of studies targeting Arabic-speaking individuals with speech disorders.

Q3, what preprocessing procedures are employed in constructing machine learning models?

Across the retained studies, researchers have used different pre-processing steps depending on the used dataset and the features they intend to employ. In general, the most used pre-processing techniques include:

- Normalization: This process involves normalizing audio signals to standard amplitude levels to ensure consistency and improve system durability. Methods used for normalization may involve scaling the signal within a specific range, such as between −1 and 1,[70,105] or employing z-score[131] normalization or techniques like peak normalization.
- Noise Removal / Filtering: Since Automatic Speech Recognition (ASR) systems are sensitive to ambient noise, adversely affecting recognition accuracy, noise reduction techniques are crucial. These techniques, including spectral subtraction or adaptive filtering, are applied to reduce the impact of ambient noise.

- Signal Segmentation: It allows the continuous audio stream to be broken into smaller segments, often based on pauses or other criteria, helps handle long audio recordings, and aligns the speech with linguistic units during recognition.
- Data Augmentation: It helps increase the training data's diversity and improve the model's robustness. Commonly applied augmentation methods include velocity perturbation, pitch shifting, time stretching/compression, noise injection, jitter, dynamic range compression, and room impulse response to simulate real-world conditions. Considering specific speech characteristics associated with Dysarthria, such as pitch, rate, and quality changes, these techniques help create a diverse and more representative training dataset, such as in.[100]
- Down-sampling recordings: This refers to the process of reducing the sampling rate of a recording. The sample rate represents the number of samples taken per second to represent a continuous audio signal digitally. Down sampling can benefit computational efficiency and resource usage by reducing the amount of information that needs to be processed, which can benefit training and inference speed, especially when working with large data sets, as in.[63,85,105]
- Signal alignment refers to synchronizing or aligning two or more signals in time. In ASR, signal alignment is often used to align the input speech signal with a reference or sample. This alignment ensures that the corresponding features or segments in the two signals match exactly, facilitating identification or comparison. Dynamic Time Warping (DTW) is a common technique to align signals.
- Signal transformation involves converting signals from one representation to another, which allows for extracting meaningful information or preparing data for analysis. Common transformations include the Fourier transform, which represents the frequency components of a signal; the Wavelet transform, suitable for analyzing signals with non-stationary characteristics; and the Mel Frequency Cepstral Coefficient (MFCC), widely used in speech processing.

In the context of speech impairment, we chose the following criteria to compare preprocessing features mentioned above:

- Effectiveness in capturing impairment characteristics: The technique can be effective in capturing speech impairment characteristics, but it usually has limited results in the literature. Effectiveness measurement is more detailed for studied references in section 5.6 using metrics like accuracy and error rates.
- Preservation of the input information (or signal): indicates whether the specified technique preserves the input signal entirely or partially by deforming a part.
- Computational Complexity: it depends on the used algorithms, but we indicate here the degree of complexity of the algorithms usually used for each technique as follows: linear (for O(n)), quadratic (for O(nn)) and quadratic (for O(nlog n)).

Table 6 shows the comparison of the preprocessing techniques and methods that were applied in the reviewed papers. Signal segmentation and alignment techniques were the most used, with 55% of the reviewed papers using them. Signal

Table 6 Comparison of Commonly Utilized Preprocessing Methods

Technique	Effectiveness	Preservation	Complexity	Reference
Normalization	Limited	Completely	Linear	[70,90,133]
Down-sampling Recordings	Limited	Partially	Linear	[12,63,131]
Signal Transformation	Limited	Completely	Linear to Quadratic	[63,131,133]
Signal Alignment	Effective	Completely	Linear to Quadratic	[26,63,71,128,131]
Data Augmentation	Effective	Partially	Linear to Quadratic	[71,102,117]
Signal Segmentation	Effective	Partially	Linear to Quadratic	[84,92,94,133,134]
Noise Removal/Filtering	Effective	Completely	Linear to Log-Linear	[81,135]

alignment, primarily through techniques such as Dynamic Time Warping (DTW), is essential in automatic speech recognition (ASR) for speech disorders because it corrects temporal irregularities and variations in speaking rate, allowing accurate comparison with reference signals. Signal transformation, such as using Mel-Frequency Cepstral Coefficients (MFCC) in,[63,85,105] is also essential to create informative feature representations that capture unique features of brain disorders. These preprocessing techniques improve the adaptability and robustness of ASR systems, allowing them to effectively recognize speech patterns affected by different types of impairments.

Normalization, down-sampling, and noise reduction are advantageous for treating speech impairment in ASR systems as they allow standardizing the amplitude of speech signals and ensuring their consistency by mitigating variations in loudness, contributing to a more uniform dataset. However, data augmentation is a crucial technique in this issue, typically in scenarios where the available data may be limited. Augmentation strategies can be tailored to reflect specific challenges posed by different impairments, making the model more robust and adaptable

Q4. What Feature Extraction Techniques are Prevalent in the Studies?

Feature extraction is crucial in automatic speech recognition (ASR) for speech disorders. This involves converting the raw audio signal into representative features that capture the information needed for recognition. Multiple feature extraction techniques are commonly used, emphasizing robust representation of speech disorders. Table 7 displays commonly employed methods in the studied works, the proportion of their utilization, and the research studies that relied on these methods.

More particularly, Mel frequency cepstral coefficients (MFCC) are frequently used in automatic speech recognition (ASR) systems to detect speech disorders due to their effectiveness in capturing essential characteristics of the signal voice, especially when something goes wrong. MFCCs mimic the sensitivity of the human auditory system to different frequencies, making them robust to variations in spectral characteristics caused by speech disorders. The ability to represent the power spectrum of speech signals in a compact and discriminatory manner makes MFCC well-suited for recognizing patterns associated with speech disorders. Additionally, MFCCs provide a good balance between capturing relevant information and reducing dimensionality, and thus, they are computationally efficient for use in ASR systems to detect speech disorders.

Table 7 shows that MFCCs-based features and Spectro-Temporal-based features constitute the most often utilized features among the researchers derived from the retained studies. The MFCC approach was used in 23 of 65 retained studies or 35.4% of the examined publications. In addition, we noted that the researchers relied heavily on spectrogram analysis in extracting the acoustic features of speech signals. Although the researchers found several ways and techniques for better detecting Dysarthria, analysis indicates that the combination of MFCC and Spectro Temporal utterance methods, such as,[87,94,124,138] achieved better accuracy than others.

Table 7 The Commonly Utilized Extraction Methods. In Some Cases, Studies Incorporate More Than One Technique. Following That, the Same Study Was Replicated, Thereby Increasing the Overall Number of Research Studies

Method/Technique	**%**	**Reference**
MFCCs/derived features from MFCCs	35.4%	[13,15,55,70,72,80,87,94,95,97–99,101,112,117,127,131,133,134]
Spectro-Temporal of utterances/keywords	26.15%	[12,14,15,26,63,72,87,89,90,97,101–103,110,112,124,136]
Articulation way/Speech timing	18.5%	[13,27,58,81,86–88,111,113,124,129]
Glottal flow	7.7%	[58,72,73,87,132]
Tongue movement/Phonation/ Speech quality	15.4%	[13,58,87,88,92,93,100,111,129,137]
Occlusion weakening	1.54%	[58]

Detecting and investigating how the patients articulated utterances or words has attracted the researcher, and we have noted many studies (18.5% of the reviewed studies). Articulation investigation involves assessing the accuracy and clarity of speech, which often occurs in speech disorders. Articulation evaluation techniques often involve analysis of formant frequencies, articulatory alignment patterns, and phonetic features extracted from speech signals. Speech timing, on the other hand, focuses on the temporal aspects of speech production and examines changes in speech rate and rhythm. Temporal features such as pause duration and speech rate are often used for speech timing analysis. The main advantage of articulatory assessment techniques is their ability to identify specific tonal distortions and inaccuracies associated with voice disorders.

Regarding speech timing, these techniques provide insight into irregularities in the temporal structure of speech that may indicate and characterize specific disorders. Combining these techniques in ASR improves the diagnostic potential of voice disorders and provides a more nuanced understanding of both articulatory accuracy and temporal dynamics in voice disorders.

Finally, other studies were based on alternate speech feature extraction methods, such as analysis of electromyography/MRI images, occlusion weakening and word features, and lexical diversity. The analysis of electromyography (EMG) and magnetic resonance imaging (MRI) images provides valuable insight into the physiological aspects of speech production and can help detect speech disorders related to muscle activity or anatomical structures. Occlusal weakness analysis, focusing on speech patterns in cases of partial impairment or weakness, can help identify Dysarthria. Integrating word features and measures of lexical diversity can reveal patterns associated with linguistic challenges and vocabulary limitations, improving the diagnostic power of ASR in such issues. Moreover, combining several techniques helps provide a comprehensive approach for detecting language impairment in ASR by including physiological, articulatory and linguistic aspects for a more accurate and nuanced assessment.

Q5. What are the existing ML algorithms for speech disorder recognition?

As most ASR approaches rely on ML techniques, we found that 67% of the studied papers used machine learning (ML) methods in their speech recognition approaches. The rest of the papers present surveys or recognition tools dedicated to people with different speech impairments, and they used existing ASR systems from the literature. The overall distribution of ML algorithms with the corresponding preprocessing and feature extraction methods according to the studied references is given in Table 8.

In Table 8, we have analyzed the different ML algorithms used. From this table, we can see that the most used algorithms are classifiers like SVM and LR. Support vector machines (SVMs) are widely used in automatic speech recognition (ASR) for speech disorders due to their effectiveness in classification tasks, especially in handling nonlinear

Table 8 Classification of References per ML Algorithm Used, Preprocessing and Feature Extraction Methods

Acronym	Algorithm	%	Preprocessing	Feature Extraction	References
Adaboost	Adaboost	1.54	Normalization	MFCC	[70]
CNN	Convolutional Neural Networks	16.92	Normalization	MFCC	[14,15,55,87,88,90,97,102,110,112,117,139]
DNN	Deep Neural Network	16.92	Down-sampling	MFCC	[12,27,71,72,81,82,92,93,113,117,131]
HMM	Hidden Markov Models	4.62	Down-sampling	MFCC/Phonation	[80,137,138]
KNN	k-Nearest Neighbors	4.62	Normalization	MFCC	[11,70,127]
LR	Linear Regression	6.15	Signal Alignment	Articulation way, Speech timing, Spectro-Temporal	[11,98,124,128]

(*Continued*)

Table 8 (Continued).

Acronym	Algorithm	%	Preprocessing	Feature Extraction	References
LRR	Linear Ridge Regression	1.54	-	Tongue movement, Phonation, Speech quality	[129]
MANN	Multi-nets Artificial Neural Networks	1.54	-	-	[101]
RF	Random Forests	4.62	Signal Alignment	Articulation way, Speech quality	[11,13,128]
RNN	Recurrent Neural Network	3.08	Spectro-Temporal	MFCC	[55,101]
SVD	Singular Value Decomposition	1.54	Signal Alignment	Spectro-Temporal	[26]
SVM	Support Vector Machines	20	All techniques	All techniques	[11,58,63,70,72,88,99,103,113,127,128,132,136]
LSR	Least Squares Regression	1.54	-	Spectro-Temporal	[124]
NB	Naive Bayes	3.08	Normalization	MFCC	[70,127]
MLP	Multilayer Perceptron	1.54	Normalization	MFCC	[70]
BRR	Bayesian Ridge Regression	1.54	-	Articulation Speech timing	[129]
DT	Decision Tree	1.54	-	-	[11]
KRR	Kernel Ridge	1.54	Normalization	MFCC	[129]
SVR	Support Vector	1.54	Normalization	MFCC	[129]

decision boundaries. SVM efficiently handles high-dimensional feature spaces, making it suitable for complex acoustic features extracted from impaired speech signals. In the context of dysarthria detection, SVM provides a robust framework for capturing complex patterns in data, allowing better classification of dysarthric speech. SVM has become a popular choice in ASR systems for speech disorders because it handles nonlinear relationships between features and can adapt to different impairment characteristics, contributing to improved accuracy and generalization.

ASR systems for speech impairment using SVM use common preprocessing methods such as normalization, filtering, and down-sampling to ensure consistent and efficient feature representation. Several feature extraction methods could be used with SVM, such as MFCC, Spectral-temporal keywords, and Speech timing, allowing the capture of the speech's relevant spectral and temporal characteristics. Because SVM operates in a high-dimensional space, it is adequate for classification tasks with complex patterns, such as distinguishing between normal and Dysarthric speech. This combination of preprocessing and feature extraction aims to create informative and discriminative feature vectors and optimize SVM performance for accurate speech recognition and fault detection.

For instance, in,[70] a machine learning system for diagnosing PD from speech signals is proposed to classify People with Parkinsonism from healthy ones. Six classifiers were used in this approach, ie, Adaboost, SVM, k-NN, MLP and NB. The experimental results indicated that SVM is the most successful classifier. In,[139] The goal is to detect PD patients by combining more than one symptom (rest tremor and voice degradation). Three classifiers were used: KNN, SVM, and NB. The majority vote technique is used to decide whether a person is PD. The proposed approach allowed them to achieve an accuracy of detection of PD up to 99.8% and confirmed that SVM classifiers take care of the outliers better than kNN and NB.

The approaches based on Artificial Neural Networks like CNN, DNN, GOP and MANN have been adopted to design a large set of ASR systems (33% of studied papers). Goodness Of Pronunciation (GOP) is a method based on both Convolutional Neural Network (CNN) and Deep Neural Network. Neural networks are widely valuable for ASR for speech disorders because they can automatically learn complex hierarchical representations from data, making them suitable for capturing complex patterns in impaired speech signals. More particularly, Convolutional neural networks (CNNs) effectively capture local patterns of spectro-temporal features, making them valuable for analyzing audio signals. Recurrent neural networks (RNNs) with sequential memory are good at modelling time dependencies and can help recognize subtle patterns in language disorders. Additionally, hybrid architectures such as GOP and MANN improve the alignment and decoding process.

With artificial neural network techniques, preprocessing often includes normalization, filtering, and down-sampling to ensure input consistency. Feature extraction methods usually used are MFCCs, and spectrogram plots provide helpful input to neural networks. The advantage lies in the ability of neural networks, especially deep learning architectures, to automatically extract hierarchical features and adapt to different characteristics of speech disorders. Also, the end-to-end learning approach minimizes the need for hand-crafted features and enables the model to recognize relevant patterns and nuances associated with different types of Dysarthria and speech impairment in general. The versatility and adaptability of neural networks make them powerful tools for ASR related to speech disorders.

For instance, inspired by the temporal processing mechanisms of the human auditory system, the authors of[131] proposed a deep learning-based dysarthric speech detection technique that separately processes the temporal envelope and fine structure signals. Two discriminative representations learned from the temporal envelope and fine structure using CNNs are then exploited for automatic dysarthric speech detection.[131] Other approaches combined the use of both traditional machine learning algorithms and ANN-based ones. In,[133] the authors combined the use of SVM and DNN to study the use of voice source information in detecting Parkinson's disease (PD) from speech using traditional pipelines and end-to-end. The traditional pipeline used SVM classifiers to classify the speech utterances into healthy or PD labels based on the extracted features. In an end-to-end approach, they trained Deep learning models on raw speech waveforms and voice source waveforms using convolutional layers and multilayer perceptron (CNN and MANN). The experimental results indicated that the SVM-based approach achieves up to 67.93% accuracy, while the CNN approach achieves 68.56%. Another hybrid approach was proposed in,[112] where authors propose a hybrid framework in which generative models are used for learning representation and discriminative models are used for classification.[112] The proposed approach outperformed conventional HMM and DNN-HMM-based approaches for various intelligibility levels.

Moreover, a multi-networks speech recognizer (DM-NSR) model is proposed in[87] using a realization of the multi-views multi-learners approach called multi-nets artificial neural networks (MANN). In particular, the DM-NSR model employs several ANNs to approximate the likelihood of ASR vocabulary words and deal with the complexity of dysarthric speech.[87] Authors trained 443 neural networks. For the speaker-independent ASR system, the DM-NSR recorded an average recognition rate of 15.69% and a decreased error rate of 6.25%.

Q6. What performance metrics have been used to gauge the efficacy of the proposed ML models?

As most studied papers used ML or Deep Learning methods for their proposed ASR approaches, Figure 7 shows that Accuracy is the most used performance metric. For 15.7% of works, they used Pearson correlation coefficient (PCC)[114] and Root mean square error (RMSE). Other evaluation metrics like sensitivity (which shows the ratio of correctly classified patients) and specificity (which indicates the percentage of correctly classified healthy people across the whole range of healthy people) were used.[70]

Common ML metrics used in the studied papers are assessed as follows:

- Accuracy: It measures the model's overall performance for all categories. It can be assessed using the following equation:

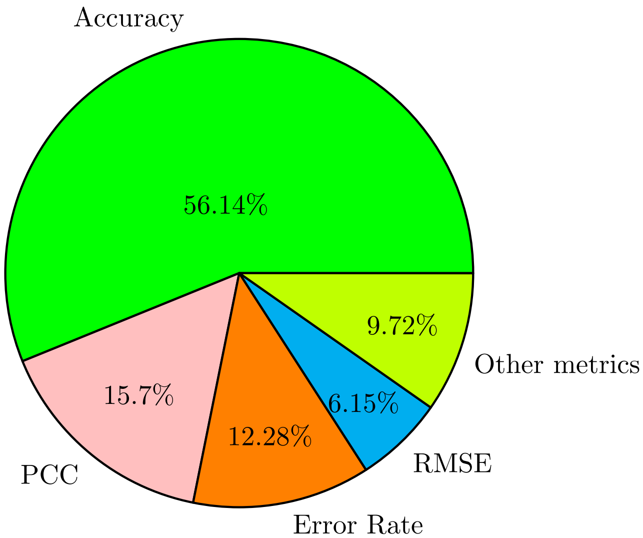

Figure 7 Performance Metrics Employed in Machine Learning-Based Approaches.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

with TP being true positive and TN being true negative. FN is false negative, and FP is false positive.

- Recall: is the proportion of all true positives predicted by the model divided by the total number of predicted values.[32] It can be evaluated using:

$$\text{Accuracy} = \frac{TP}{TP + FN} \quad (2)$$

- Precision: calculates the proportion of correctly identified positives as follows:[32]

$$\text{Accuracy} = \frac{TP}{TP + FP} \quad (3)$$

- F1-score is a summary of both recall and precision and can be assessed as:[57]

$$\text{F}_{1-\text{score}} = \frac{2*(Precision - Recall)}{(Precision - Recall)} \quad (4)$$

- PCC is a statistical method that calculates the correlation between two variables.[114]
  1. Root-mean-square deviation (RMSE) calculates the difference between the predicted values and the observed values as follows (eq. 5):[138]

$$RMSE = \sqrt{(\textstyle\sum_{i=1}^{N}(Predicted_i - Observed_i))/N} \quad (5)$$

Moreover, speech recognition accuracy metrics play a predominant role in evaluating the performance of any speech-assistive communication aid.[127] For 12.28% of the studies, they used Error Rate metrics like Word Error Rate (WER) or Sentence Error Rate or SER and phoneme Error Rate (PER) or Accuracy Recognition Rates like Word Recognition Accuracy(WRA) or Sentence Recognition Accuracy (SRA). For instance, the Accuracy Recognition Rate in ASR systems is evaluated as the number of correctly predicted words, sentences, or phonemes by persons out of all the test databases. For example, WRA can be assessed using the following formula:[127]

$$\mathrm{WRA} = \frac{NC * 100}{TC} \tag{6}$$

NC is the number of samples correctly recognized, and TC is the Total number of words per class. Word error rate (WER) is a common metric of the performance of a speech recognition or machine translation system. WER is the number of errors divided by the total words.[88]

In Table 9, we present the values of achieved accuracy in the studied works grouped by the dataset used with other assessed metrics.

We note from Table 9 that the larger the dataset size, the better the value of achieved accuracy, such as for the UA-Speech Database and the TORGO database.[78] The same observation is valid for large corpus specific to some languages, like in[120] using an Aphasia Bank corpus for 78 persons and in[124] using a specific Korean Dysarthric dataset for 174 persons. Some works were restricted to a specific language with limited datasets, like in,[12,107,123] but authors show that using techniques such as transfer learning could help generalize their approaches for multiple languages and

Table 9 Performance Metrics Values per Reference

Accuracy Value per Reference	Other Metrics	Target Database
83%[63] 80%[58]		Spanish dysarthric dataset[63]
70% for sentences96	WER of 35.3%,[97] PER of 33.3%[95]	TORGO[78]
96.3%[26]		
94.31%[99]		
80.83%[101]	WER of 30.3% 70,[97]	UASpeech[79]
64.71%[102]		
89%[98]		
67.93% for ML[88]	PCC of 78% and RMSE of 11.7 71,[87]	PDST[104]
68.56% for DL[88]	2.63% absolute (8.63% relative WER)[89]	PC-GITA[85]
82%[26]	16.4% of RMSE[97]	
89%[90]	RMSE of 16.9%[90]	
81.4%[15]		BREF[105]
72.6%[15]		C2SI[107]
95.32%[72]		C2SI[107]
20% raise for WRA,[110] 89%[98]		Nemours[108]
76% of WRA, 59% of SRA[140] Up to 98%[12]	Precision (0.88), recall (0.64),[98] WER of 9.7%, Cmd ER of 14.9%[134]	DesPhoAPaDy[111] + private corpus,[80] Other private datasets[140]
>80%[128] 94.5%[82]	RMSE <0.5[128]	
94.5%[82] >90%[94]		AphasiaBank[83]

(*Continued*)

Table 9 (Continued).

Accuracy Value per Reference	Other Metrics	Target Database
83%[11]	F-score: 80%[11] PCC: 96%[112]	Private dataset for Aphasia[11] Korean Dysarthric Dataset[112]
90%,[13] 93%[80]	WER of 26.76%[117] PCC of 95%[100]	Spanish Dysphasia datasets[13,14] SSNCE[115] UASpeech[79]
80.5%[26]		MoSpeeDi[141]

achieve better performances, such as in.[14,84,94,128] Moreover, using standard datasets for Dysarthria persons helps compare experiment results with other works and better evaluate the techniques used.[26,87,98,112]

In summary, in automatic speech recognition (ASR) systems for speech disorders, hybrid architectures such as convolutional neural networks (CNNs), recurrent neural networks (RNNs), and connectionist temporal classification (CTC) have demonstrated superior performance in terms of achieved accuracy values. CNNs are effective for feature extraction because they are exceptionally efficient at capturing local patterns in spectro-temporal features. RNNs, particularly long short-term memory (LSTM) networks, can model temporal dependencies necessary for subtle speech patterns. Hybrid models like GOP combine the strengths of CNNs and RNNs to improve the alignment and decoding process. Also, traditional methods such as Support Vector Machines (SVMs) and Hidden Markov Models (HMMs) may not match the flexibility and adaptability of neural networks in capturing complex hierarchical representations, limiting their performance in specific contexts. However, they remain robust at handling high-dimensional feature spaces, allowing them to capture complex data patterns better to classify Dysarthric speech.

Gaps and Future Directions

Applying the ML and DL techniques for implementing speech disorders classification, recognition, and prediction models has demonstrated notable improvements in performance accuracy. However, some gaps still need to be addressed as potential avenues for future exploration.

Annotated Datasets

One significant gap lies in the limited availability of large, balanced, and high-quality annotated datasets for training and evaluating machine learning models for speech disorders.[7,11,12,102,105,117,120,123,138] The scarcity of such datasets limits the generalizability and validity of the results, leading to increased ambiguity and reduced statistical control. Moreover, it hinders the ability to draw strong inferences and develop and evaluate robust algorithms. To develop a robust model, expanding the current state-of-The-art annotation values could be done by adding more human expert annotators, creating a model that would learn an outline according to a selection system, and reproducing several experts' choices. The less the divergence rate among annotators is, the more efficient the produced device is. Employing post-processing techniques would also enhance the model's robustness.[112]

Diversity

Another limitation is the lack of diversity in the accessible datasets, which hinders the development of more comprehensive and inclusive models. Most available datasets focused on specific languages and communities such as English, French, Spanish, Italian, German, Persian, Korean, Croatian, Indian languages, etc. However, only two datasets provide some level of diversity: Aphasia Bank 25 a. Aphasia Bank comprises English, Croatian, French, Italian, Mandarin, Romanian, and Spanish data. TORGO dataset[78] covers English and Italian languages. Using one speech database may not represent the diversity of speech disorders and languages and may limit the generalizability and comparison of the results with other types of speech disorders languages.[117,120,134,138] Moreover, there is a lack of generalization of the machine learning models to other datasets or populations;[11] moreover, a lack of comparison between the proposed and existing models for performance, intelligibility, and speech capability loss assessment or accuracy rate prediction.[11,85,105,135,142] More practice-driven research is required to create standard and large datasets as a feasible approach for researchers to compare different methodologies and techniques, leading to improved results.

Model

In most proposed models, the disordered speech attribute features are based on a binary representation of phonological and phonetic characteristics, which may not capture the fine-grained variations in articulation quality. Moreover, these features are sensitive to noise and recording quality, which may affect the accuracy of anomaly detection and localization and limit their applicability in real-world scenarios.[80,120,133] Most of the proposed models focused on detecting articulation, and much work is still needed for novel models to predict the correct word.[140] More effective models are required while dealing with the extreme variability of speech due to its complex nature.[26] These models need to be able to categorize types and severity of speech disorders such as dysarthric or aphasic speech into multiple categories instead of binary classification.[80,128,135,139,140] However, there are various metrics for evaluating the proposed models' performance commonly including accuracy in,[11,12,15,58,63,72,87,90,98,111,128] F1-score in,[14,111] Confidence RMSE in,[87,131] Sensitivity, and Specificity in.[70] These metrics may not entirely capture the performance of machine learning models in real-world circumstances. There is a need to develop new comprehensive evaluation metrics that are more standardized, allow for better comparison of different models, and capture the nuances and complexities of speech disorders.

Disordered Speech Features

Many studies have claimed that extracting practical acoustic and phonological features is still challenging. Disordered speech features are complex and need special tools and experts in the domain to check and revise their suitability.[125,135,142] Experts should carefully annotate these datasets by extracting phonemic and allophonic features to ensure accurate and reliable training of machine learning models. The Arabic language possesses distinct phonetic sounds, including pharyngeal, larynx, and uvula sounds, often overlooked in speech disorder research. To address this gap, we intend to employ machine learning models to tackle these challenges head-on. Our future studies will be dedicated to exploring the unique phonological characteristics of Arabic speech disorders. Doing so aims to contribute to a more comprehensive understanding of these disorders and pave the way for effective interventions.

Time and Privacy

Most studied approaches focus on the speech recognition phase to recognize wrong keywords, with no focus on deploying such solutions to the final users. Indeed, user profiles are heterogeneous, implying heterogeneity of used datasets for the training step. Even though getting good accuracy from the trained models is related to having a large and shared dataset, this can affect the assistive application's running time, which needs to operate in real-time. Another issue that can be accurate when using a shared dataset is the privacy concern related to speech disorder patients. As a new direction, overcoming these issues using Federate Learning (FL) techniques can be a promising solution. FL is a machine learning technique recently proposed by Google,[28] which is crucial in the era of ubiquitous computing, where massive IoT devices continuously generate relevant data that cannot be easily shared due to privacy and communication constraints. FL is an effective solution for training machine learning models on the growing amount of data while keeping data locale, allowing multiple clients to jointly train a learning model on their private data without revealing their local data to a centralized server.[130] This can be useful in building an adaptable, trained model for the end-user and ensuring the privacy of sensitive data.

Context of Using the Speech Assessment System

To the best of our knowledge, to predict the correct words, all studied solutions are based on a trained model that is chiefly dependent on the quality of training datasets. To enhance the output efficiency of a speech assessment system for the speech impaired, exploiting the context of the conversation between the patient and his interlocutor can be a new research direction. The context of a conversation can be, for example, the subject of the conversation, the psychological state of the interlocutors, the place and time of the conversation, etc. Detecting and using the context is a challenging task that needs more investigation.

Output of a Speech Assessment System

To our knowledge, all studied solutions mainly focused on detecting incorrect words to propose the correct ones to the end user. In real situations, speech disorder users can pronounce unclear words and sentences with unclear meanings. Dealing with this situation is a challenging task. A large language model such as ChatGPT can be a good direction for predicting a sentence.

Conclusion and Future Work

This work synthesizes and analyses research papers proposing ML-based solutions for speech disorder patients. For this end, we considered a specific number of databases, journals, conferences, and articles published between 2013 and 2023. However, the review only focuses on ML-based papers proposing assistive solutions for people with speech disorder disabilities.

ML-based assistive systems for people with speech disorders are a promising solution for better enhancing the quality of life of these people through communication networking. This work conducted an SLR on ML-based speech disorder assessment systems, aiming to provide a comprehensive understanding of the main issues related to this problem, feature extraction techniques, and ML algorithms used. Our goal was to find out how far we have come and give advice for future research on speech disorders problems. We hope this work will help researchers understand this vital research topic and continue their research.

In the future, we will contribute to this field by considering different issues and trends to follow. Our future work includes two directions. The first direction is to compare different ML-based assistive solutions for speech disorder patients in depth (experimental comparison) to detect their weaknesses and find solutions to remedy them. This will also help academics and practitioners understand how to handle the problem better. Secondly, we aim to propose a new support solution for users suffering from speech disorders by considering the limitations we have detected in existing approaches, such as the limitation of execution time, privacy preservation, diversity of datasets, and the treatment of the Arabic language. This solution will be based mainly on federated learning techniques. Another direction is to consider the user's emotional state and the subject of the conversation to enhance any assistive solution for users with speech disorders.

Acknowledgment

The authors extend their appreciation to the King Salman Center for Disability Research for funding this work.

Funding

This research was funded by King Salman Center for Disability Research, grant number KSRG −2023-542.

Disclosure

The authors declare no conflicts of interest in this work.

References

1. Pagel M. Q&A: what is human language, when did it evolve and why should we care? *BMC Biol.* 2017;15:1–6. doi:10.1186/s12915-017-0405-3
2. McGregor KK, Langenfeld N, Van Horne S, Oleson J, Anson M, Jacobson W. The university experiences of students with learning disabilities. *Learn Disabil Res Pract*. 2016;31:90–102. doi:10.1111/ldrp.12102
3. Norbury CF, Paul R. Disorders of speech, language, and communication. *Rutter's Child Adoles Psych*. 2015;2015:683–701.
4. McCormack J, McLeod S, McAllister L, Harrison LJ. A systematic review of the association between childhood speech impairment and participation across the lifespan. *Internat J Speech*. 2009;11:155–170. doi:10.1080/17549500802676859
5. Disability. World Health Oragnization; 2023. Available from: https://www.who.int/news-room/fact-sheets/detail/disability-and-health. Accessed May 17, 2024.
6. Hendriks A. UN convention on the rights of persons with disabilities. *Eur J Health Law.* 2007;14:273–298. doi:10.1163/092902707X240620

7. Hawley MS, Cunningham SP, Green PD, et al. A voice-input voice-output communication aid for people with severe speech impairment. *IEEE Trans Neural Syst Rehabil Eng*. 2012;21:23–31. doi:10.1109/TNSRE.2012.2209678
8. Hair A, Monroe P, Ahmed B, Ballard KJ, Gutierrez-Osuna R. Apraxia world: a speech therapy game for children with speech sound disorders. In Proceedings of the Proceedings of the 17th ACM Conference on Interaction Design and Children; 2018: 119–131.
9. Attwell GA, Bennin KE, Tekinerdogan B. Reference architecture design for computer-based speech therapy systems. *Comput. Speech Lang.* 2023;78:101465. doi:10.1016/j.csl.2022.101465
10. Wang P, Van Hamme H. Benefits of pre-trained mono-and cross-lingual speech representations for spoken language understanding of Dutch dysarthric speech. *EURASIP J Aud Spe Music Process*. 2023;2023:1–25. doi:10.1186/s13636-023-00280-z
11. Gu Y, Bahrani M, Billot A, et al. A machine learning approach for predicting post-stroke aphasia recovery: a pilot study. In Proceedings of the Proceedings of the 13th ACM International Conference on PErvasive Technologies Related to Assistive Environments; 2020:1–9. doi:10.1145/3389189.3389204.
12. Mulfari D, Meoni G, Marini M, Fanucci L. Machine learning assistive application for users with speech disorders. *Appl. Soft Comput.* 2021;103:107147. doi:10.1016/j.asoc.2021.107147
13. Roldan-Vasco S, Orozco-Duque A, Suarez-Escudero JC, Orozco-Arroyave JR. Machine learning based analysis of speech dimensions in functional oropharyngeal dysphagia. *Comput Methods Programs Biomed.* 2021;208:106248. doi:10.1016/j.cmpb.2021.106248
14. Sekhar SM, Kashyap G, Bhansali A, et al. Dysarthric-speech detection using transfer learning with convolutional neural networks. *ICT Express*. 2022;8:61–64. doi:10.1016/j.icte.2021.07.004
15. Abderrazek S, Fredouille C, Ghio A, Lalain M, Meunier C, Woisard V. Interpreting deep representations of phonetic features via neuro-based concept detector: application to speech disorders due to head and neck cancer. *IEEE/ACM Trans Audio Speech Lang Process*. 2023;31:200–214. doi:10.1109/TASLP.2022.3221039
16. Vashisht V, Kumar Pandey A, Prakash Yadav S. "Speech recognition using machine learning IEIE Transactions on Smart Processing & Computing 10.3; 2021:233–239.
17. Zhang Y, Pezeshki M, Brakel P, Zhang S, Laurent Yoshua Bengio C, Courville A. "Towards end-to-end speech recognition with deep convolutional neural networks. *arXiv Preprint, arXiv.* 2017;2017:1.
18. Ayanouz S, Anouar abdelhakim B, Benhmed M. A smart chatbot architecture based NLP and machine learning for health care assistance. In: Proceedings of the 3rd international conference on networking, information systems & security; 2020:1–6.
19. Qin L, Minheng N, Zhang Y, Che W. "Cosda-mL: multi-lingual code-switching data augmentation for zero-shot cross-lingual nlp. *arXiv Preprint, arXiv.* 2020;2020:1.
20. Zhang A. Human computer interaction system for teacher-student interaction model using machine learning. *Internat J Hum Comp Interact*. 2022;2022:1–12.
21. Esteva A, Chou K, Yeung S, et al. Deep learning-enabled medical computer vision. *Npj Digital Med*. 2021;4:5. doi:10.1038/s41746-020-00376-2
22. Tyagi AK, Mannoj Nair M. Deep learning for clinical and health informatics. Computational Analysis and Deep Learning for Medical Care: Principles, Methods, and Applications; 2021:107–129.
23. Yanes N. A machine learning-based recommender system for improving students learning experiences. IEEE Access 8; 2020:201218–201235.
24. Zhang B. Integrating an attention mechanism and convolution collaborative filtering for document context-aware rating prediction. IEEE Access 7; 2018: 3826–3835.
25. Jefferson M. Usability of automatic speech recognition systems for individuals with speech disorders: past, present, future, and a proposed model; 2019.
26. Janbakhshi P, Kodrasi I, Bourlard H. Subspace-based learning for automatic dysarthric speech detection. *IEEE Signal Process Lett*. 2020;28:96–100. doi:10.1109/LSP.2020.3044503
27. Tripathi A, Bhosale S, Kopparapu SK. Automatic speaker independent dysarthric speech intelligibility assessment system. *Comput. Speech Lang.* 2021;69:101213. doi:10.1016/j.csl.2021.101213
28. McMahan B, Moore E, Ramage D, Hampson S, Arcas BA Communication-efficient learning of deep networks from decentralized data. In Proceedings of the Artificial intelligence and statistics. PMLR, 2017, pp. 1273–1282.
29. Sitaula C, He J, Priyadarshi A, et al. Neonatal bowel sound detection using convolutional neural network and Laplace hidden semi-Markov model. *IEEE/ACM Trans Audio Speech Lang Process*. 2022;30:1853–1864. doi:10.1109/TASLP.2022.3178225
30. Subramanian AS, Weng C, Watanabe S, Yu M, Yu D. Deep learning based multi-source localization with source splitting and its effectiveness in multi-talker speech recognition. *Comput. Speech Lang.* 2022;75:101360. doi:10.1016/j.csl.2022.101360
31. Landrigan J-F, Zhang F, Mirman D. A data-driven approach to post-stroke aphasia classification and lesion-based prediction. *Brain*. 2021;144:1372–1383. doi:10.1093/brain/awab010
32. Jothi K, Mamatha V A systematic review of machine learning based automatic speech assessment system to evaluate speech impairment. In Proceedings of the 3rd International Conference on Intelligent Sustainable Systems (ICISS). IEEE, 2020, pp. 175–185.
33. Bharti K, Das PK A Survey on ASR Systems for Dysarthric Speech. In Proceedings of the 2022 4th International Conference on Artificial Intelligence and Speech Technology (AIST). IEEE, 2022, pp. 1–6.
34. Usha GP, Alex JSR. Speech assessment tool methods for speech impaired children: a systematic literature review on the state-of-The-art in Speech impairment analysis. *Multimedia Tools Appl*. 2023;1–38. doi:10.1007/s11042-023-14913-0
35. Attwell GA, Bennin KE, Tekinerdogan B. A systematic review of online speech therapy systems for intervention in childhood speech communication disorders. *Sensors*. 2022;22:9713. doi:10.3390/s22249713
36. de Freitas MP, Piai VA, Farias RH, Fernandes AM, de Moraes Rossetto AG, Leithardt VRQ. Artificial intelligence of things applied to assistive technology: a systematic literature review. *Sensors*. 2022;22:8531. doi:10.3390/s22218531
37. Smith E, Hokstad S, Næss KAB. Children with Down syndrome can benefit from language interventions; Results from a systematic review and meta-analysis. *J Communic Dis*. 2020;85:105992. doi:10.1016/j.jcomdis.2020.105992
38. Cera ML, Ortiz KZ, Bertolucci PHF, Tsujimoto T, Minett T. Speech and phonological impairment across Alzheimer's disease severity. *J Communic Dis*. 2023;105:106364. doi:10.1016/j.jcomdis.2023.106364
39. Resende EDPF, Nolan AL, Petersen C, et al. Language and spatial dysfunction in Alzheimer disease with white matter thorn-shaped astrocytes. *Neurology.* 2020;94:e1353–e1364. doi:10.1212/WNL.0000000000008937

40. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *Int j Surg*. 2021;88:105906. doi:10.1016/j.ijsu.2021.105906
41. Defining Speech and Language Disorders; 2023. Available from: https://speechandlanguagedisabilities.weebly.com/. Accessed December 11, 2023.
42. Dysarthria. American Speech-Language-Hearing Association; 2024. Available from: https://www.asha.org/public/speech/disorders/dysarthria/. Accessed May 17, 2024.
43. What is Aphasia? National Aphasia Association; 2024. Available from: https://www.aphasia.org/aphasia-definitions/. Accessed May 17, 2024.
44. Voice impairment has many causes. Dysphonia International. Available from: https://dysphonia.org/voice-conditions/overview-of-vocal-disorders/; 2023. Accessed May 17, 2023.
45. Sachin S, Sachin S, Shukla G, et al. Clinical speech impairment in Parkinson *Parkinson's disease*, progressive supranuclear palsy, and multiple system atrophy. *Neurol India*. 2008;56:122–126. doi:10.4103/0028-3886.41987
46. What-is-parkinsons. Parkinson's Foundation; 2024. Available from: https://www.parkinson.org/understanding-parkinsons/what-is-parkinsons. Accessed May 10, 2024.
47. Apaxia. National Organization for Rare Disorders; 2003. Available from: https://rarediseases.org/rare-diseases/apraxia. Accessed May 17, 2024.
48. What is Stammering. Stamma; 2024. Available from: https://stamma.org/about-stammering/stammering-facts/what-is-stammering. Accessed May 17, 2024.
49. Aphasia. National Institute on Deafness and Other Communication Disorders; 2017. Available from: https://www.nidcd.nih.gov/health/aphasia. Accessed May 17, 2024.
50. Quick Statistics About Voice, Speech, Language. National Institute on Deafness and Other Communication Disorders; 2024. Available from: https://www.nidcd.nih.gov/health/statistics/quick-statistics-voice-speech-language. Accessed May 17, 2024.
51. Speech And Language Disorders Statistics. Gitnux; 2023. Available from: https://blog.gitnux.com/speech-and-language-disorders-statistics. Accessed May 17, 2024.
52. Ravi SK, Sumanth P, Saraswathi T, Chinoor MAB, Ashwini N, Ahemed E. Prevalence of communication disorders among school children in Ballari, South India: a cross-sectional study. *Clin Epidemiol Global Health*. 2021;12:100851. doi:10.1016/j.cegh.2021.100851
53. Bosch R, Pagerols M, Rivas C, et al. Neurodevelopmental disorders among Spanish school-age children: prevalence and sociodemographic correlates. *Psychological Med*. 2022;52:3062–3072. doi:10.1017/S0033291720005115
54. Sung TW, Tsai PW, Gaber T, Lee CY. Artificial Intelligence of Things (AIoT) technologies and applications. *Wireless Communications and Mobile Computing*. 2021;2021:1–2. doi:10.1155/2021/9781271
55. Liu Y, Penttilä N, Ihalainen T, Lintula J, Convey R, Räsänen O. Language-independent approach for automatic computation of vowel articulation features in dysarthric speech assessment. *IEEE/ACM Trans Audio Speech Lang Process*. 2021;29:2228–2243. doi:10.1109/TASLP.2021.3090973
56. Dhouib A, Othman A, El Ghoul O, et al. "Arabic automatic speech recognition: a systematic literature review. *Appl Sci*. 2022;12:8898. doi:10.3390/app12178898
57. Mehrish A, Majumder N, Bharadwaj R, Mihalcea R, Poria S. A review of deep learning techniques for speech processing. *Information Fusion*. 2023;2023:101869.
58. Novotný M, Rusz J, Mejla Cˇ, Ru° R, Žicˇka E. Automatic evaluation of articulatory disorders in parkinson's disease. *ACM Trans Audio Speech and Lang Proc*. 2014;22:1366–1378. doi:10.1109/TASLP.2014.2329734
59. Davis S, Mermelstein P. Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences. *IEEE Trans Acoust*. 1980;28:357–366. doi:10.1109/TASSP.1980.1163420
60. Abeysinghe A, Fard M, Jazar R, Zambetta F, Davy J. Mel frequency cepstral coefficient temporal feature integration for classifying squeak and rattle noise. *J Acoust Soc Am*. 2021;150:193–201. doi:10.1121/10.0005201
61. Corcoran P, Hensman A, Kirkpatrick B Glottal Flow Analysis in Parkinsonian Speech. In Proceedings of the BIOSIGNALS, 2019, pp. 116–123.
62. Cmejla R, Rusz J, Bergl P, Vokral J. Bayesian changepoint detection for the automatic assessment of fluency and articulatory disorders. *Speech Commun*. 2013;55:178–189. doi:10.1016/j.specom.2012.08.003
63. Kodrasi I, Bourlard H. Spectro-temporal sparsity characterization for dysarthric speech detection. *IEEE/ACM Trans Audio Speech Lang Process*. 2020;28:1210–1222. doi:10.1109/TASLP.2020.2985066
64. Gowdy JN, Tufekci Z. Mel-scaled discrete wavelet coefficients for speech recognition. 2000 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings (Cat. No. 00CH37100). IEEE; 2000.
65. Sanz H, Valim C, Vegas E, Oller JM, Reverter F. SVM-RFE: selection and visualization of the most relevant features through non-linear kernels. *BMC Bioinf*. 2018;19:1–18. doi:10.1186/s12859-018-2451-4
66. Cai Y, Huang T, Hu L, Shi X, Xie L, Li Y. Prediction of lysine ubiquitination with mRMR feature selection and analysis. *Amino Acids*. 2012;42:1387–1395. doi:10.1007/s00726-011-0835-0
67. Cilia ND, De Stefano C, Fontanella F, Di Freca AS. A ranking-based feature selection approach for handwritten character recognition. *Pattern Recognit Lett*. 2019;121:77–86. doi:10.1016/j.patrec.2018.04.007
68. Sun Y, Todorovic S, Goodison S. Local-learning-based feature selection for high-dimensional data analysis. *IEEE Transact Patter Analy Mach Intellig*. 2009;32:1610–1626.
69. Tulshiram R. Regression shrinkage and selection via the lasso. *J Royal Statist Soc Series B*. 1996;58:267–288. doi:10.1111/j.2517-6161.1996.tb02080.x
70. Cantürk I, Karabiber F. A machine learning system for the diagnosis of Parkinson's disease from speech signals and its application to multiple speech signal types. *Arab J Sci Eng*. 2016;41:5049–5059. doi:10.1007/s13369-016-2206-3
71. Liu S, Geng M, Hu S, et al. Recent progress in the CUHK dysarthric speech recognition system. *IEEE/ACM Trans Audio Speech Lang Process*. 2021;29:2267–2281. doi:10.1109/TASLP.2021.3091805
72. Azadi H, Akbarzadeh-T MR, Kobravi HR, Shoeibi A. Robust voice feature selection using interval type-2 Fuzzy AHP for automated diagnosis of parkinson's disease. *IEEE/ACM Trans Audio Speech Lang Process*. 2021;29:2792–2802. doi:10.1109/TASLP.2021.3097215
73. Hegde S, Shetty S, Rai S, Dodderi T. A survey on machine learning approaches for automatic detection of voice disorders. *J Voice*. 2019;33:947–e11. doi:10.1016/j.jvoice.2018.07.014

74. Kaur J, Singh A, Kadyan V. Automatic speech recognition system for tonal languages: state-of-The-art survey. *Arch. Comput. Methods Eng.* 2021;28:1039–1068. doi:10.1007/s11831-020-09414-4
75. ACM; 2024. Available from: https://dl.acm.org/. Accessed May 17, 2024.
76. ScienceDirect; 2024. Available from: https://www.sciencedirect.com/. Accessed May 17, 2024.
77. Springer; 2024. Available from: https://link.springer.com. Accessed May 17, 2024.
78. Rudzicz F, Namasivayam AK, Wolff T. The TORGO database of acoustic and articulatory speech from speakers with Dysarthria. *Langu Resourc Evalu*. 2012;46:523–541. doi:10.1007/s10579-011-9145-0
79. Kim H, Hasegawa-Johnson M, Perlman A, et al. Dysarthric speech database for universal access research. In:Proceedings of the Ninth Annual Conference of the International Speech Communication Association; 2008.
80. Laaridh I, Fredouille C, Meunier C. Automatic detection of phone-based anomalies in dysarthric speech. *ACM Transact Accessib Comput*. 2015;6:1–24. doi:10.1145/2739050
81. Franciscatto MH, Del Fabro MD, Lima JCD, et al. Towards a speech therapy support system based on phonological processes early detection. *Comput Speech Lang*. 2021;65:101130. doi:10.1016/j.csl.2020.101130
82. Jong NS, Phukpattaranont P. A speech recognition system based on electromyography for the rehabilitation of dysarthric patients: a Thai syllable study. *Biocybernetics Biomed Eng*. 2019;39:234–245. doi:10.1016/j.bbe.2018.11.010
83. Available from: https://aphasia.talkbank.org/. Accessed May 17, 2024.
84. Yue Z, Loweimi E, Christensen H, Barker J, Cvetkovic Z. Acoustic modelling from raw source and filter components for dysarthric speech recognition. *IEEE/ACM Trans Audio Speech Lang Process*. 2022;30:2968–2980. doi:10.1109/TASLP.2022.3205766
85. Pellegrini T, Fontan L, Mauclair J, et al. Automatic assessment of speech capability loss in disordered speech. *ACM Transact Accessib Comput*. 2015;6:1–14. doi:10.1145/2739051
86. Nagarajan T, Vijayalakshmi P. Dysarthric speech corpus in Tamil for rehabilitation research. In: Proceedings of the 2016 IEEE Region 10 Conference (TENCON). IEEE; 2016:2610–2613.
87. Shahamiri SR, Salim SSB. A multi-views multi-learners approach towards dysarthric speech recognition using multi-nets artificial neural networks. *IEEE Trans Neural Syst Rehabil Eng*. 2014;22:1053–1063. doi:10.1109/TNSRE.2014.2309336
88. Celin TM, Nagarajan T, Vijayalakshmi P. Data augmentation using virtual microphone array synthesis and multi-resolution feature extraction for isolated word dysarthric speech recognition. *IEEE Journal of Selected Topics in Signal Processing*. 2020;14:346–354. doi:10.1109/JSTSP.2020.2972161
89. Mohammed SY, Sid-ahmed S, Brahim-Fares Z, Asma B. Improving dysarthric speech recognition using empirical mode decomposition and convolutional neural network. *EURASIP J Audio*. 2020;1. doi:10.1186/s13636-019-0169-5
90. Narendra N, Schuller B, Alku P. The detection of Parkinson's disease from speech using voice source information. *IEEE/ACM Trans Audio Speech Lang Process*. 2021;29:1925–1936. doi:10.1109/TASLP.2021.3078364
91. The TORGO Database: Acoustic and articulatory speech from speakers with dysarthria. Toronto; 2012. Available from: https://www.cs.toronto.edu/~complingweb/data/TORGO/torgo.html. Accessed May 17, 2024.
92. Christensen H, Rudzicz F, Portet F, Alexandersson J. Perspectives on speech and language interaction for daily assistive technology: introduction to part 1 of the special issue; 2015.
93. Shah M, Tu M, Berisha V, Chakrabarti C, Spanias A. Articulation constrained learning with application to speech emotion recognition. *EURASIP J Audio*. 2019;2019:1–17. doi:10.1186/s13636-019-0157-9
94. Takashima Y, Takashima R, Takiguchi T, Ariki Y. Knowledge transferability between the speech data of persons with Dysarthria speaking different languages for dysarthric speech recognition. *IEEE Access*. 2019;7:164320–164326. doi:10.1109/ACCESS.2019.2951856
95. Woisard V, Astésano C, Balaguer M, et al. C2SI corpus: a database of speech disorder productions to assess intelligibility and quality of life in head and neck cancers. *Langu Resourc Evalu*. 2021;55:173–190. doi:10.1007/s10579-020-09496-3
96. UA-Speech; 2024. Available from: http://www.isle.illinois.edu/sst/data/UASpeech/. Accessed May 17, 2024.
97. Fritsch J, Magimai-Doss M. Utterance verification-based dysarthric speech intelligibility assessment using phonetic posterior features. *IEEE Signal Process Lett*. 2021;28:224–228. doi:10.1109/LSP.2021.3050362
98. Shahamiri SR. Speech vision: an end-to-end deep learning-based dysarthric automatic speech recognition system. *IEEE Trans Neural Syst Rehabil Eng*. 2021;29:852–861. doi:10.1109/TNSRE.2021.3076778
99. Lamel LF, Gauvain JL, Eskénazi M, et al. Bref, a large vocabulary spoken corpus for French. *Training*. 1991;22:50.
100. Little M. Parkinsons. *UCI Mach Learn Reposit*. 2008. doi:10.24432/C591C07774
101. Menendez-Pidal X, Polikoff JB, Peters SM, Leonzio JE, Bunnell HT. The Nemours database of dysarthric speech. In Proceedings of the Proceeding of Fourth International Conference on Spoken Language Processing. ICSLP'96. IEEE; 1996:1962–1965.
102. Geng M, Xie X, Ye Z, et al. Speaker adaptation using spectro-temporal deep features for dysarthric and elderly speech recognition. *IEEE/ACM Trans Audio Speech Lang Process*. 2022;30:2597–2611. doi:10.1109/TASLP.2022.3195113
103. Fougeron C, Crevier-Buchman L, Fredouille C, et al. Developing an acoustic-phonetic characterization of dysarthric speech in French. In Proceedings of the 7th International Conference on Language Resources, Technologies and Evaluation (LREC). Nicoletta Calzolari (Conference Chair) and Khalid Choukri and Bente Maegaard, 2010, Vol. 1, pp. 2831–2838.
104. Mauclair J, Koenig L, Robert M, Gatignol P Burst-based features for the classification of pathological voices. In Proceedings of the INTERSPEECH, 2013, pp. 2167–2171.
105. Parnandi A, Karappa V, Lan T, et al. Development of a remote therapy tool for childhood apraxia of speech. *ACM Transact Accessib Comput*. 2015;7:1–23. doi:10.1145/2776895
106. BREF-120 - A large corpus of French read speech. Elra; 2005. Available from: https://catalogue.elra.info/en-us/repository/browse/ELRA-S0067. Accessed May 17, 2024.
107. Vacher M, Caffiau S, Portet F, et al. Evaluation of a context-aware voice interface for ambient assisted living: qualitative user study vs. quantitative system evaluation. *ACM Transact Accessib Comput*. 2015;7:1–36. doi:10.1145/2738047
108. Pradhan A, Mehta K, Findlater L. "Accessibility Came by Accident" Use of Voice-Controlled Intelligent Personal Assistants by People with Disabilities. In Proceedings of the Proceedings of the 2018 CHI Conference on human factors in computing systems; 2018:1–13. doi:10.1145/3173574.3174033.

109. Kominek J, Black AW. The CMU Arctic speech databases. In: Proceedings of the Fifth ISCA workshop on speech synthesis; 2004.
110. Dudy S, Bedrick S, Asgari M, Kain A. Automatic analysis of pronunciations for children with speech sound disorders. *Comput Speech Lang*. 2018;50:62–84. doi:10.1016/j.csl.2017
111. Gupta S, Patil AT, Purohit M, et al. Residual neural network precisely quantifies dysarthria severity-level based on short-duration speech segments. *Neural Networks*. 2021;139:105–117. doi:10.1016/j.neunet.2021.02.008
112. Chandrakala S, Rajeswari N. Representation learning based speech assistive system for persons with Dysarthria. *IEEE Trans Neural Syst Rehabil Eng*. 2016;25:1510–1517. doi:10.1109/TNSRE.2016.2638830
113. Kurematsu A, Takeda K, Sagisaka Y, Katagiri S, Kuwabara H, Shikano K. ATR Japanese speech database as a tool of speech recognition and synthesis. *Speech Commun*. 1990;9:357–363. doi:10.1016/0167-6393(90)90011-W
114. Sedgwick P. Pearson's correlation coefficient. *BMJ*. 2012;2012:345.
115. Narendra NP, Alku P. Automatic intelligibility assessment of dysarthric speech using glottal parameters. *Speech Commun*. 2020;123:1–9. doi:10.1016/j.specom.2020.06.003
116. The SSNCE Database of Tamil Dysarthric Speech. P. Vijayalakshmi, T. A. Mariya Celin, T. Nagarajan; 2021. Available from: https://catalog.ldc.upenn.edu/LDC2021S04. Accessed May 17, 2024.
117. Asaei A, Cernak M, Bourlard H. Perceptual information loss due to impaired speech production. *IEEE/ACM Transacti Aud Spe Langu Process*. 2017;25:2433–2443. doi:10.1109/TASLP.2017.2738445
118. MoSpeeDi. Universite Fe Geneve; 2021. Available from: https://www.unige.ch/fapse/mospeedi/sous-projets. Accessed May 17, 2024.
119. Orozco–Arroyave JR, Arias-Londoño JD, Vargas-Bonilla JF, Gonzalez-Rátiva MC, Nöth E New Spanish speech corpus database for the analysis of people suffering from Parkinson's disease. In: Proceedings of the LREC; 2014:342–347.
120. Conn P. Distribution of Language Measures among Individuals with and without Non-Fluent Aphasia. In: Proceedings of the Proceedings of the 10th International Conference on PErvasive Technologies Related to Assistive Environments, 2017: 252–253. doi:10.1145/3056540.3076214.
121. TIMIT Acoustic-Phonetic Continuous Speech Corpus. John S. Garofolo, Lori F. Lamel, William M. Fisher, et al; 1993. Available from: https://catalog.ldc.upenn.edu/LDC93S1. Accessed May 17, 2024.
122. CMU_ARCTIC speech synthesis databases. Language Technologies Institute at Carnegie Mellon University; 2003. Available from: http://www.festvox.org/cmu_arctic/. Accessed May 17, 2024.
123. Hair A, Ballard KJ, Markoulli C, et al. A longitudinal evaluation of tablet-based child speech therapy with Apraxia World. *ACM Transact Accessib Comput*. 2021;14:1–26. doi:10.1145/3433607
124. Kim MJ, Kim Y, Kim H. Automatic intelligibility assessment of dysarthric speech using phonologically-structured sparse linear model. *IEEE/ACM Trans Audio Speech Lang Process*. 2015;23:694–704. doi:10.1109/TASLP.2015.2403619
125. Middag C, Martens JP, Van Nuffelen G, De Bodt M. Automated intelligibility assessment of pathological speech using phonological features. *EURASIP J Adv Signal Process*. 2009;2009:1–9. doi:10.1155/2009/629030
126. Oxford Parkinson's Disease Detection Dataset. Max A. Little, P. McSharry, S. Roberts, at al; 2007. Available from: https://archive.ics.uci.edu/dataset/174/parkinsons. Accessed May 17, 2024.
127. Dhanalakshmi M, Mariya Celin T, Nagarajan T, Vijayalakshmi P. Speech-input speech-output communication for dysarthric speakers using HMM-based speech recognition and adaptive synthesis system. *Circuit Syst Signal Proc*. 2018;37:674–703. doi:10.1007/s00034-017-0567-9
128. Bhat C, Strik H. Automatic assessment of sentence-level dysarthria intelligibility using BLSTM. *IEEE Journal of Selected Topics in Signal Processing*. 2020;14:322–330. doi:10.1109/JSTSP.2020.2967652
129. Garofolo JS. *TIMIT Acoustic Phonetic Continuous Speech Corpus*. Linguistic Data Consortium; 1993.
130. Marfoq O, Neglia G, Kameni L, Vidal R Federated Learning for Data Streams. In Proceedings of the Proceedings of The 26th International Conference on Artificial Intelligence and Statistics. Proceedings of Machine Learning Research; 2023:8889–8924.
131. Kodrasi I. Temporal envelope and fine structure cues for dysarthric speech detection using CNNs. *IEEE Signal Process Lett*. 2021;28:1853–1857. doi:10.1109/LSP.2021.3108509
132. Lee S, Yildirim S, Kazemzadeh A, Narayanan S. An articulatory study of emotional speech production. In: Proceedings of the Ninth European Conference on Speech Communication and Technology; 2005.
133. Liu Y, Reddy MK, Penttilä N, Ihalainen T, Alku P, Räsänen O. Automatic Assessment of Parkinson's Disease Using Speech Representations of Phonation and Articulation. *IEEE/ACM Trans Audio Speech Lang Process*. 2023;31:242–255. doi:10.1109/TASLP.2022.3212829
134. Ballati F, Corno F, De Russis L. Assessing virtual assistant capabilities with Italian dysarthric speech. In: Proceedings of the Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility; 2018:93–101. doi:10.1145/3234695.3236354
135. Martínez D, Lleida E, Green P, Christensen H, Ortega A, Miguel A. Intelligibility assessment and speech recognizer word accuracy rate prediction for dysarthric speakers in a factor analysis subspace. *ACM Transact Accessib Comput*. 2015;6:1–21. doi:10.1145/2746405
136. Meunier C, Fougeron C, Fredouille C, et al. The TYPALOC Corpus: a collection of various dysarthric speech recordings in read and spontaneous styles. In: Proceedings of the Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16), Portorož, Slovenia; 2016:4658–4665.
137. Busso C, Bulut M, Lee CC, et al. IEMOCAP: interactive emotional dyadic motion capture database. *Language Res Evalu*. 2008;42:335–359. doi:10.1007/s10579-008-9076-6
138. Le D, Licata K, Persad C, Provost EM. Automatic assessment of speech intelligibility for individuals with aphasia. *IEEE/ACM Trans Audio Speech Lang Process*. 2016;24:2187–2199. doi:10.1109/TASLP.2016.2598428
139. Sajal MSR, Ehsan MT, Vaidyanathan R, Wang S, Aziz T, Mamun KAA. Telemonitoring Parkinson's disease using machine learning by combining tremor and voice analysis. *Brain Informat*. 2020;7:1–11. doi:10.1186/s40708-020-00113-1
140. Ramou N, Guerti M. Automatic detection of articulations disorders from children's speech preliminary study. *J Communicat Technol Elect*. 2014;59(11):1274–1279. doi:10.1134/S1064226914110187
141. MoSpeeDi. Universite Fe Geneve; 2021. Available from: https://www.unige.ch/fapse/mospeedi/sous-projets. Accessed May 17, 2024.
142. Vikram C, Adiga N, Prasanna SM. Detection of nasalized voiced stops in cleft palate speech using epoch-synchronous features. *IEEE/ACM Trans Audio Speech Lang Process*. 2019;27:1189–1200. doi:10.1109/TASLP.2019.2913089