TOPICAL REVIEW

The Detection of Dysarthria Severity Levels Using AI Models: A Review

AFNAN AL-ALI[1], **SOMAYA AL-MAADEED**[1], (Senior Member, IEEE), **MOUTAZ SALEH**[1],
RANI CHINNAPPA NAIDU[2], **ZACHARIAH C. ALEX**[2], **PRAKASH RAMACHANDRAN**[2],
RAJEEV KHOODEERAM[3], AND **RAJESH KUMAR M**[2], (Member, IEEE)

[1]Computer Science and Engineering Department, Qatar University, Doha, Qatar
[2]Vellore Institute of Technology, Vellore 632 014, India
[3]Faculty of Sustainable Development and Engineering, Université des Mascareignes, Beau Bassin-Rose Hill 71203, Mauritius

Corresponding author: Afnan Al-Ali (aa1805360@qu.edu.qa)

ABSTRACT Dysarthria, a speech disorder stemming from neurological conditions, affects communication and life quality. Precise classification and severity assessment are pivotal for therapy but are often subjective in traditional speech-language pathologist evaluations. Machine learning models offer objective assessment potential, enhancing diagnostic precision. This systematic review aims to comprehensively analyze current methodologies for classifying dysarthria based on severity levels, highlighting effective features for automatic classification and optimal AI techniques. We systematically reviewed the literature on the automatic classification of dysarthria severity levels. Sources of information will include electronic databases and grey literature. Selection criteria will be established based on relevance to the research questions. The findings of this systematic review will contribute to the current understanding of dysarthria classification, inform future research, and support the development of improved diagnostic tools. The implications of these findings could be significant in advancing patient care and improving therapeutic outcomes for individuals affected by dysarthria.

INDEX TERMS Dysarthria, classification, severity levels, artificial intelligence (AI)-based models, intelligibility.

I. INTRODUCTION

Speech is a distinctive, intricate, dynamic motor activity that enables us to articulate our thoughts and emotions and interact with and regulate our surroundings. Furthermore, it constitutes one of the five heritable verbal traits, encompassing speech, language, reading, writing, and spelling. Speech involves integrating neurocognitive processes for organizing thoughts into language, motor speech planning for executing verbal messages, and neuromuscular execution for coordinating speech muscles. Together, these processes constitute motor speech activities.

When neurologic impairments affect these motor speech activities, a speech disorder will result, which can also be known as Motor speech disorder (MSD) [1], [2].

The associate editor coordinating the review of this manuscript and approving it for publication was Rongbo Zhu.

There are two main types of MSD: dysarthrias and apraxia of speech. Dysarthria is a group of neurological speech disorders involving abnormalities in speech production and movement. These disorders manifest as changes in strength, speed, range, steadiness, tone, or accuracy of movements required for breathing, phonation, resonance, articulation, or prosody. Sensorimotor abnormalities, such as weakness, spasticity, incoordination, involuntary movements, or variations in muscle tone, underlie dysarthria. The condition is specifically neurologic and can be categorized into distinct types based on perceptual characteristics and underlying neuropathophysiology. It is essential to define dysarthria accurately to differentiate it from other speech and language disorders and to ensure its meaningful application in research and clinical settings [2]. Verbal apraxia, also known as apraxia of speech, is a contentious condition that some view as a deficit in the motor planning of speech. This disorder

is marked by "higher order" errors, including metathesis and segment addition, alongside errors that suggest a lack of coordination in articulation. These characteristics suggest a relatively significant level of damage to the neural system [3].

Traditionally, dysarthria is classified into several subtypes, including spastic, flaccid, hypokinetic, hyperkinetic, ataxic, and mixed. More recently, additional subtypes such as unilateral upper motor neuron dysarthria and undetermined dysarthria have been recognized [4]. Dysarthria can occur at any stage of life. Common causes encompass stroke, severe head injury, brain tumors, Parkinson's disease, multiple sclerosis, motor neuron disease, cerebral palsy, Down's syndrome, and certain medications, including those used to treat epilepsy, which may induce a side effect [2]. The severity of dysarthria is determined by the degree of involvement in the affected body regions caused by the underlying condition. Assessment of dysarthria means mainly grading its severity. This is typically performed by speech-language pathologists (SLPs) using some descriptive terms, but it can be a time-consuming and labor-intensive process with variations between different SLPs. Thus, it is essential to have objective methods of evaluating the level of intelligibility in dysarthria cases [5]. For this purpose, machine learning-based models were developed for automatic assessment of dysarthrias' levels of severity to achieve enhanced diagnostic accuracy, consistency, and reliability, all while maintaining cost-effectiveness and expediency [6]. In both paths, features are extracted from the candidate samples to help the underlying system for classification, where there are specific sets of features for each type of dysarthria. To our knowledge, this is the first comprehensive review that analyzes the works of classifying dysarthria cases based on the levels of severity. In the literature, several gaps warrant further research in conducting a systematic review on classifying dysarthria based on severity levels. These gaps include the absence of a review focused explicitly on severity-based classification, limited research on severity levels within specific populations [7], [8] [9], [10] (population-specific considerations), a lack of standardized measures for assessing severity [8], [11] [12], [13] (measurement challenges), the integration of technology in severity classification [14] (technological advancements), a need for a comparative analysis of existing classification systems [15] (evaluation of existing approaches), specific causes of dysarthria [16] (etiological factors), general treatment helping clinicians for stable dysarthria [17] (treatment approaches) and automated intelligibility assessment [18] (technology-driven assessment methods). Addressing these gaps through a systematic review would offer valuable insights for clinicians, researchers, and stakeholders involved in dysarthria assessment and treatment.

In this systematic review, we aim to answer the below research questions:

- What are the optimal set and type of features that are consistently effective across various severity levels

of dysarthria, enabling automated classification of patients?
- Which artificial intelligence techniques are most suitable for accurately classifying dysarthria patients, considering factors such as training time efficiency and high accuracy?

The rest of the paper is organized as follows: Section 2 will show the strategy followed in our research, Section 3 presents how dysarthric cases are classified based on clinical techniques, Section 4 will show the same for Section 3, but based on machine learning techniques, Section 5 discusses the results shown in the two previous sections and analyses them. Section 6 refers to some common limitations of the related work. Section 7 suggests a few points to solve the gap in this research area, and finally, the Conclusion in Section 8 will summarize our work and highlight our findings in this review.

II. SEARCH STRATEGY

A comprehensive exploration was conducted on various electronic databases, including ACM, EMBASE, SpringerLink, PubMed, Scopus, IEEE, MDPI, Elsevier, and some other conferences popular in the area, based on the search keywords: "classification" AND "Dysarthria" AND "severity levels" or "Assessment" AND "Dysarthria" AND "Intelligibility." Figure 1 shows the search strategy process.

A total of 978 publications were found. After deleting the duplicates, 733 publications passed through a screening process where only the articles that contained the keywords in their titles, abstracts, and keywords were selected, and the excluded publications were based on the following criteria:(1) reviews and surveys are not considered, (2) articles do not have the main keywords in their titles and abstracts, (3) general chapters related to motor speech disorders are not considered, and (4) other types of disorders like Dysphagia or Dementia.

The second stage of the study search is knowing the eligibility of each publication to be selected for the analysis later on in this study. The exclusion criteria in this stage are (1) The tools or software for rehabilitation or therapy, (2) automatic recognition systems or the identification of dysarthria, (3) reviews and surveys related to dysarthria or one of its specific types or causes, (4) the impact of specific features of dysarthric cases, and (5) the binary classification of dysarthria.

The final publications are 44 articles mainly related to classifying dysarthria patients based on severity levels and assessing the intelligibility factor as severity levels.

As our primary goal is to classify dysarthric cases based on severity levels, we will refer to both types: clinical or human-based and AI or machine learning-based techniques. We will highlight the methods, extracted features, and the datasets they evaluated the models on. Figure 2 shows the taxonomy of our review in classifying dysarthria based on the severity levels.
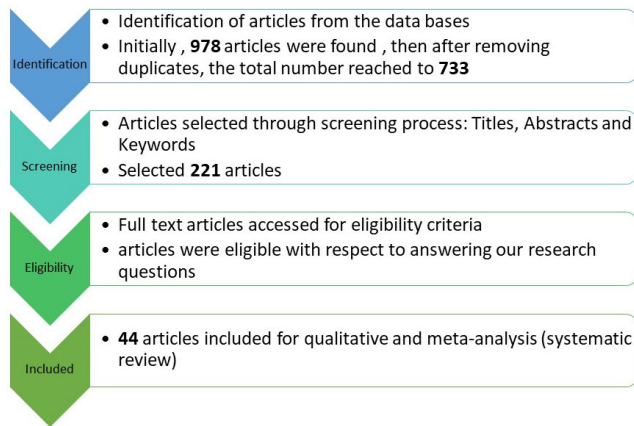
FIGURE 1. Search study process.

III. CLASSIFICATION BASED ON CLINICAL TECHNIQUES

When a dysarthric patient is admitted to the hospital, a specific procedure will be followed based on the clinical setting, available resources, and individual needs. The main points regarding each procedure are explained in the sub-sections below:

A. METHODS

As a clinical procedure, speech-language pathologists (SLPs) follow up on dysarthrias cases. Other rehabilitation professionals, physicians, and nurses may also be involved in such treatment. Each SLPs will search for some descriptive patterns in the patient to confirm his diagnosis and decide the type of treatment based on the severity of the case [19].

The procedure followed by speech-language pathologists (SLPs) to measure the severity of dysarthria in a patient typically involves employing a comprehensive approach to assess dysarthria in patients. This process includes gathering the patient's medical history to provide context, visually observing their speech and oral motor movements, making perceptual assessments of speech characteristics and severity, evaluating speech intelligibility through tests and measurements, assessing the impact of dysarthria on functional communication abilities, and collaborating with patients, family members, and healthcare professionals to gain a comprehensive understanding of the condition. By following this systematic approach, SLPs can develop tailored treatment plans to address the specific challenges faced by each patient [2], [20].

B. FEATURES

Speech-language pathologists (SLPs) use various features to assess the severity of dysarthria in patients. These features include articulation, by assessing the accuracy and precision of speech sound production; phonation, by evaluating voice quality, pitch, loudness, and presence of abnormalities; resonance, by examining the control of the velopharyngeal mechanism for speech clarity; prosody, by analyzing rhythm, stress, intonation, and melodic contour; speech rate,

by observing speed, pacing, and pauses; and intelligibility, by determining the percentage of intelligible speech.

SLPs employ clinical observation, perceptual judgments, and instrumental assessments like acoustic analysis to assess these features. By considering these aspects, SLPs can determine dysarthria severity and develop customized treatment plans to address patients' specific speech challenges [19], [21].

C. EVALUATION

Several standardized rating scales and perceptual judgments are commonly used to assess dysarthria's severity and specific features. These include formal and informal assessments. The formal ones are represented by the most famous Frenchay Dysarthria Assessment (FDA) [22], which evaluates respiration, phonation, articulation, and prosody; other measurements like the Assessment of Intelligibility of Dysarthric Speech (AIDS), which measures overall speech intelligibility [23] and dysarthria profile [24]; the and Voice Physical Disability Index [25]. Informal assessments, such as oral motor examinations [26], are often used with formal assessments. Perceptual assessment [27] is used by speech-language pathologists, relying on their expertise in active listening and analyzing speech. It's important to note that these perceptual judgments are subjective, emphasizing the need for skilled clinicians to accurately assess and interpret the speech characteristics of individuals with dysarthria. Finally, Communication Activities of Daily Living-Second Edition (CADL-2) [28], Although not specific to dysarthria, this assessment measures functional communication abilities in daily life situations. It evaluates the impact of dysarthria on communication in various contexts [29].

The inclusion of the section on clinical assessment in our review serves two essential purposes. Firstly, it provides a comprehensive understanding of the traditional approaches and methods used in dysarthria classification based on severity levels. By examining the techniques clinicians employ and the features they consider, we gain insights into the established practices and evaluation tools that have been relied upon for decades. This knowledge is crucial for contextualizing and appreciating the advancements brought by AI-based techniques.

Secondly, incorporating the clinical assessment perspective allows for a comparative analysis between human-based approaches and AI-based methods. By juxtaposing the strengths and limitations of clinical judgment with the capabilities of AI models, we can critically evaluate the potential of AI to augment and enhance dysarthria classification. This comparative analysis enables us to identify the unique contributions of AI techniques, such as increased objectivity, scalability, and potential for automated assessment.

IV. CLASSIFICATION BASED ON AI TECHNIQUES

Despite the effectiveness of SLPs' efforts, this process is time-consuming, mainly subjective, and suffers from
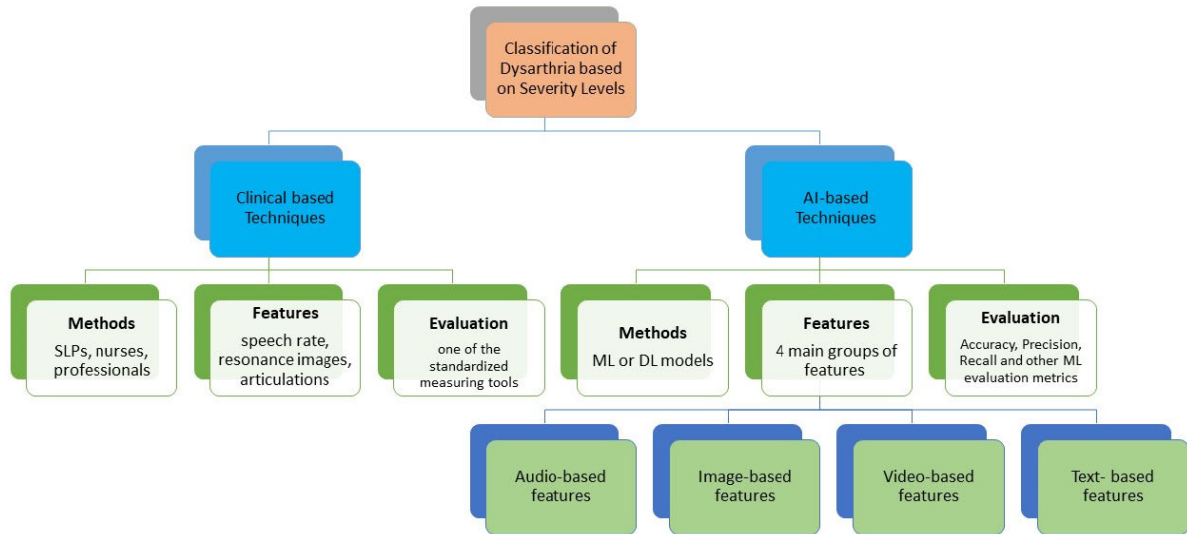
FIGURE 2. Taxonomy of the study.

differences among individual opinions. This gap inspired the researchers to seek more accurate models, away from the subjective perspective, using AI-based models, as explained below.

A. METHODS

Machine learning and deep learning-based models have shown promise in assessing the severity of dysarthria, offering alternative approaches to traditional assessments conducted by SLPs. These models leverage the power of artificial intelligence to analyze speech patterns and extract meaningful information for severity evaluation [30].

Several common machine learning approaches, such as support vector machines (SVM), Random Forests (RF), or Artificial Neural Networks (ANN), have been utilized to develop predictive models [31]. These models are trained on a dataset of acoustic features or other forms of acoustic features extracted from speech samples of individuals with dysarthria and corresponding severity ratings provided by SLPs. The models learn patterns and relationships between these features and the severity ratings, enabling them to predict/classify the severity of dysarthria in new, unseen cases.

Deep learning, a subset of machine learning, has gained attention in dysarthria severity assessment. Deep Neural Networks (DNN), specifically Long-Short Term Memory (LSTM) [32] and convolutional neural networks (CNNs) [33], have been employed to analyze speech signals and capture intricate temporal and spectral patterns. These models can process raw speech data directly or extract features automatically through multiple layers of computation. They can learn complex representations and make predictions based on the learned patterns, allowing for more accurate severity assessment.

Research studies have demonstrated the potential of these machine learning and deep learning models in objectively quantifying the severity of dysarthria. These models offer advantages such as increased efficiency, consistency, and potential for remote or self-administered assessments. However, it is essential to note that these models are still evolving. Their performance may vary depending on factors such as the quality and diversity of training data, feature selection, and the complexity of the dysarthria cases [34], [35].

B. FEATURES

Machine learning and deep learning models for assessing the severity of dysarthria utilize a range of features extracted from speech signals. These features capture different aspects of speech production and can provide valuable information for severity assessment. We can categorize these features into four main groups: audio-based, image-based, video-based, and text-based. Let's break it down:

1) AUDIO-BASED FEATURES

Audio-based features for dysarthria classification based on severity can be broadly categorized into acoustic, prosodic, and spectral features [36], [37], [38], [39].

Acoustic features include the fundamental frequency (F0), which corresponds to the pitch of the speech, as well as measures of variability such as jitter and shimmer, reflecting the stability of vocal fold vibration. The Harmonics-to-Noise Ratio (HNR) is another acoustic feature that indicates the ratio of energy in the harmonics of the speech signal to the energy in the noise. Lower HNR values may indicate poor voice quality.

Prosodic features play a role in dysarthria classification. Speech rate, which refers to the number of syllables per unit of time, can be affected by the severity of dysarthria. Pause duration, or the length of pauses between speech segments,

can provide insights into difficulties in speech production or planning. Additionally, changes in stress patterns, including variations in intensity, duration, and pitch of syllables, can indicate the presence of dysarthria.

Spectral features offer valuable information for dysarthria classification. Mel-Frequency Cepstral Coefficients (MFCCs) capture the speech signal's short-term power spectrum, providing insights into the speaker's vocal tract shape and function. Linear Predictive Coding (LPC) coefficients represent the spectral envelope of the speech signal, offering information about the vocal tract shape and articulatory movements. Formant frequencies, such as F1, F2, F3, etc., represent the resonant frequencies of the vocal tract during speech production, revealing details about the shape and position of the articulators (tongue, lips, etc.).

The audio-based features mentioned above are not the only types in each group. They represent typical features for dysarthria classification, but additional features can be within each category [26], [40]. Table 11 summarizes some studies that used this type of feature for classifying dysarthria based on severity levels or estimating intelligibility for severity classification of dysarthria.

2) IMAGE-BASED FEATURES

Image-based features in the context of dysarthria analysis involve spectrograms, which are visual representations of the frequency content of an audio signal over time.

Spectrograms provide a 2D visual representation of the frequency content of an audio signal over time. They are commonly used in speech and audio analysis, including dysarthria classification. By analyzing spectrograms, various features can be extracted to characterize dysarthric speech. These features include spectral envelope, spectral patterns, spectral variations, and spectral entropy. Spectrograms offer valuable insights into the frequency components and dynamics of speech signals [41].

Mel spectrograms, also known as Mel-frequency spectrograms or Mel-scaled spectrograms, are a type of spectrogram that uses the Mel scale to warp the frequency axis perceptually. The mel scale better aligns with the human auditory system's frequency perception. To generate a mel spectrogram, the audio signal is divided into frames, and the power spectrum is calculated using techniques like the Fast Fourier Transform (FFT). The resulting power spectrum is then transformed to the mel scale using triangular filter banks. Mel spectrograms are particularly useful in speech analysis tasks, including dysarthria classification. They capture essential spectral information, such as formant frequencies and spectral patterns, while aligning with human perception. Mel spectrograms can be used as image-based features for dysarthria analysis or further processed to extract specific characteristics [42]. Additionally, log Mel spectrograms, obtained by taking the logarithm of the values in the Mel spectrogram, can compress the dynamic range and enhance the visualization and analysis of the mel-frequency content of the audio signal [43].

Table 2 summarizes some of the studies in the literature within the collected papers for this review that have utilized image-based features to classify dysarthria based on severity levels.

3) VIDEO-BASED FEATURES

Video-based features in dysarthria analysis involve analyzing the movement of the lips during speech. Lip Movement Analysis aims to extract visual features that provide information about lip shape, lip dynamics, and lip synchronization. By analyzing the visual cues from lip movements, researchers can gain insights into the articulatory aspects of speech production [44]. However, it is essential to note that classifying the severity of dysarthria solely based on video-based features can be challenging. Dysarthria affects speech mechanisms, including articulation, phonation, resonance, and respiration, which may not be directly observable in video footage. While lip movements can provide valuable information [45], they may not capture the complete picture of dysarthria symptoms.

It is worth mentioning that research and advancements in computer vision techniques, such as facial landmark tracking and optical flow analysis, continue to improve the estimation of video-based features for dysarthria. These techniques enable more precise extraction and interpretation of lip movements and other facial cues, enhancing the potential for video-based research in dysarthria assessment.

4) TEXT-RELATED FEATURES

Text-related features play a crucial role in classifying dysarthria, providing valuable insights into the phonetic characteristics and speech production patterns of individuals with this condition. These features are derived from the analysis of voice signals and focus on the phonetic content of the speech rather than the textual information. Phoneme-level intelligibility serves as a prominent text-related feature, assessing the accuracy of phoneme production by transcribing and comparing the phonetic content of the speech. Additionally, phonetic distance and articulation errors contribute to the classification process, quantifying the dissimilarity between intended and produced phonemes and identifying specific articulation difficulties. By incorporating these text-related features, dysarthria classification models can capture and analyze the phonetic intricacies of speech, enabling improved understanding and identification of different dysarthria subtypes and severity levels. Such as in [46], [47], and lexical frequency, phonological neighborhood, word class, and lexical familiarity in [48].

C. EVALUATION

The evaluation techniques used in machine learning and deep learning models play a crucial role in assessing their performance and effectiveness. These techniques provide insights into the model's predictive capabilities and help practitioners make informed decisions [49], [50], [51], [52]. These evaluation metrics include *Accuracy*, which measures

the overall correctness of the model's predictions. *Precision and Recall* are commonly used in binary classification tasks, where precision measures the proportion of correctly predicted positive instances and recall measures the proportion of correctly predicted positive instances among all actual positives. The *F1 Score* combines precision and recall into a harmonic mean. *Mean Squared Error (MSE)* is used for regression tasks and measures the average squared difference between predicted and actual values. *Area Under the Curve (AUC)* evaluates binary classifiers by calculating the area under the ROC curve. *Cross-Validation* helps assess model performance across multiple iterations and reduces the risk of overfitting. The *Confusion Matrix* provides a detailed evaluation of the model's performance by showing true positives, true negatives, false positives, and false negatives. These evaluation techniques assist practitioners in assessing model performance, identifying areas for improvement, and comparing different models for a given task.

V. DISCUSSION

Let's analyze these categories of features to compare the effectiveness of different features for classifying dysarthria based on severity levels.

Table 1 summarizes studies that have used audio-based features for dysarthria classification based on severity levels. The studies utilize various audio-based features such as fundamental frequency (F0), jitter, shimmer, harmonics-to-noise ratio (HNR), speech rate, pause duration, stress patterns, Mel-Frequency Cepstral Coefficients (MFCCs), Linear Predictive Coding (LPC) coefficients, and formant frequencies (F1, F2, F3, etc.). Different classification techniques have been used, including Random Forest (RF), Support Vector Machine(SVM), Artificial Neural Network (ANN), Classification and Regression Tree (CART), Naive Bayes (NB), DNN, CNN, LSTM, Residual Networks (ResNet), multi-layer perception(MLP), Extreme Gradient Boosting (XGBoost), Gaussian Mixture Model(GMM), Probabilistic linear discriminant analysis (PLDA), Hidden Markov Model (HMM), and k-nearest neighbor(KNN).

It can be seen that most of the researchers' works were evaluated based on standard publicly available datasets such as TORGO [79], UA Speech [80], Qolt [81], and Numours datasets [82], as well as some other locally collected datasets or other less common foreign languages datasets.

Figure 3 shows the distribution of the included papers chosen for this study based on Categorized Features and Dataset Groups. The figure reveals interesting patterns in the distribution of papers across different categorized features and dataset groups in dysarthria classification. Among the featured categories, audio-based techniques demonstrate the highest representation across all dataset groups, including TORGO, UA Speech, Nemours, Qolt, and others. This indicates the prevalent use of audio features in dysarthria classification research across diverse datasets. The prominence of audio-based techniques suggests that researchers prioritize

capturing the acoustic characteristics of dysarthric speech for accurate classification.

In contrast, image-based and text-based features have a limited presence, with only a few studies exploring their potential within the UA Speech and other dataset groups. This highlights a potential research gap and suggests the need for further investigation into the utilization of visual information for improved dysarthria classification. Video-based features, on the other hand, are not extensively explored in the selected papers, indicating a less prominent role in dysarthria classification across all dataset groups. We added the mixed category as well, which incorporates multiple feature types and shows a moderate presence in some dataset groups, underscoring the potential benefits of integrating different modalities to enhance classification performance

The majority of audio-based features are used due to several advantages. Firstly, they provide a direct measurement of speech by capturing important properties of speech signals, including pitch, intensity, and spectral characteristics. These features are essential in evaluating dysarthria, primarily affecting a person's ability to produce clear and intelligible speech.

Secondly, acoustic features allow for objective and quantitative analysis through signal-processing techniques. This objectivity and quantifiability enhance the consistency and reliability of dysarthria severity assessments. By relying on concrete acoustic properties of speech, these measures provide a more robust evaluation of the condition.

Another benefit is that acoustic analysis is non-invasive and accessible. It can be conducted using standard methods such as recording speech samples with a microphone. This practicality makes acoustic analysis suitable for various settings, including clinics, research studies, and telemedicine. It eliminates the need for invasive procedures or specialized equipment, increasing the ease of implementation.

Furthermore, acoustic features enable longitudinal monitoring of dysarthria progression and treatment outcomes. By analyzing changes in the acoustic properties of speech over time, clinicians and researchers can assess the effectiveness of interventions and track the impact of dysarthria on an individual's communication abilities. This longitudinal perspective provides valuable insights into the management and prognosis of dysarthria [83], [84], [85].

While using acoustic features in dysarthria assessment offers several advantages, there are some limitations to consider. Acoustic features partially represent dysarthric speech, as dysarthria encompasses various communication aspects beyond acoustic properties, such as articulation and prosody. Additionally, acoustic analysis lacks the broader context of communication, failing to capture factors like facial expressions or situational cues that influence speech intelligibility.

The generalizability of acoustic features is limited, as they are derived from controlled environments and may not account for real-world variations in acoustic conditions. The acoustic analysis also primarily focuses on objective

TABLE 1. Summary of the related works used the audio-based features for classifying dysarthria based on the severity levels.

Method	Feature	Classifier	Dataset
[5]	Breathiness features like HNR	SVM	UA Speech and Kannada Database
[31]	prosody, voice quality, MFCC and their combination	RF, SVM, MLP	TORGO, QoLT
[40]	features extracted by VGGish net	ML, TF	Healthy, ataxia subjects
[53]	Acoustic characteristics, articulatory movements, glottal functioning, and low-dimensional representations	DNN, CNN, gated recurrent units (GRU),and LSTM	UA-Speech and TORGO
[54]	Hand-crafted acoustic features, as well as raw waveforms	Multi task learning (MLT), SVM, MLP, and XGBoost	QoLT
[55]	Spectral domain representation features, Cepstral domain representation features, and Frame-level features	PLDA, and ANN	UA speech, TORGO
[56]	MFCCs, Voice Quality Features, Prosody Features	SVM, ANN	QoLT
[57]	eleven prosodic features selected by LDA	GMM, SVM	Nemours database
[58]	four acoustic features: prosody, spectral, cepstral, and voice quality	SVM, LDA, ANN, CART, NB, RF	Nemours database
[59]	Multi-tapered spectral estimation-based features, Acoustic descriptors for timbre	ANN	UA Speech and TORGO
[60]	Standard Prosodic Features and Rhythm-Based Features	RF, SVM, MLP	QoLT, and TORGO
[61]	Cepstral Coefficients (TECC, LFCC), and spectral features.	CNN, Light CNN, ResNet	UA Speech
[62]	Temporal Features, Spectral Features, Voice Quality Features	Mahalanobis distance	UA Speech
[63]	Language-independent features from diverse speech dimensions, and language-unique features specific to each language	XGBoost	TORGO, QoLT, and SS-NCE
[64]	Prosodic features	GMM and SVM	Nemours database
[65]	prosodic features: (MFCCs), skewness, and formants.	SVM	UA Speech
[66]	10 speech features related to phonetic quality, prosodic quality, and voice quality	SVM	QoLT
[67]	MFCCs	CNN	UA Speech and home service Corpus
[68]	Acoustic/Signal-based Features (SECC, TECC)	CNN, Light CNN, ResNet	UA speech
[69]	10 acoustic features	Discriminant analysis	Dutch speakers and English speakers
[70]	Single Frequency Filter Bank-based Instantaneous Frequency Cepstral Coefficients (SFFB-IFCC)	I-Vector with PLDA Classifier	UA Speech
[71]	MFCCs and their first two derivatives (a total of 39 features).	SVM, DNN, CNN, LSTM	UA Speech, and TORGO
[72]	Glottal Parameters and basic acoustic features	SVM	UA Speech
[73]	Spectral, prosodic, Teager Energy Operator (TEO), and glottal waveform features	SVM	UA Speech
[74]	traditional acoustic features like MFCC and auditory-based cues.	GMM, SVM, and hybrid GMM/SVM	the Nemours database and Torgo
[75]	Rhythm Metrics Various rhythm metrics, such as SRVC (speaking rate of VC segments)	SVM	Nemours corpus
[76]	prosody, articulation, phonation, and glottal	SVM, NB, kNN, and RF	UA Speech and TORGO
[77]	wav2vec features	SVM	UA Speech
[78]	Frame-level features with global statistics	Forward selection method (FSM)and boosting algorithm	Dataset of 39 dysarthria patients

measures, potentially overlooking subjective experiences and perceptions of individuals with dysarthria. Moreover, distinguishing between dysarthria subtypes solely based on acoustic analysis can be challenging due to shared acoustic characteristics [86], [87]. Despite the promising results of the studies in Table 1, it shows some limitations due to the above-mentioned drawbacks of the audio-based features and other reasons related to the type and size of the dataset used
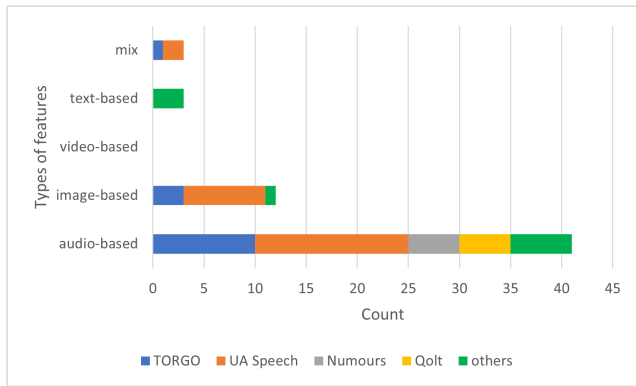
FIGURE 3. The percentage of each feature's category.

for their training and evaluation and the type of the classifiers used.

To overcome these limitations, researchers and clinicians often explore multimodal approaches incorporating other data types, such as images, videos, linguistic features, or perceptual assessments. Integrating multiple modalities can provide a more holistic understanding of dysarthria and improve the accuracy and robustness of the assessment process. As can be seen, some of these attempts are shown in Table 2 for studies that have used image-based features, where most achieved high accuracies around 90% and above. Despite the high performance in most studies in Table 2, there are still some limitations, including data scarcity, challenges in collecting dysarthric speech data, lower accuracy due to limited datasets, and confinement to isolated utterances.

Some researchers have derived benefits from employing text-based features within the feature types discussed earlier. These features, extracted from the patient's voice signal and converted into text, have demonstrated promising outcomes. However, their usage has significant limitations, rendering exclusive reliance on these types impractical. These limitations are represented by the sound quality, which plays a crucial role in the transcription process.

Despite sophisticated speech models employed by Speech-to-Text applications, accuracy can be compromised by the type and specifications of the microphones used. Audio input quality might be degraded if users speak too close to or too far from the microphone, which can subsequently impact the transcription's precision. Another fundamental limitation is the presence of background and environmental noise in the audio input. Non-speech sounds can interfere with the audio, leading to less accurate transcriptions. The complexity of transcription can also escalate when multiple speakers speak simultaneously or when there is background speech while the primary user is speaking. Specialized vocabulary presents another challenge. Even though Speech-to-Text models can recognize a wide variety of words, they may stumble upon unique terms or industry-specific jargon not included in the model's vocabulary, resulting in potential transcription errors. Accents and dialects within the same language can pose

another difficulty. If a speaker's accent deviates substantially from the norm the model has been trained on; transcription accuracy might decline. Finally, language mismatch can considerably affect transcription accuracy. If the language the user speaks differs from what the Speech-to-Text application expects, the transcription will likely be less accurate. For example, if the system is set to transcribe English, but the user speaks Arabic, the output will likely be flawed [88].

Some approaches may combine audio with image or video features to capture a more comprehensive representation of dysarthric speech, such as in [6] where audio-visual joint features are used as input to the CNN and evaluated on UA Speech dataset with an accuracy up to 99.5%. Despite the high performance of their work, their techniques face a few limitations related to manual data pre-processing and high computational power requirements. In [89], the authors applied several types of features rather than mixing them as one input to the classifier to check which type is more effective in classifying dysarthria based on severity levels, where they used Generalized Morse Wavelet (GMW)-based scalogram features (for low-frequency areas) and Mel spectrogram-based features with CNN. They evaluated their work on the UA Speech dataset and achieved the highest accuracy using Scalgram features at 95.17%. Another work where the same has been done in [90] where several experiments on MFCC, the audio combination of features including(ZCR (Zero Crossing Rate), Spectral centroid, spectral roll-off) and Mel spectrogram using KNN, and SVM classifiers and evaluated their work on Torgo with an accuracy of 95% using SVM for mel spectrogram.

It's important to note that the choice of features depends on the specific goals and requirements of the dysarthria severity assessment task. Different features may carry different levels of information and may be more or less suitable for various applications. Researchers and practitioners often select and combine features based on their relevance and effectiveness in capturing the characteristics of dysarthric speech.

A. DISCUSSION OF RESEARCH QUESTION 1

Several types of features were used across the studies: The most common type of feature is the Acoustic feature. These include MFCCs (Mel Frequency Cepstral Coefficients), voice quality, and spectral and prosodic features (such as pitch and rhythm). In a general analysis, most studies used a combination of these features. Some studies focused on features Domain-specific features of speech disorders, such as breathiness features [5] and parameters related to glottal function [53]. Hand-crafted and raw waveform features were used by one study that combined these two approaches [54].

As shown in Figure 5, the mean accuracy obtained by the most utilized classification techniques representing the highest accuracy for each group of the datasets based on audio features reached up to 90s. Let's take, for example, how different levels of severity are shown when pronouncing the same sentence or the same word, as shown in Figure 4
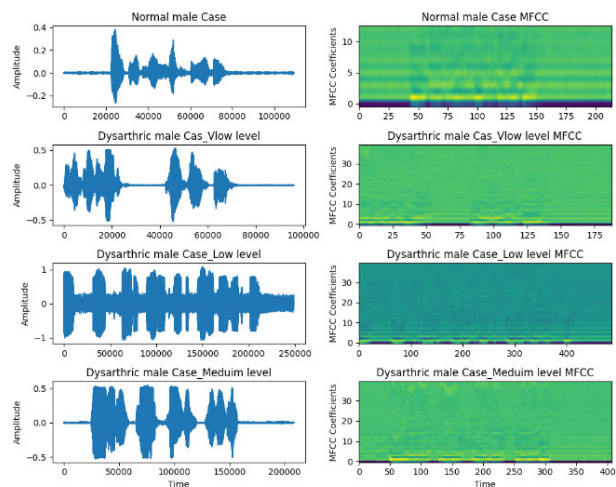
FIGURE 4. Audio signals vs their 40-MFCC features for pronouncing the sentence: 'Well, he is nearly ninety-three years old' extracted from control and three levels of dysarthria severity from TORGO dataset.

The dysarthric cases show variability in amplitude, with the "Low level" having the most variability. As the severity level increases from "Very Low" to "Medium," the overall amplitude of the waveform seems to increase, suggesting louder vocal efforts or disruptions in speech. Concerning the MFCCs, the "Very Low" and "Low" levels have MFCCs that appear more diffuse, with fewer distinct patterns. The "Medium level" appears slightly more structured than the other dysarthric cases but is still less defined than the normal case. This indicates that relying solely on MFCCs might not capture all the nuances associated with the different severity levels of dysarthria, such that it can easily explore the medium level but is difficult to differentiate the very low and low levels. As the severity level progresses, the distinctiveness in the MFCC patterns seems to decrease, which may indicate a loss in speech clarity or more speech abnormalities. This fact is applied to each feature extraction type where we cannot guarantee catching all the distinct patterns among different levels of severity.

As a proposed solution to differentiate these highly diffused types of severity levels, using some specific features, [53] like pitch features, temporal features, and specific spectral features. Also, Complex, multidimensional feature sets (a combination of features) tend to perform better and are associated with high accuracy rates that could solve the issue of variability among several levels of dysarthria severity. For example, the study [55] used a mix of spectral, cepstral, and frame-level features and achieved 99.9% with the UA Speech dataset. While combining multiple features and techniques can result in high accuracies, it can also introduce limitations such as increased computational time [53] and difficulties in interpreting feature importance [56].

Prosodic features, including articulation, often demonstrate high performance in dysarthria classification. These features capture essential aspects of speech production, such as rhythm, intonation, and speech clarity, which are crucial for assessing dysarthria severity. In [58] reported that

prosodic features are the best for classifying the mild cases (which are better classified compared to severe and moderate levels due to having more common features among speakers), and the moderate cases over the spectral features, which achieved remarkable results for classifying severe cases. Cepstral features are less effective for classifying severity levels. This is also approved by [57] and [76].

Some other techniques explored individual features for the classification purposes of ineligibility levels or dysarthria severity levels and discovered some significant features that can achieve a high accuracy rate, such as in [5], which relied on several types of breathiness features represented by Jitter, Shimmer, Harmonic-to-Noise Ratio, Harmonic Energy, Harmonic Energy of Residue, Harmonic-to-Signal Ratio, and Glottal-to-Noise Excitation Ratio and reached 96%. In their work, they discovered that Harmonic-related features could distinguish intelligibility levels and achieve the highest accuracy compared to other features but struggled with mild dysarthria. Also, in [67], where the authors rely only on the MFCC features set and achieved up to 99% accuracy rate after experiencing several trails with different utterances lengths, 50, 200, and 300 and found that increasing the length of utterances will lead to better performance. A similar attempt relying on MFCC is in [91].

Concerning image-based features, the Mel spectrograms and their log or derivatives are the primary types of features and are primarily evaluated on TORGO and UA Speech datasets, as shown in Figure 6. This is a benefit of using only one type to save the preparation and extraction of too many features, as the network is responsible for achieving this task.

The mean accuracies for classifying dysarthria based on severity levels varied across different datasets. Among the selected datasets, the TORGO dataset exhibited moderate mean accuracies ranging from 73.99% (KNN) to 92.47% (DL), with a notable performance from CNN (88.15%). The UA Speech dataset showed higher mean accuracies, with MLP achieving 98.17% accuracy, followed closely by LSTM (96.94%) and CNN (96.23%). The Qolt dataset had lower mean accuracies, with SVM and RF achieving approximately 70% accuracy. Notably, the Numours dataset demonstrated high mean accuracies of 94.45% (SVM) and 95.8% (RF). The Others dataset had an outstanding performance, with SVM and CNN achieving mean accuracies of 96% and 91.8%, respectively.

These mean accuracies provide insights into the effectiveness of classifiers for different datasets in the context of severity-based dysarthria classification. It is important to note that the performance of classifiers can vary depending on the dataset characteristics, such as the number of samples, data quality, and diversity of dysarthria types.

B. DISCUSSION OF RESEARCH QUESTION 2

Regarding the classification techniques, advanced algorithms such as Support Vector Machines (SVM), Discriminant analysis, and deep learning models (DNN, CNN, LSTM)

TABLE 2. Summary of the related works used image-based features for classifying dysarthria based on the severity levels.

Method	Feature	Classifier	Dataset
[43]	Mel spectrograms	Multi-head attention mechanism (MHA), Multi-task learning, and ResNet	UA Speech
[92]	Mel spectrograms	Squeeze-and-excitation (SE) networks	UA Speech
[93]	perceptually enhanced Fourier transform spectrograms and constant-Q transform (CQT) spectrograms	CNN	UA speech and TORGO
[94]	log Mel spectrograms	CNN	60 patients from each amyotrophic lateral sclerosis (ALS), Parkinson's disease (PD), and healthy controls
[95]	Log-mel spectrograms	LSTM networks with Attention mechanism	UA Speech
[96]	2D image spectrograms	DySARNet, a densely squeezed-and-excited attention-gated residual neural network	TORGO and UA-Speech
[97]	Spectrograms of short speech segments	ResNet (CNNs, GMM, and Light CNNs for comparison)	UA Speech
[98]	Spectrogram and its derivatives	CNN	UA Speech and TORGO

were frequently seen and generally delivered a strong performance. For example, SVM was widely utilized across various studies, including [31], [40], [54], [56], [57], [58], [60], [63], [66], [72], [74], and [76].

ANN-based models, seen in studies such as [31], [55], [56], [58], [59], and [60]. RF showed in [31], [58], [60], and [76], and Deep learning techniques (CNN, DNN, LSTM, etc.), applied in studies [53], [61], [71], [77], yielded high accuracies, typically above 90%.

Newer techniques like self-supervised learning with Wav2vec 2.0 XLS-R, as seen in [54], didn't perform as well, achieving only a 65.52% accuracy rate. This highlights the potential challenges in adapting these models for dysarthria severity classification.

From Figure 4 , audio-based classification, SVM and RF demonstrated relatively high mean accuracies across most datasets. In the TORGO dataset, SVM achieved a mean accuracy of 77.28%, while RF achieved 79.17% of mean accuracy. In the UA Speech dataset, both SVM and RF achieved accuracies above 80%, with SVM reaching 80.25% and RF reaching 82.69%. However, in the Qolt dataset, the mean accuracies for SVM and RF dropped to 70% and 70%, respectively. It's important to note that RF achieved a high accuracy of 95.8% in the Numours dataset, while SVM achieved a remarkable accuracy of 96% in the Others dataset.

For image-based classification in Figure 5, CNN and LSTM were the prominent techniques, as deep learning techniques is a direct way to extract features and classify them into several severity levels using their different layers

functions. In the TORGO dataset, CNN achieved an accuracy of 88.15%, while LSTM achieved 99.2% accuracy. In the UA Speech dataset, CNN demonstrated a high accuracy of 96.23%, and LSTM achieved 96.94% accuracy. Notably, ResNet only yields accuracy for the UA Speech datasets around 97%.

These results suggest that SVM and RF show consistent performance across various datasets for audio-based classification. Despite the effectiveness of using SVM, it is reported in kadi2016fully that it is not adequate to process utterances with diverse time lengths and needs to unify the time lengths for all speech utterances. For image-based classification, CNN and LSTM exhibit high accuracies. However, further investigation and evaluation are needed to determine these techniques' generalizability and performance on more extensive and diverse dysarthria datasets.

The mean accuracies presented in the tables were obtained by calculating the average of the highest accuracies reported in the original papers for each dataset group. To ensure accuracy and reliability, only the highest accuracy value from each paper was considered. This approach allows for a comprehensive analysis of the performance of the classification techniques across different datasets, taking into account the number of papers that reported the respective accuracies. By aggregating the highest accuracies, we provide a representative measure of the overall performance of the techniques for each dataset group, considering the number of papers that contributed to calculating the mean accuracies.
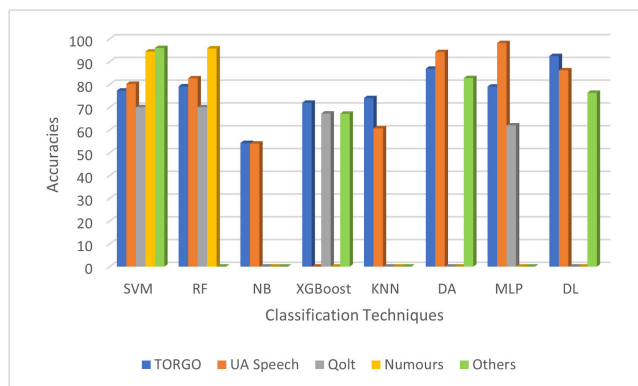
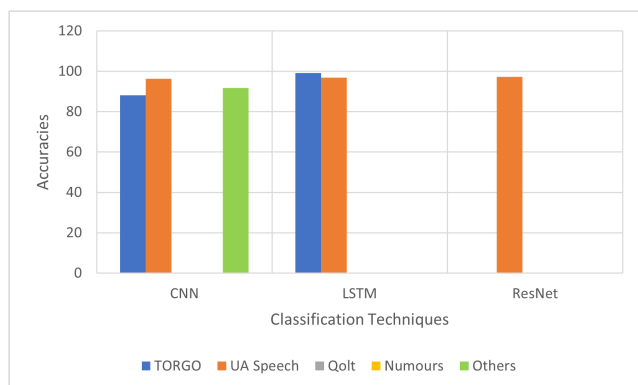**FIGURE 5.** Mean accuracy for audio-based features for common datasets.

FIGURE 6. Mean accuracy for image-based features for common datasets.

VI. LIMITATIONS

The audio-based features used in dysarthria classification have certain limitations, which can be compared to image-based features. By analyzing these limitations, we can gain insights into the techniques and accuracies associated with each feature type.

Regarding audio-based features, one limitation mentioned in [5] is the struggle to accurately classify cases with mild dysarthria. This implies that audio-based features might be more effective in capturing severe dysarthria characteristics, while their discriminatory power decreases with milder cases. Additionally, there are difficulties in interpreting the features and handling data diversity [31]. This suggests that some audio-based features may lack clear interpretability and may not be robust enough to handle the diverse dysarthria characteristics. Furthermore, there are limitations related to small datasets and the focus on a single type of dysarthria [58], which can hinder the generalizability of the findings.

On the other hand, image-based features also have their limitations. For instance, inter-subject variability and gender bias can impact the performance of image-based models [92]. These limitations suggest that certain image-based features might be more influenced by individual differences, making them less reliable for general dysarthria classification.

Another limitation is the computational time required for specific image-based techniques [93], which can hinder their real-time application.

When comparing the techniques and accuracies, it is essential to note that the performance varies across different studies and datasets. Audio-based features have reported accuracies ranging from 40.41% to 95.80% [58], whereas image-based features have achieved accuracies ranging from 72% to 99.20% [93], [96]. These variations highlight the influence of the chosen techniques and datasets on the achieved accuracies.

In summary, both audio-based and image-based features are limited in the classification of dysarthria. Audio-based features struggle with mild dysarthria cases, interpretation difficulties, and limited datasets, while image-based features face challenges related to inter-subject variability, gender bias, and computational time. These limitations influence the techniques used with these features and the resulting accuracies, emphasizing the need to carefully consider the feature type, technique selection, and dataset characteristics in dysarthria classification research.

VII. FUTURE DIRECTIONS

Differentiating different dysarthria severity levels faces many challenges, as discussed earlier. Researchers must follow several directions to deal with different cases, which may be shown with less variability, making classifying them difficult or causing high false positive results. Some of these directions:

Solving the Issue of Larger Data Size: Instead of utilizing only the standard machine learning techniques or the deep learning techniques which require large amounts of the dataset, we suggest using rule-based models, which outperform the other methods in terms of the necessity to small data size for training either in terms of features or samples, the more interpretable model explicitly compared to deep learning, adaptable and generalized model due to the generated rules which can be adjusted based on the specific application, the used data nature and the non-linearity and noise handling [99], [100].

Experiencing New Features: Some new features are suggested to be used by researchers in this field, including Nonlinear Dynamical Features, Vocal Tract Resonance Estimation, Source-Filter Separation Features, Microprosody Analysis, and Articulatory Kinematics.

Nonlinear Dynamical Features: Dysarthria affects the coordination and control of speech production, resulting in altered dynamics. Nonlinear dynamical features, such as recurrence plots(can help visualize temporal patterns in speech features that could indicate varying levels of severity [101]), Variations in the fractal dimension of speech signals could distinguish between different severity levels of dysarthria [102], or entropy measures (e.g., permutation entropy, sample entropy) (lower entropy indicates a less complex, more predictable one, and dysarthria's speech signals may become less complicated due to muscle weakness

and coordination difficulties, can capture the underlying complexity and temporal organization of dysarthric speech signals [103]). We know these features have not been utilized to classify dysarthria. Still, it is used with speech analysis, such as in [104], and it is helpful to be calculated and combined with video-based features.

Vocal Tract Resonance Estimation: Investigate techniques for estimating the vocal tract resonances directly from the speech signal. This could involve formant tracking, subspace-based analysis, or model-based estimation. This feature type was utilized for classifying dysarthria subtypes in [105]. These features alone may not be sufficient for classifying dysarthria based on severity levels. They primarily capture information about the resonant characteristics of the vocal tract, which can be influenced by various factors such as vocal fold movement, articulatory precision, and vocal tract shape. So, combining them with other acoustic features may help build a successful classification model.

Source-Filter Separation Features: Dysarthria can affect the speaker's source (glottal excitation) and filter (vocal tract) components. Extracting features that specifically focus on the characteristics of the glottal source, such as measures related to the glottal flow derivative or instantaneous frequency, in combination with standard spectral features, can provide a comprehensive representation for dysarthria classification. These features have been utilized in [106] for automatic dysarthric speech recognition and evaluated on the TORGO dataset, which contains diverse severity samples. Still, there is no specific classification model based on these features.

Microprosody Analysis: Investigate features that capture fine-grained temporal variations within short speech segments, called microporosity. This could involve analyzing pitch fluctuations, intensity modulations, or spectral changes at very short time scales (e.g., using wavelet transforms or time-frequency representations). These features can potentially capture subtle dysarthric speech dynamics. These types of features used mainly with speech synthesis and its variations were also explored in many studies such as [107].

Articulatory Kinematics: Consider exploring articulatory and kinematic features, such as lip or jaw motion tracking. Using techniques like optical motion capture or electromagnetic articulography, features can be extracted that describe articulatory gestures' spatial and temporal characteristics, which may be informative for dysarthria classification. These features have a notable correlation with the severity levels of dysarthria as discussed in [108], and they have been used to classify vocal segments into control, PD, and ALS, but still no specific work on classifying dysarthria based on severity utilizing this type of features.

VIII. CONCLUSION

Accurate classification of dysarthria according to its severity levels holds significant implications for clinicians and researchers. Such precision offers profound insights into the ramifications of the disorder on one's communicative

capabilities. Through this classification, clinicians can devise bespoke therapeutic strategies, employ objective evaluation metrics, and meticulously track progress, enhancing the life quality of those with dysarthria. This comprehensive review endeavored to critically evaluate the stratification of dysarthria based on its severity, with particular emphasis on the efficacy of diverse features and the prowess of artificial intelligence methodologies. An in-depth perusal of relevant literature revealed a marked preference for audio-based features, predominantly in acoustic analysis, as the principal modality in dysarthria classification. Nevertheless, no isolated feature can delineate all gradations of dysarthria severity. The most promising approach thus emerges as a synergistic amalgamation of features designed to holistically address all severity levels while mitigating the challenges posed by limited data availability.

Interestingly, the selection of AI methodologies doesn't inherently guarantee superior accuracy. Both conventional and cutting-edge techniques have demonstrated fluctuating efficacies. Yet, it is noteworthy that deep learning algorithms, especially when paired with image-based features, have proven adept at extracting intricate, high-dimensional features, thereby facilitating the facile classification of diverse severity strata. As we gaze into the horizon of dysarthria research, it becomes imperative to pivot towards gender-agnostic and language-neutral models. Current literature indeed signals an inadvertent bias skewed towards these dimensions, emphasizing the exigency of redressing this oversight in future endeavors.

DECLARATION OF COMPETING INTEREST

The authors declare no conflicts of interest.

REFERENCES

[1] L. D. Shriberg, E. A. Strand, K. J. Jakielski, and H. L. Mabie, "Estimates of the prevalence of speech and motor speech disorders in persons with complex neurodevelopmental disorders," *Clin. Linguistics Phonetics*, vol. 33, no. 8, pp. 707–736, Aug. 2019.

[2] J. R. Duffy, *Motor Speech Disorders: Substrates, Differential Diagnosis, and Management*. Amsterdam, The Netherlands: Elsevier, 2019.

[3] R. D. Kent and J. C. Rosenbek, "Acoustic patterns of apraxia of speech," *J. Speech, Lang., Hearing Res.*, vol. 26, no. 2, pp. 231–249, Jun. 1983.

[4] M. Laganaro, C. Fougeron, M. Pernon, N. Levúque, S. Borel, M. Fournet, S. C. Chiuvé, U. Lopez, R. Trouville, L. Ménard, P. R. Burkhard, F. Assal, and V. Delvaux, "Sensitivity and specificity of an acoustic- and perceptual-based tool for assessing motor speech disorders in French: The MonPaGe-screening protocol," *Clin. Linguistics Phonetics*, vol. 35, no. 11, pp. 1060–1075, Nov. 2021.

[5] H. M. Chandrashekar, V. Karjigi, and N. Sreedevi, "Breathiness indices for classification of dysarthria based on type and speech intelligibility," in *Proc. Int. Conf. Wireless Commun. Signal Process. Netw. (WiSPNET)*, Mar. 2019, pp. 266–270.

[6] H. Tong, "Automatic assessment of dysarthric severity level using audio-video cross-modal approach in deep learning," M.S. thesis, Unitec Inst. Technol., Auckland, New Zealand, 2020.

[7] G. Noffs, T. Perera, S. C. Kolbe, C. J. Shanahan, F. M. C. Boonstra, A. Evans, H. Butzkueven, A. van der Walt, and A. P. Vogel, "What speech can tell us: A systematic review of dysarthria characteristics in multiple sclerosis," *Autoimmunity Rev.*, vol. 17, no. 12, pp. 1202–1209, Dec. 2018.

[8] P. Gandhi, S. Tobin, M. Vongphakdi, A. Copley, and K. Watter, "A scoping review of interventions for adults with dysarthria following traumatic brain injury," *Brain Injury*, vol. 34, no. 4, pp. 466–479, Mar. 2020.

[9] C. Mackenzie, "Dysarthria in stroke: A narrative review of its description and the outcome of intervention," *Int. J. Speech-Language Pathol.*, vol. 13, no. 2, pp. 125–136, Apr. 2011.

[10] J. Lee, A. Madhavan, E. Krajewski, and S. Lingenfelter, "Assessment of dysarthria and dysphagia in patients with amyotrophic lateral sclerosis: Review of the current evidence," *Muscle Nerve*, vol. 64, no. 5, pp. 520–531, Nov. 2021.

[11] E. Finch, A. F. Rumbach, and S. Park, "Speech pathology management of non-progressive dysarthria: A systematic review of the literature," *Disability Rehabil.*, vol. 42, no. 3, pp. 296–306, Jan. 2020.

[12] C. Whillans, M. Lawrie, E. A. Cardell, C. Kelly, and R. Wenke, "A systematic review of group intervention for acquired dysarthria in adults," *Disability Rehabil.*, vol. 44, no. 13, pp. 3002–3018, Jun. 2022.

[13] Y.-J. Park and J.-M. Lee, "Effect of acupuncture intervention and manipulation types on poststroke dysarthria: A systematic review and meta-analysis," *Evidence-Based Complementary Alternative Med.*, vol. 2020, pp. 1–17, Sep. 2020.

[14] A. K. Shanmugam and R. Marimuthu, "A critical analysis and review of assistive technology: Advancements, laws, and impact on improving the rehabilitation of dysarthric patients," in *Handbook of Decision Support Systems for Neurological Disorders*. Amsterdam, The Netherlands: Elsevier, 2021, pp. 263–281.

[15] H. P. Rowe, S. E. Gutz, M. F. Maffei, K. Tomanek, and J. R. Green, "Characterizing dysarthria diversity for automatic speech recognition: A tutorial from the clinical perspective," *Frontiers Comput. Sci.*, vol. 4, p. 43, Apr. 2022.

[16] R. Chiaramonte and M. Vecchio, "A systematic review of measures of dysarthria severity in stroke patients," *PM&R*, vol. 13, no. 3, pp. 314–324, Mar. 2021.

[17] R. Palmer and P. Enderby, "Methods of speech therapy treatment for stable dysarthria: A review," *Adv. Speech Lang. Pathol.*, vol. 9, no. 2, pp. 140–153, Jan. 2007.

[18] A. Huang, K. Hall, C. Watson, and S. R. Shahamiri, "A review of automated intelligibility assessment for dysarthric speakers," in *Proc. Int. Conf. Speech Technol. Human-Comput. Dialogue (SpeD)*, Oct. 2021, pp. 19–24.

[19] K. L. Stipancic, K. M. Palmer, H. P. Rowe, Y. Yunusova, J. D. Berry, and J. R. Green, "'You say severe, I say mild': Toward an empirical classification of dysarthria severity," *J. Speech, Lang., Hearing Res.*, vol. 64, no. 12, pp. 4718–4735, Dec. 2021.

[20] M. R. McNeil, K. J. Ballard, J. R. Duffy, J. Wambaugh, P. van Lieshout, B. Maassen, and H. Terband, "Apraxia of speech theory, assessment, differential diagnosis, and treatment: Past, present, and future," in *Speech Motor Control in Normal and Disordered Speech: Future Developments in Theory and Methodology*. Rockville, MD, USA: ASHA Press, 2017, pp. 195–221.

[21] J. M. Barkmeier-Kraemer and H. M. Clark, "Speech–language pathology evaluation and management of hyperkinetic disorders affecting speech and swallowing function," *Tremor Other Hyperkinetic Movements*, vol. 7, p. 489, Sep. 2017.

[22] P. Enderby, "Frenchay dysarthria assessment," *Brit. J. Disorders Commun.*, vol. 15, no. 3, pp. 165–173, 1980.

[23] K. M. Yorkston, D. R. Beukelman, and C. Traynor, "Assessment of intelligibility of dysarthric speech," Austin, TX, USA, Tech. Rep., 1984.

[24] S. J. Robertson, "Dysarthria profile," in *Communication Skill Builders*. Buckinghamshire: Winslow, Communication Skill Builders, 1987.

[25] B. H. Jacobson, A. Johnson, C. Grywalski, A. Silbergleit, G. Jacobson, M. S. Benninger, and C. W. Newman, "The voice handicap index (VHI) development and validation," *Amer. J. Speech-Language Pathol.*, vol. 6, no. 3, pp. 66–70, 1997.

[26] D. B. Freed, *Motor Speech Disorders: Diagnosis and Treatment*. San Diego, CA, USA: Plural Publishing, 2018.

[27] S. Blaustein and A. Bar, "Reliability of perceptual voice assessment," *J. Commun. Disorders*, vol. 16, no. 2, pp. 157–161, Mar. 1983.

[28] S. A. S. A. Yaari and N. Almaflehi, "Validation of communication activities of daily living- second edition (CADL-2) on Arab aphasics: Controlled study," *J. Study English Linguistics*, vol. 1, no. 2, p. 99, Dec. 2013.

[29] A. M. Altaher, S. Y. Chu, R. B. M. Kam, and R. A. Razak, "A report of assessment tools for individuals with dysarthria," *Open Public Health J.*, vol. 12, no. 1, pp. 384–386, Sep. 2019.

[30] J. I. Godino-Llorente and P. Gomez-Vilda, "Automatic detection of voice impairments by means of short-term cepstral parameters and neural network based detectors," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 2, pp. 380–384, Feb. 2004.

[31] A. Hernandez, S. Kim, and M. Chung, "Prosody-based measures for automatic severity assessment of dysarthric speech," *Appl. Sci.*, vol. 10, no. 19, p. 6999, Oct. 2020.

[32] A. Gallardo-Antolín and J. M. Montero, "An auditory saliency pooling-based LSTM model for speech intelligibility classification," *Symmetry*, vol. 13, no. 9, p. 1728, Sep. 2021.

[33] J. C. V. Correa, T. Arias, J. R. Orozco-Arroyave, and E. Nöth, "A multitask learning approach to assess the dysarthria severity in patients with Parkinson's disease," in *Proc. INTERSPEECH*, Sep. 2018, pp. 456–460.

[34] L. Parisi, N. RaviChandran, and M. L. Manaog, "Feature-driven machine learning to improve early diagnosis of Parkinson's disease," *Expert Syst. Appl.*, vol. 110, pp. 182–190, Nov. 2018.

[35] L. Xu, J. Liss, and V. Berisha, "Dysarthria detection based on a deep learning model with a clinically-interpretable layer," *JASA Exp. Lett.*, vol. 3, no. 1, Jan. 2023, Art. no. 015201.

[36] R. D. Kent, G. Weismer, J. F. Kent, H. K. Vorperian, and J. R. Duffy, "Acoustic studies of dysarthric speech: Methods, progress, and potential," *J. Commun. Disorders*, vol. 32, no. 3, pp. 141–186, 1999.

[37] S. Sapir, L. O. Ramig, J. L. Spielman, and C. Fox, "Formant centralization ratio: A proposal for a new acoustic measure of dysarthric speech," *J. Speech, Lang., Hearing Res.*, vol. 53, no. 1, pp. 114–125, Feb. 2010.

[38] H. Kim, M. Hasegawa-Johnson, and A. Perlman, "Vowel contrast and speech intelligibility in dysarthria," *Folia Phoniatrica Logopaedica*, vol. 63, no. 4, pp. 187–194, 2011.

[39] J. R. Orozco-Arroyave, J. D. Arias-Londoño, J. F. Vargas-Bonilla, M. C. Gonzalez-Rátiva, and E. Nöth, "New Spanish speech corpus database for the analysis of people suffering from Parkinson's disease," in *Proc. LREC*, 2014, pp. 342–347.

[40] G. Tartarisco, R. Bruschetta, S. Summa, L. Ruta, M. Favetta, M. Busà, A. Romano, E. Castelli, F. Marino, A. Cerasa, T. Schirinzi, M. Petrarca, E. Bertini, G. Vasco, and G. Pioggia, "Artificial intelligence for dysarthria assessment in children with ataxia: A hierarchical approach," *IEEE Access*, vol. 9, pp. 166720–166735, 2021.

[41] D. Kempler and D. V. Lancker, "Effect of speech task on intelligibility in dysarthria: A case study of Parkinson's disease," *Brain Lang.*, vol. 80, no. 3, pp. 449–464, Mar. 2002.

[42] U. Ayvaz, H. Gürüler, F. Khan, N. Ahmed, T. Whangbo, and A. A. Bobomirzaevich, "Automatic speaker recognition using mel-frequency cepstral coefficients through machine learning," *Comput., Mater. Continua*, vol. 71, no. 3, pp. 5511–5521, 2022.

[43] A. A. Joshy and R. Rajan, "Dysarthria severity classification using multi-head attention and multi-task learning," *Speech Commun.*, vol. 147, pp. 1–11, Feb. 2023.

[44] A. Bandini, J. R. Green, B. Taati, S. Orlandi, L. Zinman, and Y. Yunusova, "Automatic detection of amyotrophic lateral sclerosis (ALS) from video-based analysis of facial movements: Speech and non-speech tasks," in *Proc. 13th IEEE Int. Conf. Autom. Face Gesture Recognit. (FG)*, May 2018, pp. 150–157.

[45] H. Ackermann, I. Hertrich, and G. Scharf, "Kinematic analysis of lower lip movements in ataxic dysarthria," *J. Speech, Lang., Hearing Res.*, vol. 38, no. 6, pp. 1252–1259, Dec. 1995.

[46] W. Xue, R. van Hout, C. Cucchiarini, and H. Strik, "Assessing speech intelligibility of pathological speech in sentences and word lists: The contribution of phoneme-level measures," *J. Commun. Disorders*, vol. 102, Mar. 2023, Art. no. 106301.

[47] W. Xue, R. van Hout, C. Cucchiarini, and H. Strik, "Assessing speech intelligibility of pathological speech: Test types, ratings and transcription measures," *Clin. Linguistics Phonetics*, vol. 37, no. 1, pp. 52–76, Jan. 2023.

[48] K. Lehner and W. Ziegler, "The impact of lexical and articulatory factors in the automatic selection of test materials for a Web-based assessment of intelligibility in dysarthria," *J. Speech, Lang., Hearing Res.*, vol. 64, no. 6S, pp. 2196–2212, Jun. 2021.

[49] N. Japkowicz and M. Shah, *Evaluating Learning Algorithms: A Classification Perspective*. Cambridge, U.K.: Cambridge Univ. Press, 2011.

[50] D. M. W. Powers, "Evaluation: From precision, recall and F-measure to ROC, informedness, markedness and correlation," 2020, *arXiv:2010.16061*.

[51] A. P. Bradley, "The use of the area under the ROC curve in the evaluation of machine learning algorithms," *Pattern Recognit.*, vol. 30, no. 7, pp. 1145–1159, Jul. 1997.

[52] D. Ruppert, "The elements of statistical learning: Data mining, inference, and prediction," *J. Amer. Stat. Assoc.*, vol. 99, no. 466, p. 567, Jun. 2004.

[53] A. A. Joshy and R. Rajan, "Automated dysarthria severity classification: A study on acoustic features and deep learning techniques," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 30, pp. 1147–1157, 2022.

[54] E. J. Yeo, K. Choi, S. Kim, and M. Chung, "Automatic severity classification of dysarthric speech by using self-supervised model with multi-task learning," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Jun. 2023, pp. 1–5.

[55] H. M. Chandrashekar, V. Karjigi, and N. Sreedevi, "Speech intelligibility assessment of dysarthria using Fisher vector encoding," *Comput. Speech Lang.*, vol. 77, Jan. 2023, Art. no. 101411.

[56] E. J. Yeo, S. Kim, and M. Chung, "Automatic severity classification of Korean dysarthric speech using phoneme-level pronunciation features," in *Proc. Interspeech*, Aug. 2021, pp. 4838–4842.

[57] K. L. Kadi, S. A. Selouani, B. Boudraa, and M. Boudraa, "Automated diagnosis and assessment of dysarthric speech using relevant prosodic features," in *Transactions on Engineering Technologies: Special Volume of the World Congress on Engineering*. New York, NY, USA: Springer, 2013, pp. 529–542.

[58] B. A. Al-Qatab and M. B. Mustafa, "Classification of dysarthric speech according to the severity of impairment: An analysis of acoustic features," *IEEE Access*, vol. 9, pp. 18183–18194, 2021.

[59] C. Bhat, B. Vachhani, and S. K. Kopparapu, "Automatic assessment of dysarthria severity level using audio descriptors," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Mar. 2017, pp. 5070–5074.

[60] A. Hernandez, E. J. Yeo, S. Kim, and M. Chung, "Dysarthria detection and severity assessment using rhythm-based metrics," in *Proc. Interspeech*, Oct. 2020, pp. 2897–2901.

[61] A. Kachhi, A. Therattil, A. T. Patil, H. B. Sailor, and H. A. Patil, "Teager energy cepstral coefficients for classification of dysarthric speech severity-level," in *Proc. Asia–Pacific Signal Inf. Process. Assoc. Annu. Summit Conf. (APSIPA ASC)*, Nov. 2022, pp. 1462–1468.

[62] M. S. Paja and T. H. Falk, "Automated dysarthria severity classification for improved objective intelligibility assessment of spastic dysarthric speech," in *Proc. 13th Annu. Conf. Int. Speech Commun. Assoc.*, Sep. 2012, pp. 62–65.

[63] E. J. Yeo, K. Choi, S. Kim, and M. Chung, "Cross-lingual dysarthria severity classification for English, Korean, and Tamil," in *Proc. Asia–Pacific Signal Inf. Process. Assoc. Annu. Summit Conf. (APSIPA ASC)*, Nov. 2022, pp. 566–574.

[64] K. Kadi, S. Selouani, B. Boudraa, and M. Boudraa, "Discriminative prosodic features to assess the dysarthria severity levels," in *Proc. World Congr. Eng.*, vol. 3, 2013, pp. 2201–2205.

[65] G. Vyas, M. K. Dutta, J. Prinosil, and P. Harár, "An automatic diagnosis and assessment of dysarthric speech using speech disorder specific prosodic features," in *Proc. 39th Int. Conf. Telecommun. Signal Process. (TSP)*, Jun. 2016, pp. 515–518.

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

[67] M. Purohit, M. Parmar, M. Patel, H. Malaviya, and H. A. Patii, "Weak speech supervision: A case study of dysarthria severity classification," in *Proc. 28th Eur. Signal Process. Conf. (EUSIPCO)*, Jan. 2021, pp. 101–105.

[68] A. Kachhi, A. Therattil, A. T. Patil, H. B. Sailor, and H. A. Patil, "Significance of energy features for severity classification of dysarthria," in *Proc. 24th Int. Conf. Speech Comput. (SPECOM)*, Gurugram, India. New York, NY, USA: Springer, 2022, pp. 325–337.

[69] V. M. Ramos, A. Lowit, L. Van den Steen, H. A. K. Hernandez-Diaz, M. E. H.-D. Huici, M. De Bodt, and G. Van Nuffelen, "Acoustic identification of sentence accent in speakers with dysarthria: Cross-population validation and severity related patterns," *Brain Sci.*, vol. 11, no. 10, p. 1344, Oct. 2021.

[70] K. Gurugubelli and A. K. Vuppala, "Analytic phase features for dysarthric speech detection and intelligibility assessment," *Speech Commun.*, vol. 121, pp. 1–15, Aug. 2020.

[71] A. A. Joshy and R. Rajan, "Automated dysarthria severity classification using deep learning frameworks," in *Proc. 28th Eur. Signal Process. Conf. (EUSIPCO)*, Jan. 2021, pp. 116–120.

[72] N. P. Narendra and P. Alku, "Automatic intelligibility assessment of dysarthric speech using glottal parameters," *Speech Commun.*, vol. 123, pp. 1–9, Oct. 2020.

[73] S. Gillespie, Y.-Y. Logan, E. Moore, J. Laures-Gore, S. Russell, and R. Patel, "Cross-database models for the classification of dysarthria presence," in *Proc. Interspeech*, Aug. 2017, pp. 3127–3131.

[74] K. L. Kadi, S. A. Selouani, B. Boudraa, and M. Boudraa, "Fully automated speaker identification and intelligibility assessment in dysarthria disease using auditory knowledge," *Biocybernetics Biomed. Eng.*, vol. 36, no. 1, pp. 233–247, 2016.

[75] H. Dahmani, S. A. Selouani, N. Doghmane, D. O'Shaughnessy, N. A. Shaughnessy, and M. Chetouani, "On the relevance of using rhythmic metrics and SVM to assess dysarthric severity," *Int. J. Biometrics*, vol. 6, no. 3, pp. 248–271, 2014.

[76] A. A. Joshy, P. N. Parameswaran, S. R. Nair, and R. Rajan, "Statistical analysis of speech disorder specific features to characterise dysarthria severity level," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Jun. 2023, pp. 1–5.

[77] F. Javanmardi, S. Tirronen, M. Kodali, S. R. Kadiri, and P. Alku, "Wav2vec-based detection and severity level classification of dysarthria from speech," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Jun. 2023, pp. 1–5.

[78] A. DeMino, R. Kubichek, and K. Caves, "Assessing dysarthria severity using global statistics and boosting," in *Proc. Conf. Rec. 45th Asilomar Conf. Signals, Syst. Comput. (ASILOMAR)*, Nov. 2011, pp. 1103–1106.

[79] F. Rudzicz, A. K. Namasivayam, and T. Wolff, "The TORGO database of acoustic and articulatory speech from speakers with dysarthria," *Lang. Resour. Eval.*, vol. 46, no. 4, pp. 523–541, Dec. 2012.

[80] H. Kim, M. Hasegawa-Johnson, A. Perlman, J. Gunderson, T. S. Huang, K. Watkin, and S. Frame, "Dysarthric speech database for universal access research," in *Proc. 9th Annu. Conf. Int. Speech Commun. Assoc.*, 2008, pp. 1741–1744.

[81] D.-L. Choi, B.-W. Kim, Y.-W. Kim, Y.-J. Lee, Y. Um, and M. Chung, "Dysarthric speech database for development of QoLT software technology," in *Proc. LREC*, 2012, pp. 3378–3381.

[82] X. Menendez-Pidal, J. B. Polikoff, S. M. Peters, J. E. Leonzio, and H. T. Bunnell, "The Nemours database of dysarthric speech," in *Proc. 4th Int. Conf. Spoken Lang. (ICSLP)*, vol. 3, Oct. 1996, pp. 1962–1965.

[83] R. Kent and K. Hustad, "Speech production: Development," in *Encyclopedia of Neuroscience*. 2009, pp. 255–264.

[84] K. López-de-Ipiña, J.-B. Alonso, C. Travieso, J. Solé-Casals, H. Egiraun, M. Faundez-Zanuy, A. Ezeiza, N. Barroso, M. Ecay-Torres, P. Martinez-Lage, and U. Lizardui, "On the selection of non-invasive methods based on speech analysis oriented to automatic Alzheimer disease diagnosis," *Sensors*, vol. 13, no. 5, pp. 6730–6745, May 2013.

[85] S. Luz, "Longitudinal monitoring and detection of Alzheimer's type dementia from spontaneous speech data," in *Proc. IEEE 30th Int. Symp. Comput.-Based Med. Syst. (CBMS)*, Jun. 2017, pp. 45–46.

[86] G. S. Turner, K. Tjaden, and G. Weismer, "The influence of speaking rate on vowel space and speech intelligibility for individuals with amyotrophic lateral sclerosis," *J. Speech, Lang., Hearing Res.*, vol. 38, no. 5, pp. 1001–1013, Oct. 1995.

[87] G. Weismer, J.-Y. Jeng, J. S. Laures, R. D. Kent, and J. F. Kent, "Acoustic and intelligibility characteristics of sentence production in neurogenic speech disorders," *Folia Phoniatrica Logopaedica*, vol. 53, no. 1, pp. 1–18, 2001.

[88] S. Modale, G. Sable, and R. Deshmukh, "A review: Devnagri speech to text for Marathwada region," in *Proc. Int. Conf. Wireless Commun. (ICWiCOM)*. New York, NY, USA: Springer, 2019, pp. 525–532.

[89] A. Kachhi, A. Therattil, P. Gupta, and H. A. Patil, "Continuous wavelet transform for severity-level classification of dysarthria," in *Proc. (SPECOM)*, Gurugram, India. New York, NY, USA: Springer, 2022, pp. 312–324.

[90] S. A. Kumar, T. Keerthivasan, and S. Sasikala, "Towards improving the performance of dysarthric speech severity assessment system," in *Proc. Int. Conf. Comput. Commun. Informat. (ICCCI)*, Jan. 2022, pp. 1–6.

[91] S. H. Lee, M. Kim, H. G. Seo, B.-M. Oh, G. Lee, and J.-H. Leigh, "Assessment of dysarthria using one-word speech recognition with hidden Markov models," *J. Korean Med. Sci.*, vol. 34, no. 13, pp. 1–9, 2019.

[92] A. A. Joshy and R. Rajan, "Dysarthria severity assessment using squeeze-and-excitation networks," *Biomed. Signal Process. Control*, vol. 82, Apr. 2023, Art. no. 104606.

[93] H. Chandrashekar, V. Karjigi, and N. Sreedevi, "Investigation of different time-frequency representations for intelligibility assessment of dysarthric speech," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 12, pp. 2880–2889, Dec. 2020.

[94] B. Suhas, J. Mallela, A. Illa, B. Yamini, N. Atchayaram, R. Yadav, D. Gope, and P. K. Ghosh, "Speech task based automatic classification of ALS and Parkinson's disease and their severity using log mel spectrograms," in *Proc. Int. Conf. Signal Process. Commun. (SPCOM)*, Jul. 2020, pp. 1–5.

[95] M. Fernández-Díaz and A. Gallardo-Antolín, "An attention long short-term memory based system for automatic classification of speech intelligibility," *Eng. Appl. Artif. Intell.*, vol. 96, Nov. 2020, Art. no. 103976.

[96] F. J. Montalbo, "Dysarnet: A densely squeezed-and-excited attention-gated residual deep learning model for dysarthric speech recognition and severity estimation," *SSRN J.*, Jan. 2023.

[97] S. Gupta, A. T. Patil, M. Purohit, M. Parmar, M. Patel, H. A. Patil, and R. C. Guido, "Residual neural network precisely quantifies dysarthria severity-level based on short-duration speech segments," *Neural Netw.*, vol. 139, pp. 105–117, Jul. 2021.

[98] H. M. Chandrashekar, V. Karjigi, and N. Sreedevi, "Spectro-temporal representation of speech for intelligibility assessment of dysarthria," *IEEE J. Sel. Topics Signal Process.*, vol. 14, no. 2, pp. 390–399, Feb. 2020.

[99] S. Akkoç, "An empirical comparison of conventional techniques, neural networks and the three stage hybrid adaptive neuro fuzzy inference system (ANFIS) model for credit scoring analysis: The case of Turkish credit card data," *Eur. J. Oper. Res.*, vol. 222, no. 1, pp. 168–178, Oct. 2012.

[100] A. Subasi, "Application of adaptive neuro-fuzzy inference system for epileptic seizure detection using wavelet feature extraction," *Comput. Biol. Med.*, vol. 37, no. 2, pp. 227–244, Feb. 2007.

[101] O. Geman, "Data processing for Parkinson's disease: Tremor, speech and gait signal analysis," in *Proc. E-Health Bioengineering Conf. (EHB)*, Nov. 2011, pp. 1–4.

[102] A. P. Accardo and E. Mumolo, "An algorithm for the automatic differentiation between the speech of normals and patients with Friedreich's ataxia based on the short-time fractal dimension," *Comput. Biol. Med.*, vol. 28, no. 1, pp. 75–89, Jan. 1998.

[103] F. Rudzicz, "Towards a noisy-channel model of dysarthria in speech recognition," in *Proc. Workshop Speech Lang. Process. Assistive Technol. (NAACL HLT)*, 2010, pp. 80–88.

[104] E. S. Jackson, M. Tiede, M. A. Riley, and D. H. Whalen, "Recurrence quantification analysis of sentence-level speech kinematics," *J. Speech, Lang., Hearing Res.*, vol. 59, no. 6, pp. 1315–1326, Dec. 2016.

[105] N. Lévêque, A. Slis, L. Lancia, G. Bruneteau, and C. Fougeron, "Acoustic change over time in spastic and/or flaccid dysarthria in motor neuron diseases," *J. Speech, Lang., Hearing Res.*, vol. 65, no. 5, pp. 1767–1783, May 2022.

[106] Z. Yue, E. Loweimi, and Z. Cvetkovic, "Raw source and filter modelling for dysarthric speech recognition," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2022, pp. 7377–7381.

[107] M. Vainio and T. Altosaar, "Modeling the microprosody of pitch and loudness for speech synthesis with neural networks," in *Proc. ICSLP*, 1998, Paper 0886.

[108] J. Lee, M. Bell, and Z. Simmons, "Articulatory kinematic characteristics across the dysarthria severity spectrum in individuals with amyotrophic lateral sclerosis," *Amer. J. Speech-Language Pathol.*, vol. 27, no. 1, pp. 258–269, Feb. 2018.

AFNAN AL-ALI was born in Iraq. She received the Master of Science degree in computer engineering from the University of Basra, Basra, Iraq, and the Ph.D. degree in computing from Qatar University. Her research interests include machine learning, AI, computer vision, object tracking, health informatics, and fuzzy logic.

SOMAYA AL-MAADEED (Senior Member, IEEE) received the Ph.D. degree in computer science from the University of Nottingham, U.K., in 2004. She is currently with the Department of Computer Science and Engineering, Qatar University, Doha, Qatar, and the Computer Vision and AI Research Group Coordinator. She enjoys excellent collaboration with national and international institutions and industry. She is a Principal Investigator of several funded research projects generating approximately five million. She has published extensively in the field of pattern recognition and delivered workshops on teaching programming for undergraduate students. She has attended workshops on higher education strategy, assessment methods, and interactive teaching. She was elected as the IEEE Chair for the Qatar Section, in 2015.

MOUTAZ SALEH received the Ph.D. degree in computer science from Universiti Kebangsaan Malaysia, in 2008. He is currently an Assistant Professor with the Department of Computer Science and Engineering, Qatar University. He has published over 70 refereed papers in reputable international journals and conferences. His research interests include assistive technology, adaptive learning, computer vision, and artificial intelligence. He has served as a Technical Program Committee for *Computer Communications* (Elsevier), IEEE TRANSACTIONS ON LEARNING TECHNOLOGIES, *Education and Information Technologies* (Springer), *Multimedia Tools and Applications* (Springer), IEEE ACCESS, and IEEE (LCN, ICT, AICCSA, APCC, GCC, ICC, EDUCON, and TALE) conferences. He received several professional awards and collaborated as a Principal Investigator in multiple internal and externally funded research projects.

RANI CHINNAPPA NAIDU received the B.Eng. (Hons.) and M.Tech. degrees in electrical engineering from VIT University, Vellore, India, in 1998 and 2006, respectively, and the Ph.D. degree in electrical engineering from Northumbria University, Newcastle upon Tyne, U. K., in 2014. After that, she joined the Northumbria Photovoltaic Applications Centre, Northumbria University, as a Postdoctoral Researcher. She is currently an Associate Professor with VIT University. She leads an appreciable number of research groups and projects in areas, such as solar photovoltaic, wind energy, power generation dispatch, power system optimization, and artificial intelligence techniques.

ZACHARIAH C. ALEX was born in Kerala, India, in March 1964. He received the B.Sc. degree in physics and the M.Sc. degree in physics with electronics from Kerala University, Thiruvananthapuram, Kerala, in 1984 and 1986, respectively, and the Ph.D. degree from Jamia Millia Islamia Central University, New Delhi, India, in 1996. He is currently a Professor with the Higher Academic Level with the Department of Sensor and Biomedical Technology, School of Electronics Engineering, VIT University, Vellore, India. His current research interests include antenna design, microwave electronics, metamaterials, MEMS, and fiber optic sensors for health care.

PRAKASH RAMACHANDRAN received the bachelor's degree in engineering from the Government College of Technology (GCT), Coimbatore, India, the master's degree in engineering (embedded systems) from the College of Engineering Guindy, Anna University, and the Ph.D. degree in compressive sensing from Vellore Institute of Technology (VIT), Chennai. He is currently an Associate Professor with the School of Electronics Engineering, VIT, Vellore. He is also a Tech Lead with Visteon in a speech recognition project for two years. His research interests include image processing, machine learning, artificial intelligence, data science, embedded systems, and the Internet of Things (IoT).

RAJEEV KHOODEERAM received the B.Eng. degree (Hons) in computer science and engineering from the University of Mauritius in 2001, the M.Sc. degree in information and communication technologies from the Electrical and Engineering Department, University of Mauritius in 2005, and the Ph.D. degree in computer science (specialisation bioinformatics) from the Université de Côte d'Azur, France, in 2021. Prior to joining the University, he has worked in the industry for almost three years both full time and as freelancer as a Software Engineer. He is working as a Senior lecturer with the Université des Mascareignes (UdM). He is currently supervising undergraduate projects and is the Founder/Head of the Emerging Technologies Department, Faculty of Sustainable Development and Engineering. Being enthusiastic and futurist, he founded the Center of Research on Emerging Technologies and Innovation in 2023, under which there are three laboratory's namely: Remote Sensing, Intelligent Systems and BioInformatics (he is the Head of this laboratory). His main areas of expertise are artificial intelligence, the Internet of Things, and bioinformatics.

RAJESH KUMAR M (Member, IEEE) received the B.E. degree in electronics and communication engineering from the University of Madras, Chennai, India, in 2000, the M.E. degree in electrical drives and embedded control from the College of Engineering Guindy, Anna University, Chennai, in 2005, and the M.B.A. degree in systems from Alagappa University, Karaikudi, India, in 2008. He was a Senior Research Assistant with the Department of Mathematics and Information Science, Northumbria University, Newcastle upon Tyne, U.K. Since 2000, he has been teaching and researching in India and the U.K. He is currently an Associate Professor with VIT University, Vellore, India. He has published appreciable several papers in refereed international journals and IEEE conferences. His research group includes speech processing, biomedical signal processing, acoustic signal, image security, data analytics, perceptual hashing, pattern recognition, machine intelligence, and optimization techniques. He received the Research Scholarships Award for his Ph.D. degree from Northumbria University, from 2010 to 2013. He also received an e-COST Grant 2013 to participate in the 10th edition of the International Summer School on Biometrics, Alghero, Italy. He was awarded the VIT University SEED Grant 2019. He received a grant from the Royal Academy of Engineering, U.K. He is a Chartered Engineer of IET, U.K.

• • •