Indonesian Journal of Electrical Engineering and Computer Science
Vol. 36, No. 2, November 2024, pp. 826~836
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v36.i2.pp826-836

Exploring the landscape of dysarthric speechrecognition: a survey of literature

Aneeta S. Antony[1], Rohini Nagapadma[1], Ajish K. Abraham[2]
[1]Department of Electronics and Communication Engineering, The National Institute of Engineering, Affiliated to Visvesvaraya Technological University, Mysuru, India
[2]Department of Electronics, All India Institute of Speech and Hearing, Mysuru, India

Article Info

Article history:

Received Feb 4, 2024
Revised May 21, 2024
Accepted Jun 5, 2024

Keywords:

ASR
Deep learning
Dysarthria speech recognition
Machine learning

ABSTRACT

Automatic speech recognition (ASR) is a valued tool for individuals with dysarthria, a speech impairment characterized by various pathological traits that differ from healthy speech. However, recognizing dysarthric speech, which is spoken by individuals with speech impairments, poses unique challenges due to its diverse characteristics such as rugged pronunciation, loudness that varies at different intervals, speech that has lot of delays, pauses that are inpredictable, excessive nasal sounds, explosive pronunciation, and airflow noise. The survey reveals the various models for dysarthric speech recognition. Deep learning technologies, unfurls an improved ASR performance leaps and bounds breaking the fluency and pronunciation barriers. Various feature extractions and identification of different types of dysarthria, including spastic, mixed, ataxic, hypokinetic, and hyperkinetic are explored. The performance of contemporary deep learning approaches in dysarthric speaker recognition (DSR) is tested using various datasets to determine accuracy. In conclusion the most effective DSR strategies are identified and areas for future investigation is suggested. However, speaker-dependent difficulties restrict the generalizability of acoustic models, and a lack of speech data impedes training on large datasets. The study throws light on how the effectiveness of ASR for dysarthric speech can be improved and further areas of research in the area are highlighted.

Corresponding Author:

Aneeta S. Antony
Department of Electronics and Communication, The National Institute of Engineering
Mysuru, Karnataka, India
Email: aneetaantony7@gmail.com

1. INTRODUCTION

Dysarthria, a disability affecting articulation, resonance, phonation, and prosody, is one of the speech diseases brought on by neurological abnormalities in the speech muscles [1]. Accurate speech recognition systems struggle with dysarthric speech due to its varied manifestations, including spastic, flaccid, ataxic, hypokinetic, and hyperkinetic types [2]. While subjective diagnosis is frequently used among speech-language pathologists, subjective diagnosis, although it is costly and time-consuming [3]-[5]. Consequently, a great deal of research has been done to enhance automatic speech recognition (ASR) applications for dysarthric speakers because of advancements in artificial intelligence (AI) [6]-[8] and computational capabilities.

Additionally, the pronunciation of words by dysarthric speakers varies significantly due to the underlying causes of their dysarthria [9]-[11]. Rather than normal speech dysarthric speech has more intricate differences which are more challenging to comprehend [12]. Therefore, comprehending the various types of dysarthria and the difficulties associated with feature extraction is crucial for creating a high speech recognition rate that is both customized and efficient. The dysarthria eventually becomes apparent using machine learning (ML) based approaches like support vector machines (SVM) and other models [13], [14]. Owing to the chaotic and unpredictable nature of dysarthric speech patterns, recognition algorithms frequently have low accuracy [15] and require fine-tuning to improve their precision. Also, it was discovered that longer training periods were effective for marginally to severely dysarthric speakers when employing speaker-dependent and ASR systems. Yet, this strategy was shown to be incredibly time-consuming for both clinicians and end users [16], [17].

Deep learning models being prominent in the area of dysarthric recognition, the major goal of this review is to examine a selection of studies that explore the use of the same [18], [19]. One is to detect dysarthria in voice recordings [20]-[23] besides performing recognition as well. This involves utilizing acoustic-phonetic, prosodic, and voice quality features to identify different types of dysarthria, including spastic, mixed, ataxic, hypokinetic, and hyperkinetic dysarthria. The survey from 2016 through 2024, concentrate on research using a variety of datasets that correspond to the five different forms of dysarthria. The goal is to assess the performance of various models and identify which features and recognition models are most efficient in diagnosing certain types of dysarthria in patients. Additionally, the analysis considers factors such as accuracy efficiency when working with dysarthria datasets.

2. METHOD

Dysarthric speech recognition entails obtaining acoustic parameters from raw speech data, which are then modelled and contrasted in a lexicon. Accurate speech recognition necessitates a controlled setting with little background noise and a decent quality microphone. Deep learning classifiers have been verified for feature extraction strategies that improve dysarthria speech recognition accuracy [24]-[26]. Recognition algorithms must be strong enough to handle the variety of dysarthria kinds. Feature extraction and recognition methods are critical for dyslexic speech recognition because they enable precise, efficient transcription, which improves communication and quality of life. More exact recognition is possible by separating distinguishing aspects of dysarthric speech using deep-learning feature extraction techniques [27]-[29]. This might speed up transcribing, saving physicians and patients time and effort while also boosting their communication skills.

2.1. Investigating advanced deep learning techniques for categorizing dysarthria features and recognizing dysarthria types

To address the challenges posed by dysarthria, researchers have explored integrating additional features, such as facial expressions and lip movements, into speech recognition systems. Feature extraction is crucial for identifying unique properties of normal and dysarthric speech-an optimal feature extraction set- ting for ASR system performance based on speaker selection. The feature extraction and ASR steps are analysed using several methods [30]-[32].

Following that, [33]used a deep belief network (DBN) to analyze acoustic features and an artificial neural network (ANN) for classification, achieving a 97% accuracy on the UA speech dataset. However, accuracy dropped to 74% with an unbalanced dataset. Subsequently, a new audio dataset of 10-second PATA test recordings from 55 patients (18 healthy, 37 with ataxia) showed the proposed HMLM achieved 90% accuracy at the first level (healthy vs patients) by identifying an optimal subset of characteristics [34]. Despite this, accuracy improvement is needed. Using mel-spectrogram analysis, empirical mode decomposition (EMD), and convolutional neural network (CNNs), the accuracy for dysarthric speaker recognition (DSR) in spastic dysarthria was 83%, though concerns about low accuracy and delays were noted. For the same condition, deep neural network (DNNs) alone achieved a lower accuracy of 57%, highlighting recognition challenges. Ataxic dysarthria recognition accuracy improved to 90% with a pre-trained network. Despite advancements, all studies encountered persistent challenges in achieving high accuracy, reflecting the ongoing efforts and obstacles in DSR Affected individuals often produce phonemes that are notably imprecise, characterized by pitch pauses in vocalic segments and inaccuracies in the articulation of consonants [35]-[37]. So, Shahamiri [38] presented a transfer learning approach using CNN with voice gram for extracting the phonemes and recognizing the spastic dysarthria with 67% accuracy on UA-speech. Robust dysarthric speech recognition, focusing on individuals with spastic and mixed dysarthria, utilizing a sequential contrastive learning framework, using the TORGO dataset, the study achieved a 78% accuracy rate [39]. However, the study noted low accuracy, indicating ongoing challenges in this area. To improve the accuracy, phonemes feature extraction [40] was analyzed in the publicly available dataset to recognize spastic dysarthria with efficient frequency, time, acoustic, and phoneme feature extraction through a perceptual linear prediction approach. Hence, 82% accuracy was achieved, yet limited availability of pathological speech data was indeed something to be looked into.

Another approach where a discriminative learning based classifier was used. The log likelihood scores from an example specific hidden markov models (ESHMMs) was fed in vector form into a discriminative classifier [41]. The accuracy superseded that of conventional HMM model or even DNN HMM models. Word recognition accuracy was the best way to calculate word level accuracy. The data scarcity problem is addressed by using meta learning approach [42]. These methods incorporate knowledge from different dysarthric speakers, allowing the model to adapt rapidly to new speakers. The results on the UASpeech dataset show that the adapted models achieve a significant word error rate reduction. Furthermore, dysarthria severity-level analysis using spectrogram analysis and a residual neural network (RNN) reached an astounding 98% accuracy rate [43]. This result marks a huge step forward in comprehending dysarthria severity levels and evaluating speech patterns. Comparable research employing conditional information feature extraction (CIFE) on auditory characteristics yielded an impressive 95% accuracy [44]. Assessing the level of impairment may be beneficial for selecting sorting or creating variability in training data. However, the study noted issues with low accuracy in particular areas.

The CNN trained on the mel-spectrogram was shown to be excellent in detecting slurred speech [45]. Whilst other models outperformed the DNN-HMM model, when trained on the TORGO dataset to recognize spastic, mixed dysarthria speech, it reached 89% recognition accuracy [46]. It has dysarthria severity levels, pause insertion, pitch, energy, and duration controls. Despite the quality of the voice, acoustic characteristics were not fully evaluated. The difficulty of a limited data set and its consequences can be mitigated by employing a transfer learning strategy [47] with CNN for identifying spastic, mixed dysarthria. In addition, a mel-spectrogram was employed to recognize auditory, texture, and articulatory information, resulting in 97.73% speech recognition accuracy. Distinguishing between normal and dysarthric speech and identifying the specific type of dysarthria is a significant challenge, as many systems often misinterpret it as noise. To address this, an AI model was developed to detect ataxic and hypokinetic dysarthria from normal speech, as well as from speech affected by Parkinson’s disease and cerebellar ataxia [48]. The model demonstrated higher accuracy with a gradual recognition strategy for ataxic dysarthria, but although it had high recognition speed for hypokinetic detection, its accuracy was limited. Additionally, the severity of dysarthria was assessed using DNNs and auditory characteristics [49]. The study focused on people with spastic dysarthria and used the UA Speech dataset. The findings implies that correct classification of dysarthria severity levels remains a difficulty, highlighting the need for more study and improvement in this field [50]-[52].

When the amount of available data is limited, transfer learning emerges as a promising approach. This method was applied to develop a spastic dysarthric ASR system on the UA-Speech Corpus, achieving an accuracy of 59.78%. Utilizing transfer learning in the custom construction of models proves to be an intriguing approach. Despite limited dataset availability, automatic DSR has shown improved accuracy. For instance, a study referenced in [53] achieved a significant milestone in dysarthric speech detection, reaching an accuracy of 97.45% through mel-spectrogram analysis combined with Google Net. While this accomplishment is noteworthy, similar studies have also acknowledged enduring challenges, including low accuracy rates and delays. These findings collectively underscore the need for continual refinement and enhancement in dysarthric speech recognition and voice disorder classification research. In another study [54], a fusion approach was implemented using deep learning techniques on acoustic features with CNNs. Despite challenges related to low accuracy, the study reported a 95% accuracy rate on the UA speech dataset. Similarly, [55] squeeze-and-excitation (SE) networks were employed on mel spectrograms, emphasizing spectral frequency and time features. By utilizing deep CNN models with SE blocks, this research achieved a 95% accuracy rate on the UA-Speech dataset, although errors and low accuracy were reported. These outcomes highlight the persistent challenges in achieving high accuracy in dysarthric speech recognition, despite the effective application of advanced techniques such as deep learning and SE networks.

Following this, the focus shifted to analyzing datasets like UA and TORGO for spastic mixed-dysarthria. Yue *et al.* [56] and Jolad and Khanai [57] developed a CNN and a speech enhancement generative adversarial network (SEGAN) enhanced with a fractional competitive crow search algorithm (FCCSA). This algorithm combined fractional calculus with competitive swarm optimization to improve speech signal quality, resulting in higher accuracy for SEGAN-FCCSA. To further enhance accuracy in detecting extrapyramidal dysarthria, a personalized phrase recognition system was created by [58] using the public easy call dataset, achieving a 78% accuracy rate. However, this approach did not focus on voice features specific to extrapyramidal dysarthria. Subsequently, dysarthria assessment in children with ataxia was examined using a machine learning approach. While this method resulted in better accuracy, it also introduced delays in the process. Isaev *et al.* [59], research on dysarthria assessment in children with Ataxia was undertaken using deep learning techniques on entropy, frequency, and intensity variables. Using CNNs on the Ataxic dataset, the study attained an accuracy rate of 93%. Despite the comparatively high accuracy, the study found low accuracy and evaluation delays. Farhadipour and Veisi [60], the focus was on dysarthric speech processing, using AlexNet on spectral energy characteristics. Using CNNs on the UA speech dataset resulted in a slightly lower accuracy of 91.29%. Notably, the study showed issues in effective feature extraction, demonstrating that dysarthric speech processing remains a barrier despite advances in deep learning approaches.

Table 1 in Appendix lists these aspects for a better understanding of the many models used for dysarthria detection and speech recognition, along with the databases and accuracy ratings linked to each. The primary objective is to evaluate and determine the top models and databases for voice recognition in dysarthria. Further relevant research is also examined, such as the classification of voice abnormalities and the detection of dysarthria, in order to get insight into potential techniques or combinations of approaches.

Studies demonstrate that mel-spectrogram analysis and CNNs may accurately diagnose dysarthria and assess intelligibility. But issues like inaccuracies and delays continue. SE networks, fusion approaches, and sequential contrastive learning frameworks are examples of advanced techniques that show promise but are difficult to achieve high accuracy. Larger feature sets, unbalanced datasets, and limited access to pathological speech data are among the problems. Research is moving in the direction of more accurate and efficient dysarthric speech recognition systems in spite of these obstacles

3. RESULTS AND DISCUSSION

The study investigated thoroughly the types of dysarthria and it has explored the challenges faced in the recognition of dysarthric speech. The influence of deep learning in the area of speech has copiously increased the degrees of automation in this field. The temporal distribution of articles published between 2016 and 2024 is depicted in Figure 1. Table 1 vividly gives an at a glance idea to the various methods used and types of database were extracted from each article, and the following subsections present a direct comparison of these items across the reviewed articles. Besides, Figure 2 shows the distribution of the included paper counts chosen for this study based on types of dysarthria in dataset groups. It reveals interesting patterns in the distribution of documents across different categorized dysarthria such as spastic, ataxic, hypokinetic, and hyperkinetic Dysarthria and dataset groups like TORGO [61], UA speech [62], TORGO and UA speech, public easy call dataset, independent larger dataset, IRCCS, numours dataset, and SMCND. In recent years, deep learning contributed enormous efforts in speech processing to bring the highest level of automation. These being the most widely utilized data sets to detect, analyze, classify, and recognize the dysarthric speech. Still, the performance of a deep learning-based system depends upon several constraints, and the gap from the survey of recent research in speech processing based on accuracy is listed below. The challenges of dysarthria recognition are analyzed to compare the effectiveness of classifiers for classifying types of dysarthria. Based on this, various datasets and spastic, spastic with mixed, hypokinetic, spastic-hypokinetic, extrapyramidal, and ataxic types of dysarthria with extracting feature types [63] were also analyzed using multiple deep-learning approaches in Tables 2 to 5.

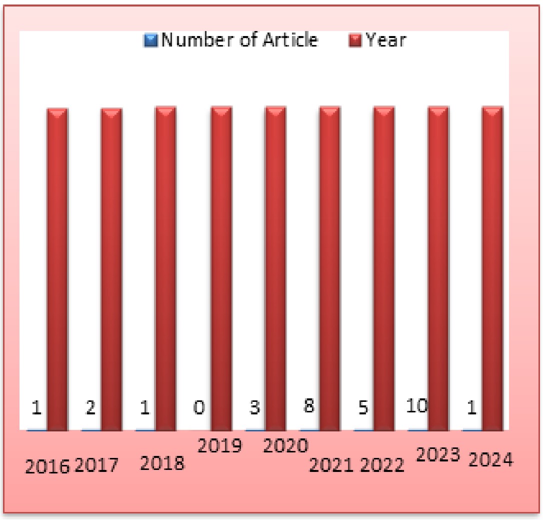

Figure 1. Number of reviewed articles based on year

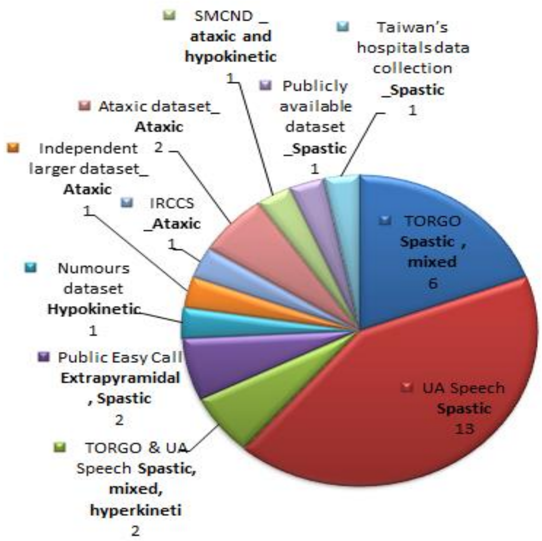

Figure 2. No of papers based on datasets and types of dysarthria

Tables 2 to 5 analyze standard datasets for dysarthria types, describing feature extraction and recognition classifiers. Feature analysis-based speech recognition is efficient for dysarthria speech. CNN-based deep learning classifiers achieve better accuracy in spastic dysarthria recognition, spastic-mixed dysarthria type recognition, and types of ataxic dysarthria analysis ANN and HMLM models. Overall, CNN-based deep learning classifiers outperform other methods. Spastic dysarthria analysis achieved low accuracy of 67% using CNN and voice gram feature extraction methods. Bi-LSTM and RNNs achieved 59% and 98% accuracy, respectively. However, energy operator feature extraction for speech recognition is low. Limited availability of pathological speech data and delays in processing or recognition are significant challenges. Additionally, crucial features in model training are absent, leading to low accuracy. Comprehensive strategies for improving model performance are needed.

Spastic dysarthria, caused by damage to upper motor neurons, is the most common type, causing slow, effortful speech with strained vocal quality and imprecise articulation. Advanced techniques like mel-spectrogram analysis and deep learning classifiers can achieve high accuracy in identifying this condition. Spastic-mixed dysarthria combines spastic and other dysarthria features, but presents unique classification difficulties. Ataxic dysarthria, characterized by incoordination of speech movements, can be identified using deep learning techniques like CNNs and HMLM. Hypokinetic dysarthria, linked to Parkinson’s disease, can be detected using 1D-CNN techniques, achieving 82% accuracy. Data augmentation, a technique in machine learning and data science, broadens dataset diversity, improving model performance, particularly in voice recognition, by considering changing speaking styles and background noise. Transformer architecture is increasingly being used in voice recognition due to its attention mechanism, bidirectional context, scalability, transfer learning, effective parallelization, and cutting-edge performance [64]. Transformers can handle longer input sequences, fine-tune models for smaller datasets, and train quicker than RNNs. A transformer encoder-decoder framework with a multiobjective training strategy, incorporating connectionist temporal classification and masked language modeling objectives, is explored for learning contextual bidirectional representations [65]. The model outperforms comparable models on multiple datasets and fine-tuning the top layers enhances performance, particularly on the fluent speech command dataset. Class attention is introduced for efficient spoken language understanding [66]. Data-driven techniques could improve word recognition, and phone-based applications could address speech-based emotion detection, suicidal tendencies [67], dysarthric speech, and stroke detection [68]. Further exploration is needed to develop a promising system. The survey of the literature reveals a dynamic landscape in dysarthric speech recognition, encompassing various types of dysarthria and employing diverse methodologies. The exploration underscores the efficacy of feature analysis-based speech recognition techniques, particularly; acoustic, prosody, spectral, and voice quality features have been efficiently extracted and utilized in dysarthria recognition classifiers.

Table 2. Analysing the accuracy for various classifiers on spastic and spastic mixed dysarthria

Dataset	Feature extraction method	Extracting feature types	Classifiers	Accuracy%
Nemours database	Mel-spectrogram	Acoustic	EMD-CNN	83
UA-Speech	Spectrogram	Teager energy operator	RNN	98
UA-Speech	Voicegram	Acoustic	CNN	67
Publicly available dataset	Perceptual linear prediction	Acoustic	CNN	82
UA-Speech	Whisper	Prosodic	Bi-lstm	59
UA-Speech	Visual representation extraction	Phoneme	SCNN- MHAT	91
UA-Speech	CNN	Acoustic	Fusion CNN	95
UA-Speech	DNN	Acoustic	DNN	93
UA-Speech	AlexNet	Spectral	CNN	91
Taiwan’s hospitals	Mel-spectrogram	Acoustic	Time delay neural	74

Table 3. Analysing the accuracy for various classifiers on spastic-mixed dysarthria type

Dataset	Feature extraction	Extracting feature types	Classifiers	Accuracy
TORGO	DNN	Prosodic:	DNN-HMM	89%
TORGO	Mel-spectrogram	Acoustic	CNN	97.73%
TORGO	CNN-LSTM	Acoustic	Sinusoidal rectified unit (SinRU) with CNN-LSTM	70.62%
TORGO	Mel-spectrogram	Spectral	Google Net	97.45%

Notably, the recognition of spastic dysarthria, deep learning classifiers of CNNs and ANN models have shown promising results in achieving rates as high as 97% across different dysarthria types. However, challenges persist, particularly in the recognition of hypokinetic and extrapyramidal dysarthria types, which have not received as much attention as spastic and ataxic dysarthria. The limited availability of pathological speech data poses a significant hurdle, leading to lower accuracy rates, especially in CNN models trained on datasets like "UA-Speech" and "Publicly available datasets." Moreover, shortcomings in model training, such as the exclusion of crucial features like voice or spectral features in the Bi-LSTM model applied to the spastic dataset, have led to lower accuracy rates.

This study focuses on the advancements as well as the unmet problems in the realm of deep learning approaches for the detection of dysarthria. The analysis leads to the following important conclusions: Spastic dysarthria is the one that has been studied the most. Current research pays less attention to other forms, such as mixed dysarthria and hypokinetic dysarthria. Mel-spectrogram analysis is a widely used technique for feature extraction that has demonstrated great accuracy rates in a number of investigations. Other methods have also been investigated, including CIFE and EMD. CNNs are widely utilised as deep learning classifiers for the identification of dysarthria. They have demonstrated strong performance in identifying various forms of dysarthria, particularly in conjunction with transfer learning methodologies. Variations exist in the accuracy rates among studies and dysarthria types, with some demonstrating remarkable outcomes (e.g., 98% accuracy for spastic dysarthria). Still, there are difficulties in continuously reaching high accuracy rates, particularly for uncommon forms of dysarthria.

Table 4. Analysis the accuracy for various classifiers on ataxic and hypokinetic dysarthria

Dataset	Feature extraction	Extracting feature types	Classifiers	Accuracy
SMCND	CNN	Spectral feature	CNN	Ataxic:90%
	feature selection approach	Acoustic	ANN	Hypokinetic:86%
IRCCS	VGG pre-trained network	Acoustic	HMLM	90%
Ataxic dataset	deep learning	Entropy, frequency, intensity features	CNN	93%
UA, TORGO	FCCSA	MKMFCC, pitch Chroma, spectral and noise	Generative adversarial network	93%

Table 5. Accuracy analysis of extrapyramidal

Dataset	Feature extraction	Extracting feature types	Classifiers	Accuracy
Public easy call	Latent vectors	Spectral	CNN	78%

4. CONCLUSION

In summary, this work has emphasized important findings and offered a clear path for further deep learning-based dysarthria identification research. With the drastic advancements in deep learning techniques. To improve the precision and resilience of recognition models, future studies should give priority to the gatheringand annotation of a wider variety of datasets, including uncommon types of dysarthria. Model architectures must be continuously improved. To enhance efficiency, researchers want to investigate innovative network designs or optimization strategies designed especially for dysarthria identification. Mel-spectrogram analysis has demonstrated potential, but more research into spectro-temporal features or deep learning features may improve the accuracy of recognition. Researching transfer learning and domain adaptation strategies may help models become more broadly applicable to a wider range of datasets and environmental factors. It is critical to develop methods for real-time dysarthria identification with shorter processing delays. Improving the effectiveness of recognition algorithms for real-time applications should be the main goal of future research. Adoption of dysarthria recognition systems requires its validation in clinical settings. To evaluate how well these technologies assist clinical decision-making and enhance patient outcomes, robust clinical trials should be conducted in future study. In the field of dysarthric voice recognition, it is critical to overcome issues brought about by small or skewed datasets. There are several strategies that may be used to lessen these difficulties. The dataset may be increased by employing data augmentation to produce different dysarthric speech sample versions. Yet again transfer learning provides a way to improve performance on smaller dysarthric speech datasets by using information from bigger, non-dysarthric speech datasets. Techniques for resampling data can aid in balancing the dataset, and by pooling predictions from several models, ensemble learning can increase the resilience of the model. Models trained on ordinary speech can be modified via domain adaptation approaches to more accurately identify dysarthric speech. Lastly, experimenting with various features or feature combinations specific to dysarthric speech might further improve recognition performance. When combined, these methods provide viable paths towards improving dysarthric speech.

Future research should focus on expanding datasets, refining model architectures, and exploring innovative feature extraction techniques to improve accuracy and robustness across diverse dysarthria types. By addressing these gaps, we can enhance the effectiveness and applicability of dysarthric speech recognition systems, ultimately benefiting individuals with speech disorders and facilitating better communication and quality of life. Overall, while significant progress has been made in dysarthria recognition using deep learning, there is still room for improvement. Future research should focus on addressing the remaining challenges and exploring new techniques to enhance the accuracy and efficiency of dysarthria recognition systems. In conclusion, while dysarthria presents significant challenges in speech recognition and classification, advances in deep learning and feature extraction techniques have shown promising results in accurately identifying and classifying different types of dysarthria. Continued research in this area is essential to further improve the accuracy and effectiveness of dysarthria detection and classification systems, ultimately the goal being to improve the quality of life for individuals affected by this condition.

APPENDIX

Table 1. Survey of dysarthria recognition using deep learning

Aim	Feature extraction	Extracting features	Deep learning classifier	Types of dysarthria	Database	Benefits	Drawback
Intelligibility assessment in dysarthria disease	Mel-spectrogram	Acoustic	Gaussian mixture models, SVM	Spastic, mixed	TORGO	Acc: 97%	Delay
DSR	feature selection approach	Acoustic, energy	ANN	ataxic	Ataxic dataset	Acc: 91%	Low accuracy, delay not taken extended features
DSR	Mel-spectrogram	Acoustic	DNN-HMM	Spastic	UA speech	Acc: 65%	Low accuracy
DSR	Mel-spectrogram	Acoustic	EMD and CNN	Spastic	Nemours database	Acc: 83%	Low accuracy
DSR	Mel-spectrogram	Acoustic	DNN	Spastic	UA-speech	Acc: 57%	Low accuracy
DSR	VGG pre-trained network	Acoustic, articulatory	HMLM	Ataxic	IRCCS	Acc:90%	Low accuracy
Automatic DSR via TL	Voicegram	Phoneme, acoustic features	CNN	Spastic	UA-speech	Acc: 67%	Low accuracy
Robust DSR		-	Sequential contrastive learning framework	Spastic, mixed	TORGO	Acc:78%	Low accuracy
DSR	Perceptual linear prediction	Frequency, time, acoustic, phoneme	CNN	Spastic	Publicly available dataset	Acc: 82%	Low accuracy limited availability of pathological speech data
Dysarthric ASR	Voicegrams	Voice quality	Spatial CNN	Spastic	UA-speech	Acc: 64%	Low accuracy
Dysarthric speech recognition	Batch normalization	Acoustic:	model-agnostic meta learning	Spastic	UA-speech	Acc: 54%	Low accuracy
dysarthria severity-level analysis	Spectrogram	Teager energy operator	RNN	Spastic	UA-speech	Acc: 98%	-
Classification of dysarthric speech	CIFE	Acoustic: prosody, spectral, and voice quality	ANN	Spastic	UA-speech	Acc: 95%	Low accuracy
Parkinson's disease detection	Mel-spectrogram	Articulation, phonation, voice and energy	1D-CNN	hypokinetic	Numours dataset	Acc: 82%	Low accuracy
Dysarthria severity level	DNN	Prosodic: Pitch, Energy, Duration	DNN-HMM	Spastic, mixed	TORGO	Acc: 89%	Low accuracy
DS detection using TL	Mel-spectrogram	Acoustic, texture and articulatory	CNN	Spastic, mixed	TORGO	Acc: 97.73%	Low accuracy High error
Detection and differentiation of ataxic and hypokinetic dysarthria	CNN	Spectral feature	CNN	Ataxic and hypokinetic	SMCND	Ataxic:90% Hypo kinetic 86%	Delay, low accuracy
Dysarthria severity classification	DNN	Acoustic	DNN	Spastic	UA Speech	Acc: 93%	Low accuracy

Table 1. Survey of dysarthria recognition using deep learning (*Continued*)

Aim	Feature extraction	Extracting features	Deep learning classifier	Types of dysarthria	Database	Benefits	Drawback
Automatic DSR using transfer learning	Whisper model	Prosodic	Bi-LSTM	Spastic	UA-Speech Corpus	Acc: 59.78%	Low accuracy Not include voice, spectral features
Voice disorder classification	CNN-LSTM	Acoustic	Sinusoidal rectified unit (SinRU) with CNN-LSTM	Spastic, mixed	TORGO	Acc: 70.62%	Low accuracy
Automatic DSR	Visual representation extraction	Phoneme	SCNN-MHAT	Spastic	UA-Speech	Acc: 94%	Low accuracy
Dysarthric speech detection	Mel-spectrogram	Spectral frequency	Google Net	Spastic, mixed	TORGO	Acc: 97.45%	Delay
Fusion approach	DL	Acoustic	CNN	Spastic	UA speech	Acc: 95%	Low accuracy
SE networks	mel spectrograms	Spectral: frequency, time	deep CNN models using SE blocks	Spastic	UA-speech	Acc: 95%	Low accuracy, error
DSR	CNN	Acoustic	CNN-RNN	Spastic, mixed	UA, TORGO	Acc: 88%	Low accuracy
Improve the quality of the speech signal	FCCSA	MKMFCC, pitch Chroma, spectral and noise	Generative adversarial network	Spastic, hyperkinetic	UA, TORGO	93% accuracy	Low accuracy
Personalized phrase recognition system	Latent vectors	Spectral	CNN	Extrapyramidal	Public easy call dataset	Acc: 78%	Low accuracy not taken voice, phonation features
Dysarthria Assessment in children with Ataxia	deep learning	Entropy, frequency, intensity features	CNN	Ataxic	Ataxic dataset	Acc: 93%	Low accuracy, delay
Dysarthric speech processing	AlexNet	Spectral: energy	CNN	Spastic	UA speech	Acc: 91.29	Low accuracy in efficient feature extraction

ACKNOWLEDGEMENTS

We wish to acknowledge the support given by Dr Reuben, Scientist from AIISH who helped in a deeper understanding of speech disorders and the patient variable's.

REFERENCES

[1] R. M. Hanifa *et al.*, "Voiced and unvoiced separation in Malay speech using zero crossing rate and energy," *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS)*, vol. 16, no. 2, pp. 775–780, Nov. 2019, doi: 10.11591/ijeecs.v16.i2.pp775-780.

[2] T. B. Adam, M. S. Salam, and T. S. Gunawan, "Wavelet cesptral coefficients for isolated speech recognition," *TELKOMNIKA Indonesian Journal of Electrical Engineering*, vol. 11, no. 5, 2013, doi: 10.11591/telkomnika.v11i5.2510.

[3] L. Rampello, L. Rampello, F. Patti, and M. Zappia, "When the word doesn't come out: a synthetic overview of dysarthria," *Journal of the Neurological Sciences*, vol. 369, pp. 354–360, Oct. 2016, doi: 10.1016/j.jns.2016.08.048.

[4] C. Ewen, H. Jenkins, C. Jackson, J. Jutley-Neilson, and J. Galvin, "Well-being, job satisfaction, stress and burnout in speech-language pathologists: a review," *International Journal of Speech-Language Pathology*, vol. 23, no. 2, pp. 180–190, Mar. 2021, doi: 10.1080/17549507.2020.1758210.

[5] S. R. Shareef and Y. F. Al-Irhayim, "Towards developing impairments arabic speech dataset using deep learning," *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS)*, vol. 25, no. 3, pp. 1400–1405, Mar. 2022, doi: 10.11591/ijeecs.v25.i3.pp1400-1405.

[6] V. Delić *et al.*, "Speech technology progress based on new machine learning paradigm," *Computational Intelligence and Neuroscience*, vol. 2019, pp. 1–19, Jun. 2019, doi: 10.1155/2019/4368036.

[7] M. S. Abu, S. R. Selamat, A. Ariffin, and R. Yusof, "Cyber threat intelligence – issue and challenges," *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS)*, vol. 10, no. 1, pp. 371–379, Apr. 2018, doi: 10.11591/ijeecs.v10.i1.pp371-379.

[8] I. Salhi, K. El Guemmat, M. Qbadou, and K. Mansouri, "Towards developing a pocket therapist: an intelligent adaptive psychological support chatbot against mental health disorders in a pandemic situation," *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS)*, vol. 23, no. 2, pp. 1200–1211, 2021, doi: 10.11591/ijeecs.v23.i2.pp1200-1211.

[9] Z. Qian and K. Xiao, "A survey of automatic speech recognition for dysarthric speech," *Electronics (Switzerland)*, vol. 12, no. 20, p. 4278, Oct. 2023, doi: 10.3390/electronics12204278.

[10] M. C. Stoian, S. Bansal, and S. Goldwater, "Analyzing ASR pretraining for low-resource speech-to-text translation," in *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2020, vol. 2020-May, pp. 7909–7913, doi: 10.1109/ICASSP40776.2020.9053847.

[11] H. A. Younis, A. S. A. Mohamed, M. N. Ab Wahab, R. Jamaludin, and S. Salisu, "A new speech recognition model in a human-robot interaction scenario using NAO robot: proposal and preliminary model," in *International Conference on Communication and Information Technology, ICICT 2021*, Jun. 2021, pp. 215–220, doi: 10.1109/ICICT52195.2021.9568457.

[12] T. F. Abidin, A. Misbullah, R. Ferdhiana, L. Farsiah, M. Z. Aksana, and H. Riza, "Acoustic model with multiple lexicon types for Indonesian speech recognition," *Applied Computational Intelligence and Soft Computing*, vol. 2022, pp. 1–16, Sep. 2022, doi: 10.1155/2022/3227828.

[13] B. Kashyap, M. Horne, P. N. Pathirana, L. Power, and D. Szmulewicz, "Automated topographic prominence based quantitative assessment of speech timing in cerebellar ataxia," *Biomedical Signal Processing and Control*, vol. 57, p. 101759, Mar. 2020, doi: 10.1016/j.bspc.2019.101759.

[14] I. Suwarno, R. Lalo, E. Costru-Tasnic, Jan, and H. Vanguru, "Sensors and transducers for stroke detection systematic literature review," *Sunan Kalijaga Journal of Physics*, 2023.

[15] V. Despotovic, O. Walter, and R. Haeb-Umbach, "Machine learning techniques for semantic analysis of dysarthric speech: an experimental study," *Speech Communication*, vol. 99, pp. 242–251, May 2018, doi: 10.1016/j.specom.2018.04.005.

[16] J.-C. Junqua, "Evaluation of ASR front ends in speaker-dependent and speaker-independent recognition," *The Journal of the Acoustical Society of America*, vol. 81, no. S1, pp. S93–S93, May 1987, doi: 10.1121/1.2024477.

[17] S. S. Nidhyananthan, R. S. S. Kumari, and V. Shenbagalakshmi, "Assessment of dysarthric speech using Elman back propagation network (recurrent network) for speech recognition," *International Journal of Speech Technology*, vol. 19, no. 3, pp. 577–583, Sep. 2016, doi: 10.1007/s10772-016-9349-1.

[18] D. Yu, G. Hinton, N. Morgan, J.-T. Chien, and S. Sagayama, "Introduction to the special section on deep learning for speech and language processing," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 20, no. 1, pp. 4–6, Jan. 2012, doi: 10.1109/TASL.2011.2173371.

[19] P. Janbakhshi, I. Kodrasi, and H. Bourlard, "Automatic dysarthric speech detection exploiting pairwise distance-based convolutional neural networks," in *ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Jun. 2021, vol. 2021-June, pp. 7328–7332, doi: 10.1109/ICASSP39728.2021.9413922.

[20] K. Hux, J. Rankin-Erickson, N. Manasse, and E. Lauritzen, "Accuracy of three speech recognition systems: case study of dysarthric speech," *AAC: Augmentative and Alternative Communication*, vol. 16, no. 3, pp. 186–196, Jan. 2000, doi: 10.1080/07434610012331279044.

[21] M. B. Akçay and K. Oğuz, "Speech emotion recognition: emotional models, databases, features, preprocessing methods, supporting modalities, and classifiers," *Speech Communication*, vol. 116, pp. 56–76, Jan. 2020, doi: 10.1016/j.specom.2019.12.001.

[22] J. Millet and N. Zeghidour, "Learning to detect dysarthria from raw speech," in *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, May 2019, vol. 2019-May, pp. 5831–5835.

[23] M. A. Motin, N. D. Pah, S. Raghav, and D. K. Kumar, "Parkinson's disease detection using smartphone recorded phonemes in real world conditions," *IEEE Access*, vol. 10, pp. 97600–97609, 2022, doi: 10.1109/ACCESS.2022.3203973.

[24] S. Nafisah, O. Wahyunggoro, and L. E. Nugroho, "Evaluating the usage of short-time energy on voice biometrics system for cerebral palsy," in *Proceedings of 2016 8th International Conference on Information Technology and Electrical Engineering: Empowering Technology for Better Future, ICITEE 2016*, Oct. 2017, pp. 1–6, doi: 10.1109/ICITEED.2016.7863303.

[25] D. L. Choi, B. W. Kim, Y. J. Lee, Y. Um, and M. Chung, "Design and creation of dysarthric speech database for development of QoLT software technology," in *2011 International Conference on Speech Database and Assessments, Oriental COCOSDA 2011 - Proceedings*, Oct. 2011, pp. 47–50, doi: 10.1109/ICSDA.2011.6085978.

[26] G. Schu, P. Janbakhshi, and I. Kodrasi, "On using the UA-speech and torgo databases to validate automatic dysarthric speech classification approaches," in *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, Jun. 2023, vol. 2023-June, pp. 1–5, doi: 10.1109/ICASSP49357.2023.10095981.

[27] A. B. Kain, J. P. Hosom, X. Niu, J. P. H. van Santen, M. Fried-Oken, and J. Staehely, "Improving the intelligibility of dysarthric speech," *Speech Communication*, vol. 49, no. 9, pp. 743–759, Sep. 2007, doi: 10.1016/j.specom.2007.05.001.

[28] W. Ziegler *et al.*, "Speech and nonspeech parameters in the clinical assessment of dysarthria: a dimensional analysis," *Brain Sciences*, vol. 13, no. 1, p. 113, Jan. 2023, doi: 10.3390/brainsci13010113.

[29] A. Benba, A. Jilbab, and A. Hammouch, "Voice analysis for detecting patients with Parkinson's disease using the hybridization of the best acoustic features," *International Journal on Electrical Engineering and Informatics (IJEEI)*, vol. 8, no. 1, pp. 108–116, Mar. 2016, doi: 10.15676/ijeei.2016.8.1.8.

[30] Kadi, Kamil Lahcene, et al. "Fully automated speaker identification and intelligibility assessment in dysarthria disease using auditory knowledge," *Biocybernetics and Biomedical Engineering*, vol. 36, no 1, p. 233-247, 2016

[31] T. B. Ijitona, J. J. Soraghan, A. Lowit, G. Di-Caterina, and H. Yue, "Automatic detection of speech disorder in dysarthria using extended speech feature extraction and neural networks classification," in *IET Conference Publications*, 2017, vol. 2017, no. CP731, pp. 15 (6 .)-15 (6 .), doi: 10.1049/cp.2017.0360.

[32] B. Vachhani, C. Bhat, B. Das, and S. K. Kopparapu, "Deep autoencoder based speech features for improved dysarthric speech recognition," in *Interspeech 2017*, Aug. 2017, vol. 2017-Augus, pp. 1854–1858, doi: 10.21437/Interspeech.2017-1318.

[33] A. Farhadipour, H. Veisi, M. Asgari, and M. A. Keyvanrad, "Dysarthric speaker identification with different degrees of dysarthria severity using deep belief networks," *ETRI Journal*, vol. 40, no. 5, pp. 643–652, Oct. 2018, doi: 10.4218/etrij.2017-0260.

[34] A. Misbullah, H. H. Lin, C. Y. Chang, H. W. Yeh, and K. C. Weng, "Improving acoustic models for dysarthric speech recognition using time delay neural networks," in *Proceedings of the International Conference on Electrical Engineering and Informatics*, Oct. 2020, vol. 2020-Octob, pp. 1–4, doi: 10.1109/ICELTICs50595.2020.9315506.

[35] M. Sidi Yakoub, S. ahmed Selouani, B. F. Zaidi, and A. Bouchair, "Improving dysarthric speech recognition using empirical mode decomposition and convolutional neural network," *Eurasip Journal on Audio, Speech, and Music Processing*, vol. 2020, no. 1, p. 1, Dec. 2020, doi: 10.1186/s13636-019-0169-5.

[36] F. Xiong, J. Barker, Z. Yue, and H. Christensen, "Source domain data selection for improved transfer learning targeting dysarthric speech recognition," in *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, May 2020, vol. 2020-May, pp. 7424–7428, doi: 10.1109/ICASSP40776.2020.9054694.

[37] G. Tartarisco *et al.*, "Artificial intelligence for dysarthria assessment in children with ataxia: a hierarchical approach," *IEEE Access*, vol. 9, pp. 166720–166735, 2021, doi: 10.1109/ACCESS.2021.3135078.

[38] S. R. Shahamiri, "Speech vision: an end-to-end deep learning-based dysarthric automatic speech recognition system," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 29, pp. 852–861, 2021, doi: 10.1109/TNSRE.2021.3076778.

[39] L. Wu, D. Zong, S. Sun, and J. Zhao, "A sequential contrastive learning framework for robust dysarthric speech recognition," in *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, Jun. 2021, vol. 2021-June, pp. 7303–7307, doi: 10.1109/ICASSP39728.2021.9415017.

[40] B. F. Zaidi, S. A. Selouani, M. Boudraa, and M. Sidi Yakoub, "Deep neural network architectures for dysarthric speech analysis and recognition," *Neural Computing and Applications*, vol. 33, no. 15, pp. 9089–9108, Aug. 2021, doi: 10.1007/s00521-020-05672-2.

[41] S. Chandrakala and N. Rajeswari, "Representation learning based speech assistive system for persons with dysarthria," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 25, no. 9, pp. 1510–1517, Sep. 2017, doi: 10.1109/TNSRE.2016.2638830.

[42] D. Wang, J. Yu, X. Wu, L. Sun, X. Liu, and H. Meng, "Improved end-to-end dysarthric speech recognition via meta-learning based model re-initialization," in *2021 12th International Symposium on Chinese Spoken Language Processing, ISCSLP 2021*, Jan. 2021, pp. 1–5, doi: 10.1109/ISCSLP49672.2021.9362068.

[43] S. Gupta *et al.*, "Residual neural network precisely quantifies dysarthria severity-level based on short-duration speech segments," *Neural Networks*, vol. 139, pp. 105–117, Jul. 2021, doi: 10.1016/j.neunet.2021.02.008.

[44] 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, doi: 10.1109/ACCESS.2021.3053335.

[45] C. Quan, K. Ren, Z. Luo, Z. Chen, and Y. Ling, "End-to-end deep learning approach for Parkinson's disease detection from speech signals," *Biocybernetics and Biomedical Engineering*, vol. 42, no. 2, pp. 556–574, Apr. 2022, doi: 10.1016/j.bbe.2022.04.002.

[46] M. Soleymanpour, M. T. Johnson, R. Soleymanpour, and J. Berry, "Synthesizing dysarthric speech using multi-speaker tts for dysarthric speech recognition," in *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, May 2022, vol. 2022-May, pp. 7382–7386, doi: 10.1109/ICASSP43922.2022.9746585.

[47] S. R. Mani Sekhar, G. Kashyap, A. Bhansali, A. Andrew Abishek, and K. Singh, "Dysarthric-speech detection using transfer learning with convolutional neural networks," *ICT Express*, vol. 8, no. 1, pp. 61–64, Mar. 2022, doi: 10.1016/j.icte.2021.07.004.

[48] J. Song *et al.*, "Detection and differentiation of ataxic and hypokinetic dysarthria in cerebellar ataxia and parkinsonian disorders via wave splitting and integrating neural networks," *PLoS ONE*, vol. 17, no. 6 June, p. e0268337, Jun. 2022, doi: 10.1371/journal.pone.0268337.

[49] A. A. Joshy and R. Rajan, "Automated dysarthria severity classification: a study on acoustic features and deep learning techniques," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 30, pp. 1147–1157, 2022, doi: 10.1109/TNSRE.2022.3169814.

[50] S. Rathod, M. Charola, and H. A. Patil, "Transfer learning using whisper for dysarthric automatic speech recognition," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 14338 LNAI, 2023, pp. 579–589.

[51] M. Chaiani, S. A. Selouani, M. Boudraa, and M. Sidi Yakoub, "Voice disorder classification using speech enhancement and deep learning models," *Biocybernetics and Biomedical Engineering*, vol. 42, no. 2, pp. 463–480, Apr. 2022, doi: 10.1016/j.bbe.2022.03.002.

[52] A. Almadhor, R. Irfan, J. Gao, N. Saleem, H. Tayyab Rauf, and S. Kadry, "E2E-DASR: end-to-end deep learning-based dysarthric automatic speech recognition," *Expert Systems with Applications*, vol. 222, p. 119797, Jul. 2023.

[53] P. Shanmugapriya and V. Mohan, "Comparative analysis of deep learning models for dysarthric speech detection," *Soft Computing*, vol. 28, no. 6, pp. 5683–5698, Mar. 2024, doi: 10.1007/s00500-023-09302-6.

[54] D. Yee *et al.*, "Latent phrase matching for dysarthric speech," in *INTERSPEECH 2023*, Aug. 2023, vol. 2023-Augus, pp. 161–165, doi: 10.21437/Interspeech.2023-1921.

[55] A. A. Joshy and R. Rajan, "Dysarthria severity assessment using squeeze-and-excitation networks," *Biomedical Signal Processing and Control*, vol. 82, p. 104606, Apr. 2023, doi: 10.1016/j.bspc.2023.104606.

[56] Z. Yue, E. Loweimi, and Z. Cvetkovic, "Dysarthric speech recognition, detection and classification using raw phase and magnitude spectra," *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, vol. 2023-Augus, pp. 1533–1537, 2023, doi: 10.21437/Interspeech.2023-222.

[57] B. Jolad and R. Khanai, "An approach for speech enhancement with dysarthric speech recognition using optimization based machine learning frameworks," *International Journal of Speech Technology*, vol. 26, no. 2, pp. 287–305, Jul. 2023, doi: 10.1007/s10772-023-10019-y.

[58] C. Lea *et al.*, "From user perceptions to technical improvement: enabling people who stutter to better use speech recognition," in *Conference on Human Factors in Computing Systems - Proceedings*, Apr. 2023, pp. 1–16, doi: 10.1145/3544548.3581224.

[59] D. Y. Isaev *et al.*, "Uncertainty of vowel predictions as a digital biomarker for ataxic dysarthria," *Cerebellum*, vol. 23, no. 2, pp. 459–470, Apr. 2024, doi: 10.1007/s12311-023-01539-z.

[60] A. Farhadipour and H. Veisi, "Gammatonegram representation for end-to-end dysarthric speech processing tasks: speech recognition, speaker identification, and intelligibility assessment," *Iran Journal of Computer Science*, Mar. 2024, doi: 10.1007/s42044-024-00175-y.

[61] V. Illner, T. Tykalova, D. Skrabal, J. Klempir, and J. Rusz, "Automated vowel articulation analysis in connected speech among progressive neurological diseases, dysarthria types, and dysarthria severities," *Journal of Speech, Language, and Hearing Research*, vol. 66, no. 8, pp. 2600–2621, Aug. 2023, doi: 10.1044/2023_JSLHR-22-00526.

[62] J. Harvill, D. Issa, M. Hasegawa-Johnson, and C. Yoo, "Synthesis of new words for improved dysarthric speech recognition on an expanded vocabulary," in *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, Jun. 2021, vol. 2021-June, pp. 6428–6432, doi: 10.1109/ICASSP39728.2021.9414869.

[63] M. Suresh, R. Rajan, and J. Thomas, "Dysarthria speech disorder classification using traditional and deep learning models," in *2023 2nd International Conference on Electrical, Electronics, Information and Communication Technologies, ICEEICT 2023*, Apr. 2023, pp. 01–06, doi: 10.1109/ICEEICT56924.2023.10157285.

[64] T. A. Mariya Celin, P. Vijayalakshmi, and T. Nagarajan, "Data augmentation techniques for transfer learning-based continuous dysarthric speech recognition," *Circuits, Systems, and Signal Processing*, vol. 42, no. 1, pp. 601–622, Jan. 2023, doi: 10.1007/s00034-022-02156-7.

[65] Q. Meeus, M. F. Moens, and H. Van hamme, "Bidirectional representations for low-resource spoken language understanding," *Applied Sciences (Switzerland)*, vol. 13, no. 20, p. 11291, Oct. 2023, doi: 10.3390/app132011291.

[66] A. M. Zaki, N. Khodadadi, W. H. Lim, and S. K. Towfek, "Predictive analytics and machine learning in direct marketing for anticipating bank term deposit subscriptions," *American Journal of Business and Operations Research*, vol. 11, no. 1, pp. 79–88, 2024, doi: 10.54216/AJBOR.110110.

[67] A. Pillai *et al.*, "Investigating generalizability of speech-based suicidal ideation detection using mobile phones," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 7, no. 4, pp. 1–38, Dec. 2023, doi: 10.1145/3631452.

[68] Z. Ou *et al.*, "Early identification of stroke through deep learning with multi-modal human speech and movement data," *Neural Regeneration Research*, vol. 20, no. 1, pp. 234–241, Jan. 2025, doi: 10.4103/1673-5374.393103.

BIOGRAPHIES OF AUTHORS

Prof. Aneeta S. Antony is a researcher and educationist who works affiliated to Visvesvaraya Technological University. Machine learning and speech/signal processing are her research foci. She's applied for two patents. She has several publications under her belt as a professional in data science and machine learning. In addition to neural networks, artificial intelligence, and machine learning, the trainer has over seven years of expertise instructing students in electronics and communication engineering. She is competent in deep learning, Keras, TensorFlow, curve fitting, regression, statistical analysis, and predictive analysis. She has completed a number of AI-based projects. She can be contacted at email: aneetaantony123@gmail.com.

Prof. Dr. Rohini Nagapadma is the esteemed Principal of The National Institute of Engineering in Mysuru, Karnataka, India, where she leads with a passion for excellence. With over 30 years of teaching experience, she has dedicated her career to the field of Electronics and Communication Engineering, enriching the lives of countless students along the way. During her tenure as the Head of the Department of Electronics and Communication Engineering, spanning over 15 years, she honed her skills and expertise, which she now generously shares as a Ph.D. thesis evaluator and examiner. Her journey has been marked by a commitment to continuous learning and growth. She has contributed significantly to the academic community through her reviews for IEEE transactions and various international journals and conferences, broadening her perspectives with each paper. A dynamic individual, she actively engages in delivering talks at workshops, conferences, and Faculty Development Programs, constantly evolving and enhancing her delivery with each session. As the branch counselor of the NIE IEEE student branch and a member of IEEE and IETE Professional society, she remains at the forefront of technological advancements. Her contributions extend beyond academics, as she actively participates in various committees such as the local inspection committee (LIC), VTU, Belagavi, and the board of studies (BoS) and board of examination (BoE) of reputed institutions, where her strong leadership and fairness are highly valued. With numerous publications in renowned journals and a passion for her work that shines through in all her endeavors, she is a true asset to the academic community. She can be contacted at email: rohininagapadma@nie.ac.in.

Prof. Dr. Ajish K. Abraham is a distinguished Professor specializing in Electronics and Acoustics. His research endeavors encompass a wide array of projects, notably the development of a robot for articulatory simulation, an automatic speech processing system for assessing stuttering, and the automation of the Malayalam articulation test for evaluating articulation disorders. With a profound focus on developing automated diagnostic systems for assessing the severity of speech disorders, he is actively engaged in creating assistive technology devices for rehabilitating individuals with communication disorders. He earned his Ph.D. in Electronics from the University of Mysore and pursued Postdoctoral Training at the University of California, San Diego, USA. With a rich academic background, he brings over 32 years of teaching and research experience to his current role. His professional journey includes 10 years at the Central Scientific Instruments Organization, Chandigarh, a constituent Laboratory of the Council of Scientific and Industrial Research (CSIR), Government of India, followed by 22 years at the All-India Institute of Speech and Hearing (AIISH), Mysore. His areas of interest span the design and development of biomedical equipment, audio and speech processing, research in hearing aid technology, and acoustic noise, among others. He can be contacted at email: ajish68@aiishmysore.in.