International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10
Article Received: *10 August 2023* ***Revised:*** *29 September 2023* ***Accepted:*** *15 October 2023*

Survey of Automatic Dysarthric Speech Recognition

Namita Kure, S. B. Dhonde
[1]Research Scholar, AISSM IOT, Pune, npachling@gmail.com
[2]Professor, AISSMS College of Engineering,Pune, India, dhondesomnath@mail.com

Abstract—The need for automated speech recognition has expanded as a result of significant industrial expansion for a variety of automation and human-machine interface applications. The speech impairment brought on by communication disorders, neurogenic speech disorders, or psychological speech disorders limits the performance of different artificial intelligence-based systems. The dysarthric condition is a neurogenic speech disease that restricts the capacity of the human voice to articulate. This article presents a comprehensive survey of the recent advances in the automatic Dysarthric Speech Recognition (DSR) using machine learning and deep learning paradigms. It focuses on the methodology, database, evaluation metrics and major findings from the study of previous approaches. From the literature survey it provides the gaps between exiting work and previous work on DSR and provides the future direction for improvement of DSR.

Keywords-Dysarthric Speech Recognition, Speech Intelligibility, Voice Pathology, Speech Recognition,

I. INTRODUCTION

Dysarthria is a speech disorder generated due to weakness in speed production muscle or when an individual is unable to control them. It frequently causes slow or slurred speech which is difficult to understand. Dysarthria can be caused due to neural disorder, troat or tongue muscle weakness, or facial paralysis [1][2]. The muscle used for speed production is controlled by the nervous system and brain. Mostly dysarthria is caused due to damage to these muscles. Dysarthria is grouped into developmental and acquired dysarthria. The developmental dysarthria normally found in children is occurred due to brain damage during or before birth. The acquired dysarthria generally occurred due to brain damage in adulthood or later in life such as brain tumors, stroke, head injury, motor neuron disease, or Parkinson's disease [3][4][5].

The term "dysarthria" refers to a variety of neurological speech abnormalities caused by injury to the central or peripheral nerve systems. Reduced stress, sluggish speech pace, hyper-nasality, muscular stiffness, spasticity, monopitch, and a limited range of speech motions are all signs of dysarthric speech. It can impact the subglottal, laryngeal, and articulatory systems, which can make speech production difficult. Stroke, Parkinson's disease, and cerebral palsys are the most common roots of motor speech difficulties. According to reports, improving human-machine interaction for persons with dysarthria is becoming increasingly important in order to boost overall wellness and independence. Physical impairments are common in people with dysarthria, making common input methods (typing, touch screen, etc.) difficult to use [6][7].

Traditionally, the language or speech therapist diagnosed dysarthria disorder by asking people to read passages loudly, recite numbers or weekdays, make various sounds or talk about any familiar topic. The traditional techniques performance is limited due to various factors such as inadequate knowledge of experts, tiredness, fatigue, etc. Dysarthria may affect phonation, breathing, prosody, articulation, resonance, and lip movement. It shows a larger variation in speech intelligibility. The scope of intelligibility is huge and may depend upon the extent of nervous system damage. The typical symptoms of the dysarthria are enlisted in Fig. 1. Because of articulatory difficulties, there is no uniformity in articulation. Pronunciation changes and speaking pace slows as a result of exhaustion. All of these distinctiveness impair the dysarthric speaker's intelligibility (the degree to which others can understand their speech) and limit verbal interactions, reducing their quality of life [10][11].

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10
Article Received: *10 August 2023* ***Revised:*** *29 September 2023* ***Accepted:*** *15 October 2023*

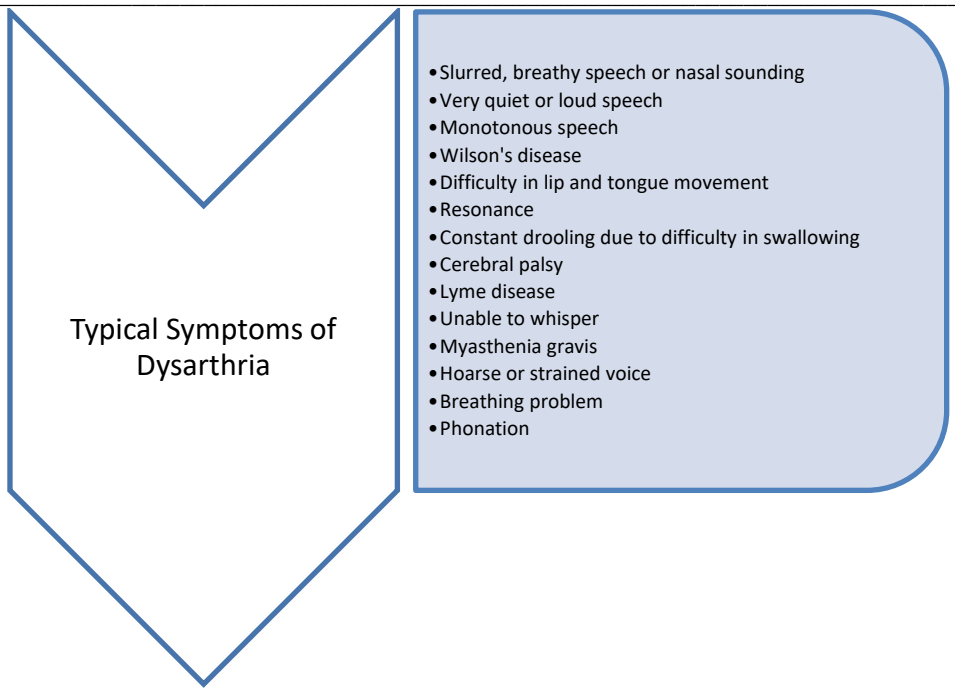

Figure 1. Typical symptoms of dysarthria

The classification system helps to narrow down the dimension of perceptual analysis of dysarthric speech. The classification of dysarthric speech is given in Fig. 2. Most clinicians find this useful to correct or reduce the deficit found in dysarthric speech production. Normal speakers typically communicate at rates between 150 to 200 words per minute. The speech is clear, timely, and contextually relevant. Speakers with severe impairments communicate at a rate of fewer than 15 words per minute; This reduction in the rate of communications has implications in the quantity and the quality. People suffering from dysarthria are generally physically challenged. It is difficult for them to handle the conventional keyboard or mouse interfaces. Dysarthric speakers experience difficulty to contribute enough samples of speech data. Some dysarthric speakers get tired soon which may lead to distress. They often fall short to utter certain sounds, which results in phonetic variation [8][9].

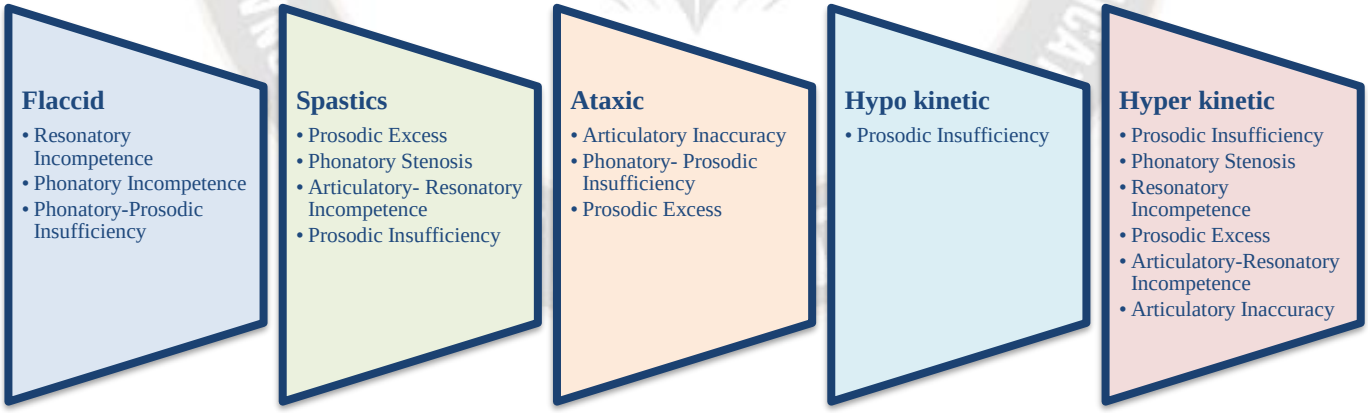

Figure 2. Types and features of dysarthria

This paper presents a comprehensive survey of distinct ML-based and DL-based DSR systems. It focuses on the DSR methodology that comprises enhancement, data augmentation, feature extraction, feature selection, and classification techniques. It analyses the dataset, experimental results and performance metrics to depict the merits, demerits and challenges of the present DSR systems.

The rest of paper is structured as follow: Section 2 depicts the generalized process of the automatic DSR. Section 3 gives the succinct survey of recent ML and DL based SER systems. Finally, section 4 concludes the paper and paves the way for future enhancement through future scope.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10
Article Received: *10 August 2023* ***Revised:*** *29 September 2023* ***Accepted:*** *15 October 2023*

II. GENERALIZED PROCESS OF DSR

The generalized process of DSR is shown in Fig. that encompasses the pre-processing, feature representation, classification and DSR.

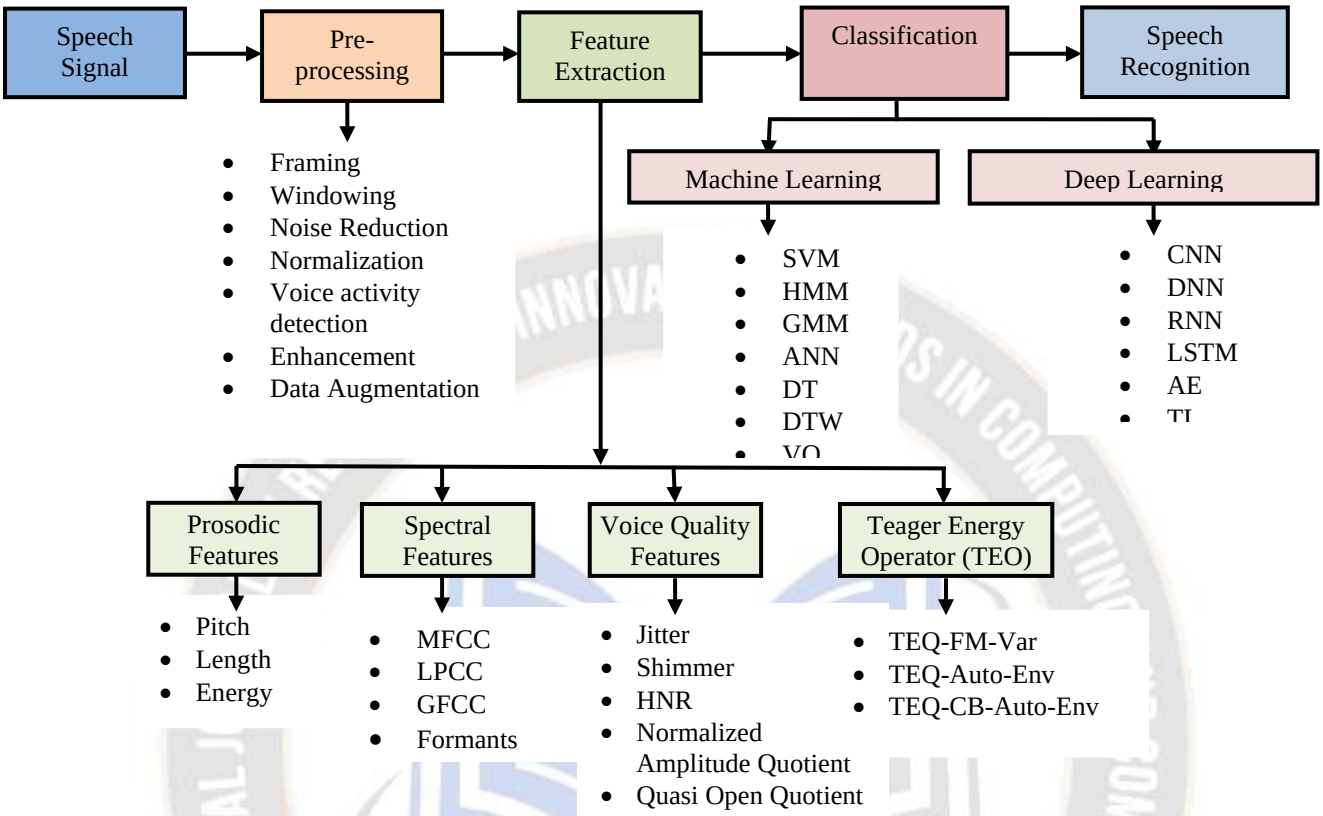

Figure 3. Generalized Process of DSR

The preprocessing phase deals with the primary processing on the dysarthric speech to improve the quality of features and performance of the classifiers. It encompasses framing, cropping, speech separation, noise suppression, windowing, normalization, speech enhancement, data augmentation, etc. The dysarthric speech contains different types of the reverberations, silent regions, stops, wide variety in pitch and energy of the signal which tends to use speech enhancement to enhance DSR effectiveness. The feature extraction is important phase to collect the distinctive and unique characteristics of the normal and dysarthric speech. The features are generally grouped into spectral, prosodic, voice quality, and Teager-Energy Operator features. Traditional machine leaning (ML) based DSR includes feature extraction followed by classification whereas in deep learning (DL) the feature extraction may not be used as DL techniques often refers to combination of hidden feature extraction layers and classification layer. However, many hybrid DL algorithms uses the traditional features as the input to boost the speech intelligibility, feature representation, and DSR accuracy.

III. SURVEY OF DSR

A. ML-based DSR

Various DSR strategies have been presented in last two decades; this section gives a quick overview of recent DSR approaches. Voice tremor has been quantified using phonation parameters that define disordered voice, such as jitter and fundamental frequency [11,12]. To avoid the gender and acoustic environment dependence of these parameters, a Pitch Period Entropy-based evaluation was developed [13]. Hypophonia has also been described using fluctuation of energy and short-time energy [14]. The Teager-Kaiser Energy Operator which provides the speech intensity measure is utilized to adjust for signal frequency [15]. To explore the influence on articulatory dynamics and speech intelligibility, acoustic cues based on the first three formants and their respective bandwidths can be studied [16].Vowel Space Area (VSA) has been investigated for assessing speech intelligibility [17]. A Support Vector Machine (SVM) classifier was used to investigate a method for distinguishing dysarthric speech from healthy speech using a collection of glottal and openSMILE characteristics [18]. In [19],the authors investigated

analytic phase characteristics generated from voice signals using the single frequency filtering approach. In [20], audio descriptor information used for determining musical instrument timbre were combined with an Artificial Neural Network (ANN) model to classify dysarthric speech severity levels. For dysarthria classification, multi-tapered spectral estimation was used to extract audio descriptor features.

Hasegawa-Johnson et al. [21] evaluate recognition performance for dysarthric speech compared with automatic speech recognition (ASR) systems based on Gaussian mixture model–hidden Markov models (GMM–HMMs) and SVMs [22]. The experimental results showed that the HMM-based model may provide robustness against large-scale word-length variances; meanwhile, the SVM-based model can alleviate the effect of deletion of or reduction in consonants. Rudzicz et al. [23, 24] investigated acoustic models of GMM–HMM, conditional random field, SVM, and artificial neural networks (ANNs) [24], and the results showed that the ANNs provided higher accuracy than other models.Revathi et al. [25] presented multiple such as Gamma Tone Energy (GFE), modified group delay cepstrum (MGDFC), and stock well features for isolated DSR. It used decision level fusion with the help of vector quantization (VQ) classifier. It used speech enhancement scheme to minimize the distortions and improve the speech intelligibility. It resulted in WER of 4% for the dysarthric subjects with 6% intelligibility. Al-Qatab et al. [26] used four types of features such as Spectral, Cepstral, Voice Quality, Prosodic and Overall Speech features along with SVM, ANN, Linear Discriminant Analysis (LDA), Classification and Regression Tree (CART), Naive Bayes (NB), and Random Forest (RF) classifier for DSR. Seven feature selection algorithms have been presented for the feature selection to select the dominant features such as Conditional Information Feature Extraction (CIFE), Double Input Symmetrical Relevance (DISR), Interaction Capping (ICAP), Conditional Mutual Information Maximization (CMIM), Conditional Redundancy (Condred), Joint Mutual Information (JMI), and Relief. It provided Average Ranking Score of 4.88 for Random Forest and Relief Feature Selection. Janbakhshi et al. [27] presented singular value decomposition (SVD) for the spectro-temporal representation of the dysarthric speech and Temporal Grassmann discriminant analysis (T-GDA) for the DSR. It outperformed the traditional MFCC-SVM based DSR. The subspace based learning shows superior discrimination between normal and dysarthric speech. The temporal subspace gives enhanced performance compared with spectral subspace.

B. DL-based DSR

Recently, deep learning technology has been widely used in many voiced based automation systems and has proven it can provide better performance than conventional ML based methods [28][29]. Fathima et al. [30] applied a multilingual Time Delay Neural Network (TDNN) system that combined acoustic modeling and language specific information to increase ASR performance. The experimental results showed that the TDNN-based ASR system achieved suitable performance, as the word error rate was 16.07% in this study. Yue et al. [31] investigated convolutional and light gated recurrent unit (LiGRU) based multi-spectra acoustic model for DSR. It used data augmentation to minimize the data scarcity problem using speed perturbation which has given 11% and 40.6% WER for normal and dysarthric speech. Futher, Yue et al. [32] developed multi-stream acoustic model based on Convolutional neural Network (CNN), LiGRU, and fully connected Multi Layer Perceptron (MLP) and optimal fusion technique for DSR. The proposed model provided a WER of 4.6% for the pre-processed data using electromagnetic articulography (EMA). The EMA preprocessing includes Butterworth filter for measurement noise minimization and down-sampling for synchronization of MFCC features.

The data efficiency is major obstacle in the DSR. Soleymanpour et al. [33] proposed text to speech (TTS) synthesizer for the data augmentation based on FastSpeech model. The augmented data provided to Deep Neural Network-HMM (DNN-HMM) with light bidirectional GRU that has given a WER improvement of 12.2% over the baseline model. Traditional data augmentation approaches majorly focuses on the temporal variations of the signal however spectral envelope remains same. Liu et al. [34] presented vocal tract length perturbation (VTLP), tempo perturbation and speed perturbation for the data augmentation that concentrates on temporal as well as spectral transformations of the dysarthric speech signal. The DNN and Neural architecture search (NAS) based DSR provides WER of 25.21 % and 5.4% for UASpeech and CUHK dataset respectively. Shahamiri [35] used voicegram to provide the correlation between phonemes and the dysarthric speech. The visual data augmentation model is used for the data augmentation to minimize data scarcity problem in DSR. The Spatial–CNN (S-CNN) provides an accuracy of 67% on UASpeech dataset. The proposed S-CNN some time causes vanishing gradient problem and provides poor results for the moderate dysarthria. The intelligibility of the speech is hugely affected due to time domain variance of dysarthric speech and background noise. Lin et al. [36] suggested that the deep learning based voice conversion (DVC) using phonetic posteriorgram (PPG) provides stable performance compared with DVC-Mel under noisy condition.

Khodrasi et al. [37] suggested that spectro-temporal sparsity using the Gini index provided better performance than shimmer, jitter, fundamental frequency, harmonics to noise ratio (HNR), and MFCC for the DSR. It is observed that spectral sparsity has proven better performance than temporal sparsity. Further, Khodrasi et al. [38] used CNN for learning the temporal spectral characteristics obtained using temporal envelope and fine structure (TEFS).The TEFS outperformed the traditional SIFT based speech signal spectrogram. The TEFS-CNN provides 85.72% accuracy for DSR whereas SIFT-CNN provides 69.76% accuracy for DSR. Chandrashekhar et al. [39] investigated the time–frequency CNN for capturing the temporal as well as spectral properties of the dysarthric speech. The spectro-temporal properties of the speech signals are obtained using Short-time Fourier Transform (SIFT), Spectrograms Using Single Frequency Filtering (SFF), and Constant Q-Transform (CQT). The DSR performance has shown higher accuracy for the female subjects compared with the male subject. The training data deficiency resulted in class imbalance problem. The time-frequency based CNN provides better spectro-temporal variation of the dysarthric speech which has shown significant improvement in DSR accuracy over the traditional ANNs [40]. Fritsch and Doss [41] presented Recurrent Neural Network (RNN) based binary and CNN based multi-feature classifier. It provided high correlation for synthesized speech generated using Text to Speech (TTS).

Table 1: provides the summary of various DSR techniques based on ML and DL approaches.

Table 1: Summary of ML and DL based DSR

Author and Year	**Speech Enhancement**	**Data Augmentation**	**Feature extraction**	**Classifier**	**Database**	**Performance metrics**	**Remark**
Yue, et al. (2022a)	Cepstral Processing to separate filter and speech element	Speed perturbation	CNN-LiGRU	Softmax	TORGO	WER- 40.6% (dysarthric), 11% (Normal)	Combination of excitation and vocal tract component can be used for speaking stylemodelling
Yue, et al. (2022b)	EMA	-	CNN-LiGRU-FCMLP	softmax	TORGO	WER-4.6%	Over-fitting problem for high level articulatory feature fusion
Soleymanpour et al. (2022)	-	TTS	DNN-HMM-BLiGRU	softmax	TORGO	WER-41.6%	The severity of dysarthric speech depeds upon energy, duration and pitch of the signal.
Liu et al. (2021)	-	VTLP, tempo perturbation and speed perturbation	Model based speaker adaptation and cross-domain generation of visual features	DNN-NAS	UASpeech and Chinese University of Hong Kong (CUHK)	-WER =25.21% (UASpeech) - WER=5.4% (CUHK)	- High WER for low intelligibility speaker

Shahamiri (2021)	-	Visual Data augmentation	Voicegram	Spatial Convolutional Neural Network (S-CNN)	UASpeech	Accuracy=67%	-Provides less temporal representation of speech -May cause vanishing gradient problem
Lin et al. (2021)	-	-	-	Convolution neural network (CNN) with a phonetic posteriorgram (PPG)	10 samples of 19 Chinese commands for 3 user	CNN–PPG-93.49 %, CNN-MFFC-65.67%, ASR based System-89.59%	- Class imbalance problem issue due to uneven dataset size
Kodrasi et al. (2020)	-	-	Spectro-temporal sparsity using the Gini index	SVM	Spanish database (PC-GITA database)	Accuracy=83.30 % (GST), 76.7% (MFCC), 60% (HNR), 57% (Shimmer) ,52% (Jitter), 54.40 % (Fo)	Less recognition rate due to less number of features - Not suitable for larger dataset
Kodrasi et al. (2021)	-	-	Temporal envelope and fine structure (TEFS)	CNN	PC-GITA database	Accuracy =85.75, AUC=0.93	-Less feature discrimination due to higher intra-class and lower interclass variability -Can not handle complex auditory models
Chandrashekar et al. (2020)	-	-	SIFT, Spectrograms Using Single Frequency Filtering (SFF), Constant Q Transform (CQT)	Time-Frequency CNN	Universal Access and TORGO	Accuracy-98.00% (Female), 95.80% (Male)	-Class imbalance problem -complexity of network -High computation time
Al-Qatab and Mustafa (2021)	-	-	Spectral, Cepstral, Voice Quality,	LDA, CART, NB, ANN, SVM, and RF	NEMOURS database	Average Ranking Score for Random Forest and Relief	-Ability to classify speech based on severity

			Prosodic, Overall Speech features,			Feature Selection (4.88)	level - Feature selection is important for DSR - Not applicable for larger dataset -less performance than deep learning approaches
Janbakhshi et al. (2021)	-	-	SVD	T-GDA	PC-GITA, MoSpeeDi, UA-speech	Accuracy-82.0±3.5 % (PC-GITA,), 80.5±4.7 % (MoSpeeDi), 96.30% (UA)	- Temporal subspaces provides better representation of normal and dysarthric speech compared with spectral subspaces
Fritsch and Doss (2021)	-	-	Pearson's correlation coefficient and Spearman's correlation coefficient	RNN	UA-Speech database	PCC (0.950), SCC (0.957)	-Provides high correlation for synthesized speech generated using Text to Speech (TTS)

IV. CONCLUSION AND FUTURE SCOPE

Thus this article presents the DSR based on various ML and DL approaches that covers the methodology, database, evaluation metrics, advantages, disadvantages, and finding from the study. It is observed that the deep learning techniques outperformed the traditional machine learning techniques because of its superior feature representation. The DL approaches are less dependent on the hand crafted features unlike traditional ML based approaches. Database generation is challenging task because of unavailability of the proper resources and proper ground truth. The DSR is very challenging ddue to variability in the speech intelligibility because of various attributes such as language, age, gender, region, noise, etc.

REFERENCES

[1] Watanabe, Shinji, Marc Delcroix, Florian Metze, and John R. Hershey. "New Era for Robust Speech Recognition." *Springer International Publishing. doi* 10 (2017): 978-3.

[2] Bhangale, KishorBarasu, and K. Mohanaprasad. "A review on speech processing using machine learning paradigm." *International Journal of Speech Technology* 24 (2021): 367-388.

[3] Vadwala, Ayushi Y., Krina A. Suthar, Yesha A. Karmakar, Nirali Pandya, and Bhanubhai Patel. "Survey paper on different speech recognition algorithm: challenges and techniques." *Int J computappl* 175, no. 1 (2017): 31-36.

[4] Sonawane, Anagha, M. U. Inamdar, and Kishor B. Bhangale. "Sound based human emotion recognition using MFCC & multiple SVM." In *2017 international conference on information, communication, instrumentation and control (ICICIC)*, pp. 1-4. IEEE, 2017.

[5] Bhangale, KishorBarasu, and MohanaprasadKothandaraman. "Survey of deep learning paradigms for speech processing." *Wireless Personal Communications* 125, no. 2 (2022): 1913-1949.

[6] Pennington, Lindsay, Naomi K. Parker, Helen Kelly, and Nick Miller. "Speech therapy for children with dysarthria acquired before three years of age." *Cochrane Database of Systematic Reviews* 7 (2016).

[7] Jamal, Norezmi, ShahnoorShanta, Farhanahani Mahmud, and M. N. A. H. Sha'abani. "Automatic speech recognition (ASR) based approach for speech therapy of aphasic patients: A review." In *AIP Conference Proceedings*, vol. 1883, no. 1, p. 020028. AIP Publishing LLC, 2017.

[8] Vachhani, Bhavik, ChitralekhaBhat, and Sunil Kumar Kopparapu. "Data Augmentation Using Healthy Speech for Dysarthric Speech Recognition." In *Interspeech*, pp. 471-475. 2018.

[9] Takashima, Ryoichi, Tetsuya Takiguchi, and YasuoAriki. "Two-step acoustic model adaptation for dysarthric speech recognition." In *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 6104-6108. IEEE, 2020.

[10] Kim, M. J., Cao, B., An, K., & Wang, J. (2018). "DSR Using Convolutional LSTM Neural Network." In *INTERSPEECH* pp. 2948-2952.

[11] Vasilakis, Miltiadis, and YannisStylianou. "Voice pathology detection based eon short-term jitter estimations in running speech." *Folia PhoniatricaetLogopaedica* 61, no. 3 (2009): 153-170.

[12] Skodda, Sabine, WenkeVisser, and Uwe Schlegel. "Short-and long-term dopaminergic effects on dysarthria in early Parkinson's disease." *Journal of Neural Transmission* 117 (2010): 197-205.

[13] Little, Max, Patrick McSharry, Eric Hunter, Jennifer Spielman, and Lorraine Ramig. "Suitability of dysphonia measurements for telemonitoring of Parkinson's disease." *Nature Precedings* (2008): 1-1.

[14] Dimitriadis, D., Potamianos, A., & Maragos, P. Shao, Jun, Julia K. MacCallum, Yu Zhang, Alicia Sprecher, and Jack J. Jiang. "Acoustic analysis of the tremulous voice: assessing the utility of the correlation dimension and perturbation parameters." *Journal of communication disorders* 43, no. 1 (2010): 35-44.

[15] Dimitriadis, Dimitrios, Alexandros Potamianos, and Petros Maragos. "A comparison of the squared energy and Teager-Kaiser operators for short-term energy estimation in additive noise." *IEEE Transactions on signal processing* 57, no. 7 (2009): 2569-2581.

[16] Allison, Kristen M., Lucas Annear, Marisa Policicchio, and Katherine C. Hustad. "Range and precision of formant movement in pediatric dysarthria." *Journal of Speech, Language, and Hearing Research* 60, no. 7 (2017): 1864-1876.

[17] Lansford, Kaitlin L., and Julie M. Liss. "Vowel acoustics in dysarthria: Speech disorder diagnosis and classification." (2014).

[18] Narendra, N. P., and PaavoAlku. "Glottal source information for pathological voice detection." *IEEE Access* 8 (2020): 67745-67755.

[19] Gurugubelli, Krishna, and Anil Kumar Vuppala. "Analytic phase features for dysarthric speech detection and intelligibility assessment." *Speech Communication* 121 (2020): 1-15.

[20] Bhat, Chitralekha, BhavikVachhani, and Sunil Kumar Kopparapu. "Automatic assessment of dysarthria severity level using audio descriptors." In *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 5070-5074. IEEE, 2017.

[21] Hasegawa-Johnson, Mark, Jonathan Gunderson, Adrienne Perlman, and Thomas Huang. "HMM-based and SVM-based recognition of the speech of talkers with spastic dysarthria." In *2006 IEEE International Conference on Acoustics Speech and Signal Processing Proceedings*, vol. 3, pp. III-III. IEEE, 2006.

[22] Cortes, Corinna, and Vladimir Vapnik. "Support-vector networks." *Machine learning* 20 (1995): 273-297.

[23] Rudzicz, Frank. "Phonological features in discriminative classification of dysarthric speech." In *2009 IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 4605-4608. IEEE, 2009.

[24] Rudzicz, Frank. "Articulatory knowledge in the recognition of dysarthric speech." *IEEE Transactions on Audio, Speech, and Language Processing* 19, no. 4 (2010): 947-960.

[25] Revathi, Arunachalam, R. Nagakrishnan, and N. Sasikaladevi. "Comparative analysis of DSR : multiple features and robust templates." *Multimedia Tools and Applications* (2022): 1-15.

[26] Al-Qatab, Bassam Ali, and Mumtaz Begum Mustafa. "Classification of dysarthric speech according to the severity of impairment: an analysis of acoustic features." *IEEE Access* 9 (2021): 18183-18194.

[27] Janbakhshi, Parvaneh, Ina Kodrasi, and HervéBourlard. "Subspace-based learning for automatic dysarthric speech detection." *IEEE Signal Processing Letters* 28 (2020): 96-100.

[28] Bhangale, Kishor B., Prashant Titare, RaosahebPawar, and SagarBhavsar. "Synthetic speech spoofing detection using MFCC and radial basis function SVM." *IOSR J. Eng.(IOSRJEN)* 8, no. 6 (2018): 55-62.

[29] Bhangale, Kishor, and K. Mohanaprasad. "Speech Emotion Recognition Using Mel Frequency Log Spectrogram and Deep Convolutional Neural Network." In *Futuristic Communication and Network Technologies*, pp. 241-250. Springer, Singapore, 2022.

[30] Fathima, Noor, Tanvina Patel, C. Mahima, and AnuroopIyengar. "TDNN-based Multilingual Speech Recognition System for Low Resource Indian Languages." In *Interspeech*, pp. 3197-3201. 2018.

[31] Yue, Zhengjun, ErfanLoweimi, and Zoran Cvetkovic. "Raw Source and Filter Modelling for DSR ." In *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 7377-7381. IEEE, 2022.

[32] Yue, Zhengjun, ErfanLoweimi, Zoran Cvetkovic, Heidi Christensen, and Jon Barker. "Multi-modal acoustic-articulatory feature fusion for dysarthric speech recognition." In *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 7372-7376. IEEE, 2022.

[33] Soleymanpour, Mohammad, Michael T. Johnson, Rahim Soleymanpour, and Jeffrey Berry. "Synthesizing Dysarthric Speech Using Multi-Speaker Tts For Dysarthric Speech Recognition." In *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 7382-7386. IEEE, 2022

[34] Liu, Shansong, Shoukang Hu, XurongXie, MengzheGeng, Mingyu Cui, Jianwei Yu, Xunying Liu, and Helen M. Meng. "Recent Progress in the CUHK DSR System." *IEEE/ACM Transactions on Audio, Speech, and Language Processing* (2021).

[35] Shahamiri, Seyed Reza. "Speech vision: An end-to-end deep learning-based dysarthric automatic speech recognition system." *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 29 (2021): 852-861.

[36] Lin, Yu-Yi, Wei-Zhong Zheng, Wei Chung Chu, Ji-Yan Han, Ying-Hsiu Hung, Guan-Min Ho, Chia-Yuan Chang, and Ying-Hui Lai. "A speech command control-based recognition system for dysarthric patients based on deep learning technology." *Applied Sciences* 11, no. 6 (2021): 2477.

[37] Kodrasi, Ina, and HervéBourlard. "Spectro-temporal sparsity characterization for dysarthric speech detection." *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 28 (2020): 1210-1222.

[38] Kodrasi, Ina. "Temporal envelope and fine structure cues for dysarthric speech detection using CNNs." *IEEE Signal Processing Letters* 28 (2021): 1853-1857.

[39] Chandrashekar, H. M., VeenaKarjigi, and N. Sreedevi. "Investigation of different time-frequency representations for intelligibility assessment of dysarthric speech." *Ieee transactions on neural systems and rehabilitation engineering* 28, no. 12 (2020): 2880-2889.

[40] Chandrashekar, H. M., VeenaKarjigi, and N. Sreedevi. "Spectro-temporal representation of speech for intelligibility assessment of dysarthria." *IEEE Journal of Selected Topics in Signal Processing* 14, no. 2 (2019): 390-399.

[41] Fritsch, Julian, and Mathew Magimai-Doss. "Utterance verification-based dysarthric speech intelligibility assessment using phonetic posterior features." *Ieee signal processing letters* 28 (2021): 224-228.