International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9s
DOI: https://doi.org/10.17762/ijritcc.v11i9s.7392
***Article Received**: 01 May 2023 **Revised:** 25 June 2023 **Accepted:** 20 July 2023*

A Comprehensive Survey of Automatic Dysarthric Speech Recognition

Chandarani Pophale[1], Dr.Shankar Chavan[2]
[1]Research Scholar ,Department of Electronics &Telecommunication,
Dr. D. Y. Patil Institute of Technology
Pimpri, Pune-18,Maharashtra,India.
Chanda.pophale@gmail.com
[2]Associate Professor , Department of Electronics &Telecommunication,
Dr. D. Y. Patil Institute of Technology
Pimpri, Pune-18,Maharashtra,India.
shankar.chavan@dypvp.edu.in

Abstract—Automatic dysarthric speech recognition (DSR) is very crucial for many human computer interaction systems that enables the human to interact with machine in natural way. The objective of this paper is to analyze the literature survey of various Machine learning (ML) and deep learning (DL) based dysarthric speech recognition systems (DSR). 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.The proposed survey presents the various challenges related with DSR such as individual variability, limited training data, contextual understanding, articulation variability, vocal quality changes, and speaking rate variations.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, Machine Learning, Deep Learning, Speech Intelligibility, Voice Pathology, Speech Recognition.

I. INTRODUCTION

Dysarthria is common disorder occurred due to neural disorder, tongue or throat muscle weakness, or facial paralysis. The dysarthria is grouped into developmental which caused due to brain damage before birth and acquired dysarthria that caused due to stroke, brain injury, brain tumor, Parkinson's disease or motor neuron disease, etc. Normal speaker can utter 150 to 200 words per minute but speakers with voice impairment can utter 15-25 words per minute. Dysarthria may affect phonation, breathing, prosody, articulation, resonance, and lip movement [1][2][3]. It shows a larger variation in speech intelligibility. The dysarthria causes mono pitch, muscular stiffness, sluggish speech, too fast or too slow speech, longer pauses in voice that limits the understanding of the content and context of speech [4][5]. The different types of the dysarthric disorders depending upon the perceptual analysis of the voice are illustrated in Fig. 1.

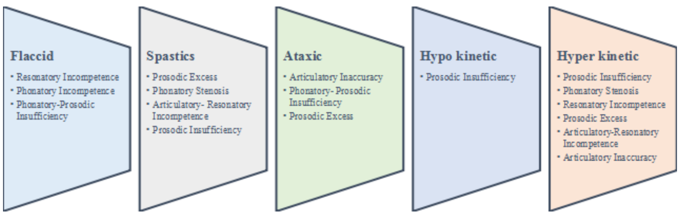

Fig. 1 Types of Dysarthria

The manual dysarthria recognition is performed by speech therapist by asking patient to recite number, read the passage loudly, make different sounds or talking something on familiar topic. However, the effectiveness of the traditional techniques is limited because of tiredness, fatigue, inadequate knowledge of expert, and larger speech as well as speaker variability [6][7].

In recent years various ML and DL techniques have been presented for the DSR to assist human computer interaction systems and affective computing systems. The machine learning based schemes used feature extraction methods to characterize the unique attributes of the voice signal and classification algorithms are utilized to distinguish the normal and dysarthric voice. The traditional ML based DSR schemes provides limited performance because of low feature representation capability, unable to work for larger database, and need of extensive re-processing strategies to minimize the noise, separate speech, minimization of reverberations, etc. The DL algorithms have capability to learn and classify the dysarthric voice using single model and provides better feature representation compared with traditional ML based DSR systems. However, the DL based DSR schemes are suffering from complex deep learning architectures, higher trainable parameters, data scarcity issue due to unavailability of data, tedious hyper-parameter tuning of deep learning algorithms,

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9s
DOI: https://doi.org/10.17762/ijritcc.v11i9s.7392
***Article Received**: 01 May 2023 **Revised:** 25 June 2023 **Accepted:** 20 July 2023*

and larger training as well as detection time. The Ml and DL based DSR schemes has shown less generalization capability due to larger variability in speech, age, gender, language, speaker, cultural and regional issues [8][9][10].

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 data set, 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 DSR methodology that described the generalized process of the automatic DSR. Section 3 gives the succinct analysis of results and discussions of ML and DL based SER systems. Finally, section 4 concludes the paper and paves the way for future enhancement through future scope.

II. METHODOLOGY

Generalized Process of DSR

The generalized process of DSR is shown in Fig. 2 that encompasses the pre-processing, feature representation, classification and DSR.

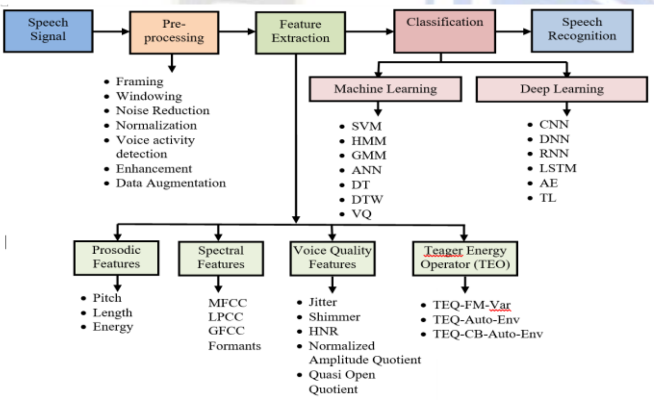

Fig. 2. Generalized Process of DSR

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, prosody, 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 performance of DSR systems [9][10].

III. DISCUSSION

Various ML and DL based techniques have been utilized for the DSR in last decade. This section focuses on the methodology, data-set, findings, merits and challenges of the existing researches on the DSR.

3.1 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. To avoid the gender and acoustic environment dependence of these parameters, a Pitch Period Entropy-based evaluation was developed. Hypophonia has also been described using fluctuation of energy and short-time energy. The Teager-Kaiser Energy Operator which provides the speech intensity measure is utilized to adjust for signal frequency. Acoustic cues based on the first three formants and their respective bandwidths has shown the influence on articulatory dynamics and speech intelligibility [11][12]. In [13], audio descriptor information used for determining musical instrument timbre were combined with an Artificial Neural Network (ANN) model to classify dysarthric speech severity levels. Speech modelling analysis of word length variance is very important in DSR to characterize the dysarthric voice. Hasegawa-Johnson et al. [14] 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. 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. [15] investigated acoustic models of GMM–HMM, conditional random field, SVM, and artificial neural networks (ANNs), and the results showed that the ANNs provided higher accuracy than other models. Revathi et al. [16] 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. [17] 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. [18] 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.

3.2 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 [19]. Fathima et al. [20] 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. [21] 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. Further, Yue et al. [22] 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. [23] proposed text to speech (TTS) synthesizer for the data augmentation based on Fast Speech 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. [24] 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 [25] 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. [26] 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. [27] 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. [28] 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. [29] 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 [30]. Fritsch and Doss [31] 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).

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9s
DOI: https://doi.org/10.17762/ijritcc.v11i9s.7392
Article Received: 01 May 2023 Revised: 25 June 2023 Accepted: 20 July 2023

IV. RESULTS AND DISCUSSIONS

Table 1.provides the summary of various DSR techniques based on ML and DL approaches based on use of preprocessing, data augmentation, feature representation, classification algorithm, dataset, performance metrics and findings of study.

Table 1.Summary of ML and DL based DSR.

Reference [Citation]	Speech Enhancement	Data Augmentation	Feature extraction	Classifier	Database	Performance metrics	Remark
Yue, et al. [21]	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 style modelling
Yue, et al. [22]	EMA	-	CNN-LiGRU-FCMLP	softmax	TORGO	WER-4.6%	Over-fitting problem for high level articulatory feature fusion
Soleymanpour et al. [23]	-	TTS	DNN-HMM-BLiGRU	softmax	TORGO	WER-41.6%	The severity of dysarthric speech depends upon energy, duration and pitch of the signal.
Liu et al. [24]	-	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 [25]	-	Visual Data augmentation	Voicegram	S-CNN	UASpeech	Accuracy=67%	-Provides less temporal representation of speech -May cause vanishing gradient problem
Lin et al. [26]	-	-	-	CNN-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. [27]	-	-	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. [28]	-	-	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. [30]	-	-	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 [17]	-	-	Spectral, Cepstral, Voice Quality, Prosodic, Overall Speech features,	LDA, CART, NB, ANN, SVM, and RF	NEMOURS database	Average Ranking Score for Random Forest and Relief Feature Selection (4.88)	-Ability to classify speech based on severity level - Feature selection is important for DSR - Not applicable for larger dataset -less performance than deep learning approaches
Janbakhshi et al. [18]	-	-	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 [31]	-	-	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)

It is observed that the CNN based architectures provides better spatial and spectral representation of the dysarthric speech helps to achieve the better DSR accuracy. The use of RNN with CNN helps to improve the temporal attributes of the speech signal that boosts DSR results. Some of the deep learning algorithms utilizing traditional handcrafted features and pre-processing techniques for representation of speech and speech enhancement respectively have shown significant improvement in the DSR performance.

V. CONCLUSION

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. Various DSR techniques faces challenges due to articulation variability of the speaker, voice quality change, intelligibility issue, speaking rate variation, limited training data, lower contextual understanding, larger individual variability, etc. The DL algorithms utilized for the DSR have shown significant improvement in the DSR accuracy however its performance is limited due to extensive hyper-parameter tuning, complex architecture, less implementation flexibility on the standalone devices due to larger training and testing time and data scarcity issue. Many researcher's have focused on the detection of the dysarthric speech only and very less focus has been given on the dysarthric speech correction.

In future, the performance of DSR can be further improved by analyzing the intelligibility of the speaker, data augmentation to minimize the data scarcity problem, affective state estimation of the dysarthric people.

Conflicts of Interest

Authors must identify and declare any personal circumstances or interest that may be perceived as inappropriately influencing the representation or interpretation of reported research results. If there is no conflict of interest, please state "The authors declare no conflict of interest."

REFERENCES

[1] Wei Xue, Catia Cucchiarini, Roeland van Hout, and Helmer Strik. 2023. Measuring the intelligibility of dysarthricspeechthrough automatic speech recognition in a pluricentric language. Speech Commun. 148, C (Mar 2023), 23–30. https://doi.org/10.1016/j.specom.2023.02.004

[2] Aisha Jaddoh, Fernando Loizides & Omer Rana (2022) Interaction between people with dysarthria and speechrecognitio systems: A review, Assistive Technology, DOI: 10.1080/10400435.2022.2061085

[3] Chen, L. Special Issue on Automatic Speech Recognition. Appl. Sci. 2023, 13, 5389. https://doi.org/10.3390/app13095389

[4] Shih, D.-H.; Liao, C.-H.; Wu, T.-W.; Xu, X.-Y.; Shih, M.-H. (2022). Dysarthria Speech Detection Using Convolutional Neural Networks with Gated Recurrent Unit. Healthcare. 2022. 10. 1956. https://doi.org/10.3390/healthcare10101956

[5] S. Alharbi et al., "Automatic Speech Recognition: Systematic Literature Review," in IEEE Access, vol. 9, pp. 131858-131876, 2021, doi: 10.1109/ACCESS.2021.31125

[6] Roger, V., Farinas, J. & Pinquier, J. (2022). Deep neural networks for automatic speech processing: a survey from large corpora to limited data. EURASIP Journal on Audio, Speech, and Music Processing. 2022. 19. https://doi.org/10.1186/s13636-022-00251-w

[7] Rista,A. & Kadriu,A.(2020).Automatic Speech Recognition: A Comprehensive Survey. SEEU Review. 15(2): 86-112. https://doi.org/10.2478/seeur-2020-0019

[8] C. Yu, X. Su and Z. Qian, "Multi-Stage Audio-Visual Fusion for Dysarthric Speech Recognition With Pre-Trained Models," in IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 31, pp. 1912-1921, 2023, doi: 10.1109/TNSRE.2023.3262001

[9] Bhangale, K. B., & Mohanaprasad, K. (2021). A review on speech processing using machine learning paradigm. International Journal of Speech Technology, 24, 367-388. https://doi.org/10.1007/s10772-021-09808-0

[10] Bhangale, Kishor Barasu, and Mohanaprasad Kothandaraman. (2022). Survey of Deep Learning Paradigms for Speech Processing. Wireless Personal Communications, 1-37. https://doi.org/10.1007/s11277-022-09640-y

[11] Narendra, N. P., & Alku, P. (2020). Glottal source information for pathological voice detection. IEEE Access, 8, 67745-67755. DOI: 10.1109/ACCESS.2020.2986171

[12] Bhavya K. R, & S. Pravinth Raja. (2023). Fruit Quality Prediction using Deep Learning Strategies for Agriculture. International Journal of Intelligent Systems and Applications in Engineering, 11(2s), 301–310. Retrieved from https://ijisae.org/index.php/IJISAE/article/view/2697

[13] Gurugubelli, K., & Vuppala, A. K. (2020). Analytic phase features for dysarthric speech detection and intelligibility assessment. Speech Communication, 121, 1-15. https://doi.org/10.1016/j.specom.2020.04.006

[14] Bhat, C., Vachhani, B., & Kopparapu, S. K. (2017, March). 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. DOI: 10.1109/ICASSP.2017.7953122

[15] Hasegawa-Johnson, M.; Gunderson, J.; Perlman, A.; Huang, T. (2006). Hmm-Based and Svm-Based Recognition of the Speech of Talkers with Spastic Dysarthria. In Proceedings of the 2006 IEEE International Conference on Acoustics Speed and Signal Processing Proceedings, Toulouse, France, 14–19 May 2006; IEEE: New York, NY, USA, 2006. DOI: 10.1109/ICASSP.2006.1660840

[16] Ms. Sweta Minj. (2012). Design and Analysis of Class-E Power Amplifier for Wired & Wireless Systems. International Journal of New Practices in Management and Engineering, 1(04), 07 - 13. Retrieved from http://ijnpme.org/index.php/IJNPME/article/view/9

[17] Rudzicz, F. (2009). Phonological features in discriminative classification of dysarthric speech. In Proceedings of the 2009 IEEE International Conference on Acoustics, Speech and Signal Processing, Taipei, Taiwan, 19–24 April 2009; IEEE: New York, NY, USA, 2009; pp. 4605–4608. DOI: 10.1109/ICASSP.2009.4960656

[18] Revathi, A., Nagakrishnan, R., & Sasikaladevi, N. (2022). Comparative analysis of Dysarthric speech recognition: multiple features and robust templates. Multimedia Tools and Applications, 81(22), 31245-31259. https://doi.org/10.1007/s11042-022-12937-6

[19] B. A. Al-Qatab and M. B. Mustafa. (2021). Classification of Dysarthric Speech According to the Severity of Impairment: an Analysis of Acoustic Features. In IEEE Access,9, 18183-18194. doi: 10.1109/ACCESS.2021.3053335.

[20] Janbakhshi, P., Kodrasi, I., & Bourlard, H. (2021). Subspace-Based Learning for Automatic Dysarthric Speech Detection. IEEE Signal Processing Letters, 28, 96–100. doi:10.1109/lsp.2020.3044503

[21] Bhangale, K., & Mohanaprasad, K. (2022). Speech emotion recognition using mel frequency log spectrogram and deep convolutional neural network. In Futuristic Communication and Network Technologies: Select Proceedings of VICFCNT 2020 (pp. 241-250). Springer Singapore. https://doi.org/10.1007/978-981-16-4625-6_24

[22] Fathima, N., Patel, T., Mahima, C., & Iyengar, A. (2018, September). TDNN-based Multilingual Speech Recognition System for Low Resource Indian Languages. In Interspeech (pp. 3197-3201).

[23] Yue, Z., Loweimi, E., & Cvetkovic, Z. (2022, May). Raw source and filter modelling for dysarthric speech recognition. In ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 7377-7381). IEEE. DOI: 10.1109/ICASSP43922.2022.9746553

[24] Smith, J., Jones, D., Martinez, J., Perez, A., & Silva, D. Enhancing Engineering Education through Machine Learning: A Case Study. Kuwait Journal of Machine Learning, 1(1). Retrieved from http://kuwaitjournals.com/index.php/kjml/article/view/86

[25] Yue, Z., Loweimi, E., Cvetkovic, Z., Christensen, H., & Barker, J. (2022, May). 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. DOI: 10.1109/ICASSP43922.2022.9746855

[26] Soleymanpour, M., Johnson, M. T., Soleymanpour, R., & Berry, J. (2022, May). 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. DOI: 10.1109/ICASSP43922.2022.9746585

[27] Liu, S., Geng, M., Hu, S., Xie, X., Cui, M., Yu, J., ... & Meng, H. (2021). Recent progress in the CUHK dysarthric speech recognition system. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 29, 2267-2281. DOI: 10.1109/TASLP.2021.3091805

[28] Shahamiri, S. R. (2021). Speech vision: An end-to-end deep learning-based dysarthric automatic speech recognition system. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 29, 852-861. DOI: 10.1109/TNSRE.2021.3076778

[29] Lin, Y. Y., Zheng, W. Z., Chu, W. C., Han, J. Y., Hung, Y. H., Ho, G. M., ... & Lai, Y. H. (2021). A speech command control-based recognition system for dysarthric patients based on deep learning technology. Applied Sciences, 11(6), 2477. https://doi.org/10.3390/app11062477

[30] Kodrasi, I., & Bourlard, H. (2020). Spectro-temporal sparsity characterization for dysarthric speech detection. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 28, 1210-1222. DOI: 10.1109/TASLP.2020.2985066

[31] Kodrasi, I. (2021). Temporal envelope and fine structure cues for dysarthric speech detection using CNNs. IEEE Signal Processing Letters, 28, 1853-1857. DOI: 10.1109/LSP.2021.3108509

[32] Chandrashekar, H. M., Karjigi, V., & Sreedevi, N. (2020). Investigation of different time-frequency representations for

intelligibility assessment of dysarthric speech. Ieee transactions on neural systems and rehabilitation engineering, 28(12), 2880-2889. DOI: 10.1109/TNSRE.2020.3035392

[33] Chandrashekar, H. M., Karjigi, V., & Sreedevi, N. (2019). Spectro-temporal representation of speech for intelligibility assessment of dysarthria. IEEE Journal of Selected Topics in Signal Processing, 14(2), 390-399. doi: 10.1109/JSTSP.2019.2949912.