Harnessing Deep Learning Techniques for Dysarthria Detection

Varalakshmi P, Sanmitha V S, Dhivyadharshni S B and Saravanan S

Department of Information Science and Technology, Anna University - Chennai, India

Email: varanip@gmail.com,sanmithasadhishkumar@gamil.com, dhivyadharshu07@gmail.com, saravanan20030205@gmail.com

***Abstract*—Dysarthria, a motor speech disorder resulting from neurological impairments. This study explores various approaches for the automated detection of dysarthria, integrating both traditional and emerging technologies. Speech assessments by acoustic analysis, machine learning models, and deep learning techniques are considered. Machine learning models are trained on datasets containing normal and dysarthric speech for males and females, while deep learning methods, including convolutional and recurrent neural networks, are employed to automatically learn features from speech data. Our goal is to develop a reliable and accessible dysarthria detection system that complements the expertise of healthcare professionals. Validation using diverse datasets is emphasized, acknowledging the importance of early detection and intervention in improving outcomes for individuals with dysarthria.**

***Index Terms*—Acoustic analysis, Convolutional neural networks, Deep learning techniques, Dysarthria, Machine learning models, Recurrent neural networks**

I. Introduction

Dysarthria stands as a significant impediment to fluent speech, characterized by challenges in articulation owing to muscle weakness or impaired muscular coordination crucial for speech production. This disorder manifests across a spectrum of severity, with individuals experiencing varying degrees of difficulty in vocalizing due to underlying neurological conditions, traumatic brain injuries, strokes, or other afflictions affecting the neuromuscular apparatus involved in speech. The ensemble of muscles indispensable for speech—comprising the lips, tongue, vocal cords, and diaphragm—may encounter disruptions in their functionality, culminating in a range of symptoms. These symptoms encompass slurred speech, compromised articulation, imprecise pronunciation, and struggles in modulating the volume or pitch of utterances. The extent and nature of these manifestations are contingent upon the specific etiology of dysarthria and the precise localization of neural impairment within the intricate network governing speech production.

Health professionals must conduct regular assessments of individuals with dysarthria to detect and address potential speech difficulties. Nevertheless, there are limitations to this traditional approach, including financial constraints, operational costs, and limitations in terms of scheduling and availability.

Deep learning models have demonstrated significant potential in tackling the issue of dysarthria detection. In this study, we introduce an innovative approach to dysarthria detection using deep learning models, specifically including Convolutional Neural Networks (CNN), CNN with Transformer, CNN with Long Short-Term Memory (LSTM), Gaussian Mixture Model (GMM), CNN with GMM, Gated Recurrent Unit (GRU)and Recurrent Neural Networks (RNN). We assess the performance of these models on a part of TORGO dataset utilizing various optimization techniques and measure their accuracy using metrics such as Mean Average Precision (MAP). Our findings indicate that deep learning models surpass conventional methods for dysarthria detection, achieving an impressive accuracy of 99.6% with the CNN combined with Transformer architecture.

The structure of this paper unfolds as follows: In Section II, we delve into a comprehensive review of prior research concerning Dysarthria detection. Section III elucidates our novel approach. Following that, in Section IV, we present a detailed examination of the performance of each model, accompanied by experimental results, and discuss the datasets employed for evaluation. Ultimately, Section V encapsulates our findings, drawing conclusions and outlining potential avenues for future research endeavors.

II. Related Works

In [1], Up to 90% of patients with Parkinson's disease (PD) suffer from hypokinetic dysarthria (HD). A univariate classification method was primarily utilized to identify Hypokinetic Dysarthria associated with Parkinson's disease, achieving an F1-score of 62.63%, while trained on Czech native speakers dataset. A bivariate general classification approach employing diverse models reported a sensitivity of 83.42%. The General Super-positional Model (GSM) in [2] analyzed F0 variability in dysarthria, particularly in Parkinson's disease (PD) speech, with a focus on prosody. The GSM method decomposes the F0 curve into phrase and accent curves, elucidating mechanisms contributing to reduced F0 variability, notably in PD, through factors such as phrase curve slope, number of feet, and accent curve height.

Unlike [3] the utilization of a CNN model in the cited study, the SincNet model was employed for severity classification, achieving accuracy rates of 95.7% and 99.6%, respectively. Additionally, in [4] attention-based neural network trained on mel-filterbanks for dysarthria detection explored the integration of a Per Channel Energy Normalization (PCEN) block or learnable time-domain filterbanks to adapt feature extraction to dysarthric speech characteristics, ultimately achieving optimized feature extraction, compression, and normalization, thus improving efficacy in dysarthria detection tasks. Furthermore, solely in [5] MFCC features were utilized for severity classification, resulting in a superior accuracy of 99.6% compared to the cited research's use of both MFCC and CQCC features, achieving an accuracy of 99.6% using CNN with Transformer architecture. In [6], An approach for automatic Dysarthria detection using Centroid Formants, weighted averages of the first four formants extracted from speech signals, distinguished Dysarthric from healthy speech samples utilizing a neural network classifier.

Moreover,Convolutional Neural Networks (CNN) were employed in [7] to predict enhanced speech from dysarthric utterances, utilizing Logarithmic Power Spectra (LPS) features. Initially, the Short-time Fourier transform (STFT) extracted complex-valued frames in the frequency domain, followed by a logarithmic operation to compute LPS. Dynamic Time Warping (DTW) aligned LPS features for supervised CNN training, optimizing model parameters to minimize mean square error (MSE) and enhance dysarthric speech. The study followed a three-stage methodology in [8], comprising feature extraction, genetic algorithm-based feature selection, and Support Vector Machines (SVMs) classification. Speech-specific features such as MFCCs, peak amplitude, skewness, kurtosis, fundamental frequency, and formants were extracted. Genetic algorithms were then employed to identify discriminant features for diagnosing dysarthric speech. SVMs with a linear kernel were utilized for both diagnosis and severity classification, with multi-class SVMs used for severity labeling. Diagnosis and assessment of dysarthric speech relied on support vector machines, primarily focusing on severity classification in [9], resulting in an accuracy of 98%. Employing CNN with Transformer architecture achieved a higher accuracy rate of 99.6%. Additionally, in [10] the citation mentions the use of Weak Speech Supervision (WSS) methodology to generate weak data from unlabeled training sets. Conversely, hand-crafted features like MFCCs were employed to train the model. While WSS achieved an accuracy of 35.68%, the approach yielded a significantly higher accuracy of 99.6%.

III. Proposed Work

Dysarthria, a motor speech disorder, significantly impacts individuals' ability to communicate effectively, posing a considerable challenge to public health. In this proposed research, we introduce an innovative approach to dysarthria detection utilizing deep learning models with diverse architectures. Specifically, we explore the effectiveness of Convolutional Neural Networks (CNN), CNN with Transformer, CNN with LSTM (Long Short-Term Memory) , CNN with Gaussian Mixture Model (GMM), Recurrent Neural Network (RNN), and Gated Recurrent Unit (GRU). The proposed architecture is shown in Fig 1. Here, the speech data set is fed into various deep learning models in order to predict the outcome and the performance of various models are compared and the model which performs the best is suggested for deployment in health centers.

We aim to compare the performance metrics of the above mentioned models on a part of Torgo Dataset using various optimization techniques, evaluating their performance based on accuracy Sensitivity, Precision, F1 Score and Specificity. Our motive is to showcase the effectiveness of deep learning in identifying dysarthria and to assess the efficacy of different models for this task. Additionally, we conduct an in-depth review of each model, providing insights that can guide future research in this domain.

A. Feature Extraction

From an audio, we can extract many features including Mel Frequency Ceptral Coefficients, rms, Spectral Bandwidth, Spectral Centroid, Spectral Contrast, Spectral Flatness, Spectral Rolloff, Zero Crossing Rate, etc

Out of these features, we use Mel Frequency Ceptral Coeffiecients (MFCCs) as MFCCs are widely used in speech recognition. They are coefficients representing the short-term power spectrum of a sound signal.

The extracted MFCCs are converted into Data Frames which is then splitted into MFCC data and their respective labels.

B. Deep Learning

Deep learning has now grown as an inevitable wealth in machine learning, providing feasible and innovative solutions for complicate challenges across multiple domains.Not like all the conventional machine learning techniques that rely on manually crafted features, deep learning employs neural networks comprising multiple layers. These networks comprises of interconnected neurons that collectively analyze input data.

In the context of dysarthria detection, deep learning models play a crucial role in analyzing speech patterns and identifying speech disorders. Convolutional Neural Networks (CNNs) are foundational in processing spectrogram images of speech signals, enabling the extraction of relevant features for dysarthria detection. Additionally, recurrent neural networks (RNNs), including Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU), excel in sequential data analysis, making them well-suited for capturing temporal dependencies in speech data.

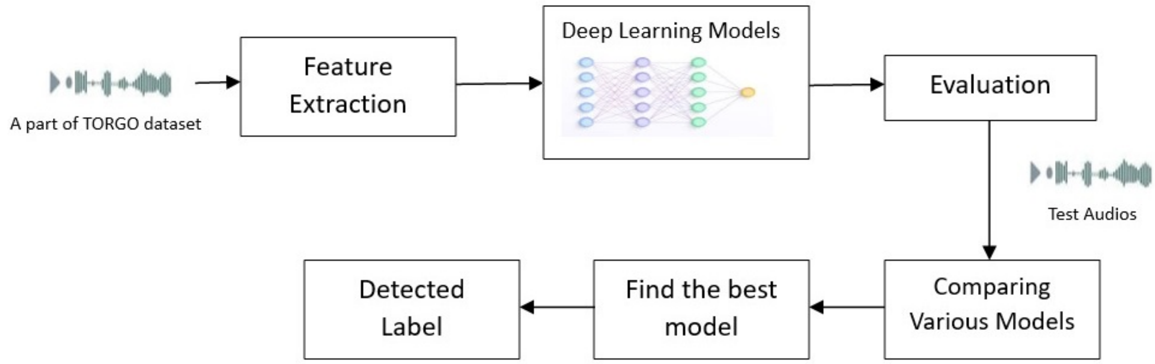

Fig. 1. Architecture diagram of proposed work

Moreover, the integration of transformers in deep learning architectures has shown promising results in various natural language processing tasks, including dysarthria detection. By leveraging self-attention mechanisms, transformer-based models can effectively capture long-range dependencies in speech sequences, enhancing the accuracy of dysarthria identification.

In addition to these deep learning approaches, models such as Gaussian Mixture Model (GMM) can also be employed in dysarthria detection tasks

In summary, the diverse range of deep learning models, including CNNs, transformers, RNNs, and GMMs, offer versatile solutions for dysarthria detection, enabling accurate analysis of speech signals and facilitating early diagnosis of speech disorders.

1) Gaussian Mixture Model: This model applies a Gaussian Mixture Model (GMM) for dysarthria prediction, leveraging statistical modeling and probability distributions to identify patterns within the speech features.

2) CNN (Convolutional Neural Networks): This model is a concise Convolutional Neural Network (CNN) model for binary classification on input data with dimensions (16, 8, 1). The architecture shown in Table I includes two convolutional layers with max pooling, followed by flattening and two dense layers.

3) CNN with Transformer: This model integrates Convolutional Neural Network (CNN) architecture with Transformer, featuring multi-head attention mechanisms to capture global dependencies and layer normalization for improved training stability. Notably, this hybrid model shown in Table II, leveraging the strengths of both convolutional and attention mechanisms, gives the best accuracy in predicting whether the patient is affected by dysarthria or not.

TABLE I
CNN MODEL ARCHITECTURE

Layer Name	Input Shape	Output Shape	Weights	Activation
InputLayer	(None, 16, 8, 1)	(None, 16, 8, 1)	0	-
Conv2D	(None, 16, 8, 1)	(None, 16, 8, 32)	320	ReLU
MaxPooling 2D	(None, 16, 8, 32)	(None, 8, 4, 32)	0	-
Conv2D	(None, 8, 4, 32)	(None, 8, 4, 64)	18496	ReLU
MaxPooling 2D	(None, 8, 4, 64)	(None, 4, 2, 64)	0	-
Flatten	(None, 4, 2, 64)	(None, 512)	0	-
dense	(None, 512)	(None, 32)	16,416	ReLU
dense	(None, 32)	(None, 1)	33	Sigmoid

4) CNN with GMM: This model combines a Convolutional Neural Network (CNN) and a Gaussian Mixture Model (GMM) as shown in Table III for classification. The CNN extracts features from input data, which are then concatenated with the original features. A GMM with two components is trained on the combined features from the training data and used for predictions on the testing data. This hybrid approach synergizes CNN's spatial pattern recognition with GMM's probabilistic modeling for enhanced classification.

TABLE II
CNN WITH TRANSFORMER - MODEL ARCHITECTURE

Layer Name	Input Shape	Output Shape	Weight	Activation
InputLayer	(None, 16, 8, 1)	(None, 16, 8, 1)	0	-
Conv2D	(None, 16, 8, 1)	(None, 16, 8, 32)	320	ReLU
MaxPooling 2D	(None, 16, 8, 32)	(None, 8, 4, 32)	0	-
Conv2D	(None, 8, 4, 32)	(None, 8, 4, 64)	18496	ReLU
MaxPooling 2D	(None, 8, 4, 64)	(None, 4, 2, 64)	0	-
Transformer Encoder	(None, 4, 2, 64)	(None, 4, 2, 64)	33344	-
Flatten	(None, 4, 2, 64)	(None, 512)	0	-
Flatten	(None, 4, 2, 64)	(None, 512)	0	-
Concatenate	(None, 512), (None, 512)	(None, 1024)	0	-
Dense	(None, 1024)	(None, 32)	32800	ReLU
Dense	(None, 32)	(None, 1)	33	Sigmoid

TABLE III
CNN WITH GMM MODEL ARCHITECTURE

Layer Name	Input Shape	Output Shape	Weights	Activation
InputLayer	(None, 16, 8, 1)	(None, 16, 8, 1)	0	-
Conv2D	(None, 16, 8, 1)	(None, 16, 8, 32)	320	ReLU
MaxPooling 2D	(None, 16, 8, 32)	(None, 8, 4, 32)	0	-
Conv2D	(None, 8, 4, 32)	(None, 8, 4, 64)	18496	ReLU
MaxPooling 2D	(None, 8, 4, 64)	(None, 4, 2, 64)	0	-
Flatten	(None, 4, 2, 64)	(None, 512)	0	-
Concatenate	(None, 512)	(None, 640)	0	-
GaussianMixture	(None, 640)	(None, 1)	0	-

5) Recurrent Neural Network: This model utilizes a standard Recurrent Neural Network architecture shown in Table IV for dysarthria prediction, considering sequential dependencies in speech signals to capture temporal patterns.

TABLE IV
RNN MODEL ARCHITECTURE

Layer Name	Input Shape	Output Shape	Weights	Activation
InputLayer	(None, 16, 8, 1)	(None, 16, 8, 1)	0	-
Simple RNN	(None, 16, 8, 1)	(None, 16, 32)	1312	ReLU
Simple RNN	(None, 16, 32)	(None, 64)	6208	ReLU
Dense	(None, 64)	(None, 32)	2080	ReLU
Dense	(None, 32)	(None, 1)	33	Sigmoid

6) Gated Recurrent Unit: This model incorporates the Gated Recurrent Unit, a type of recurrent neural network (RNN), to capture long-term dependencies in sequential speech data, enhancing the model's ability to discern temporal patterns. The architecture of this model is shown in Table V

TABLE V
GRU MODEL ARCHITECTURE

Layer Name	Input Shape	Output Shape	Weights	Activation
InputLayer	(None, 16, 8, 1)	(None, 16, 8, 1)	0	-
GRU	(None, 16, 8, 1)	(None, 16, 32)	4032	ReLU
GRU	(None, 16, 32)	(None, 64)	18816	ReLU
Dense	(None, 64)	(None, 32)	2080	ReLU
Dense	(None, 32)	(None, 1)	33	Sigmoid

7) CNN with Long Short Time Memory: This model is a hybrid of convolutional and recurrent layers as shown in Table VI, tailored for structured data with dimensions (16, 8, 1). Convolutional layers capture spatial features, followed by max pooling. The data is then reshaped and processed by Bidirectional Long Short-Term Memory (BiLSTM) layers for temporal understanding. The final dense layers with a sigmoid activation to produce a binary classification output

TABLE VI
CNN WITH LSTM MODEL ARCHITECTURE

Layer Name	Input Shape	Output Shape	Weights	Activation
InputLayer	(None, 16, 8, 1)	(None, 16, 8, 1)	0	-
Conv2D	(None, 16, 8, 1)	(None, 16, 8, 32)	320	ReLU
MaxPooling 2D	(None, 16, 8, 32)	(None, 8, 4, 32)	0	-
Conv2D	(None, 8, 4, 32)	(None, 8, 4, 64)	18496	ReLU
MaxPooling 2D	(None, 8, 4, 64)	(None, 4, 2, 64)	0	-
Conv2D	(None, 4, 2, 64)	(None, 4, 2, 128)	73856	ReLU
MaxPooling 2D	(None, 4, 2, 128)	(None, 2, 1, 128)	0	-
Flatten	(None, 2, 1, 128)	(None, 256)	0	-
Reshape	(None, 256)	(None, 16, 16)	0	-
Bidirectional	(None, 16, 16)	(None, 16, 128)	41472	-
Bidirectional	(None, 16, 128)	(None, 64)	41216	-
Dense	(None, 64)	(None, 32)	2080	ReLU
Dense	(None, 32)	(None, 1)	33	Sigmoid

IV. RESULTS AND ANALYSIS

In the realm of deep learning, leveraging platforms like Google Colab has become increasingly popular due to its accessibility and powerful computational resources. Researchers and practitioners often turn to Colab for its seamless integration with TensorFlow and Keras, enabling the utilization of deep learning models with ease. Scikit-learn serves as a fundamental tool for traditional machine learning tasks, while Keras simplifies complex neural network architectures, particularly for convolutional neural networks (CNNs), essential for feature extraction. These extracted features are then combined with other techniques, often within the Colab environment, facilitating Gaussian Mixture Model (GMM)-based classification. TensorFlow, at the core of Colab's capabilities, offers versatility in constructing sophisticated neural networks, integrating convolutional and self-attention mechanisms for robust feature learning. Thus, Colab emerges as an indispensable platform for developing and deploying deep learning models efficiently.

A. Dataset

The dataset for training and testing is taken from TORGO dataset. TORGO dataset is the result of a collaboration between Computer Science and Speech-Language Pathology departments at the University of Toronto and the Holland-Bloorview Kids Rehab hospital in Toronto. We use a part of TORGO dataset that contains 2000 audio samples [11], 500 for dysarthric male, female, non dysarthric male and female.

10% of these samples is used as testing dataset i.e 50 samples from each class.

Since this dataset is a balanced dataset, no balancing technique is required.

B. Evaluation Metrics

Performance evaluation metrics, as depicted in Table VII are utilized to gauge the effectiveness of trained models. They offer a quantitative assessment of a model's performance in a specific task and play a crucial role in evaluating the efficacy of deep learning algorithms. Within this framework, several metrics are considered to demonstrate the performance of models, highlighting their effectiveness.

1) Train and Test Dataset: The training set within the TORGO dataset consists of input data paired with corresponding accurate output values. Its purpose is to facilitate the development of a model capable of accurately predicting outcomes for new data. In contrast, the test set comprises 10% of a dataset used to assess the performance of a deep learning model. This test set is exclusively employed to evaluate the model's accuracy on previously unseen data and is not utilized during the training process.

2) Sensitivity: Sensitivity measures the accuracy of a classification model in identifying true positive instances. It's calculated by dividing the number of true positives by the sum of true positives and false negatives. Sensitivity serves as a vital indicator of a model's capability to detect genuine positive cases, thus playing a crucial role in evaluating the effectiveness of deep learning models

3) F1 Score: The F1-score is calculated as twice the product of precision and recall, divided by the sum of precision and recall. This metric represents a weighted average of precision and recall and serves as a usual measure for evaluating the overall performance of a deep learning model.

4) Precision: Precision is computed as the ratio of true positive predictions to the sum of true positives and false positives. It represents the accuracy of positive predictions made by the model. Precision is a fundamental metric used to gauge the reliability of a deep learning model's positive predictions.

5) Specificity: Specificity evaluates the model's ability to accurately identify negative cases, calculated by dividing the true negatives with the sum of true negatives and false positives. This metric provides insight into the model's proficiency in distinguishing true negative instances from false positives, contributing to a comprehensive assessment of its performance.

TABLE VII
PERFORMANCE METRICS

Model Name	**Accuracy**	**Loss/Error**	**Sensitivity**	**Precision**	**F1 score**	**Specificity**
Gaussian Mixture Model	0.23333	-	-	-	-	-
CNN	0.99179	0.04161	1.0	0.995	0.995	0.99
CNN with Transformer	0.99599	0.01669	1.0	1.0	1.0	1.0
CNNwith GMM	0.48888	-	-	-	-	-
RNN	0.97690	0.06651	0.98	0.975	0.975	1.0
GRU	0.95918	0.11317	0.99	0.985	0.985	0.98
CNN with LSTM	0.98987	0.03385	0.99	0.99	0.99	0.99

V. CONCLUSION

The quest for effective dysarthria detection has driven the exploration of diverse methodologies, encompassing both traditional clinical assessments and cutting-edge technologies. Speech and language pathologists, armed with standardized tests and keen clinical observations, play a crucial role in diagnosing dysarthria. Acoustic analysis emerges as a valuable tool, shedding light on distinct patterns associated with dysarthric speech.

Machine learning models, strategically trained on datasets that span both normal and dysarthric speech, demonstrate promising capabilities in discerning nuanced speech characteristics. The integration of deep learning techniques, including convolutional and recurrent neural networks, further showcases the potential to automatically extract relevant features from speech data. However, ensuring the robustness and generalizability of dysarthria detection systems demands a comprehensive validation process, leveraging diverse datasets.

The imperative acknowledgment arises that these technologies should function as aids to healthcare professionals, not replacements. This emphasizes the collaborative synergy between human expertise and technological advancements in the realms of early dysarthria detection and intervention. The pivotal significance of early identification remains evident, as it not only optimizes therapeutic interventions but also holds the potential to improve communication outcomes for individuals affected by this motor speech disorder. The ongoing integration of advanced technologies, coupled with collaborative efforts between professionals and automated systems, presents a promising trajectory for enhancing the overall landscape of dysarthria detection and care.

References

[1] J. Mucha et al., "Identification of hypokinetic dysarthria using acoustic analysis of poem recitation," 2017 40th International Conference on Telecommunications and Signal Processing (TSP), Barcelona, Spain, 2017, pp. 739-742, doi: 10.1109/TSP.2017.8076086. keywords: Speech;High definition video;Feature extraction;Correlation;Parkinson's disease;acoustic analysis;binary classification;hypokinetic dysarthria;Parkinson's disease;poem recitation.

[2] M. S. E. Langarani and J. van Santen, "Modeling fundamental frequency dynamics in hypokinetic dysarthria," 2014 IEEE Spoken Language Technology Workshop (SLT), South Lake Tahoe, NV, USA, 2014, pp. 272-276, doi: 10.1109/SLT.2014.7078586. keywords: Speech;Feature extraction;Mathematical model;Accuracy;Foot;Equations;Support vector machines;Hypokinetic dysarthria;Parkinson's Disease;Pitch decomposition.

[3] K. Radha and M. Bansal, "Automated Detection and Severity Assessment of Dysarthria using Raw Speech," 2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT), Delhi, India, 2023, pp. 1-7, doi: 10.1109/ICCCNT56998.2023.10307923. keywords: Deep learning;Speech analysis;Neural networks;Speech enhancement;Muscles;Paralysis;Convolutional neural networks;Dysarthria severity level assessment;Speech-related disorders;Deep learning;SincNet;Raw waveforms.

[4] J. Millet and N. Zeghidour, "Learning to Detect Dysarthria from Raw Speech," ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brighton, UK, 2019, pp. 5831-5835, doi: 10.1109/ICASSP.2019.8682324. keywords: Task analysis;Neural networks;Time-domain analysis;Speech recognition;Databases;Standards;Computational modeling;dysarthria;paralinguistic;classification;waveform;lstm.

[5] A. A. Joshy and R. Rajan, "Automated Dysarthria Severity Classification: A Study on Acoustic Features and Deep Learning Techniques," in IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 30, pp. 1147-1157, 2022, doi: 10.1109/TNSRE.2022.3169814. keywords: Deep learning;Databases;Speech recognition;Support vector machines;Neural networks;Estimation;Diseases;Deep learning;dysarthria;i-vectors;severity assessment.

[6] 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," IET 3rd International Conference on Intelligent Signal Processing (ISP 2017), London, 2017, pp. 1-6, doi: 10.1049/cp.2017.0360. keywords: Dysarthria;speech disorder;Centroid Formants;Neural Networks.

[7] S. Wang et al., "Dysarthric Speech Enhancement Based on Convolution Neural Network," 2022 44th Annual International Conference of the IEEE Engineering in Medicine Biology Society (EMBC), Glasgow, Scotland, United Kingdom, 2022, pp. 60-64, doi: 10.1109/EMBC48229.2022.9871531. keywords: Training;Time-frequency analysis;Neural networks;Training data;Focusing;Speech enhancement;Internet.

[8] G. Vyas, M. K. Dutta, J. Prinosil and P. Harár, "An automatic diagnosis and assessment of dysarthric speech using speech disorder specific prosodic features," 2016 39th International Conference on Telecommunications and Signal Processing (TSP), Vienna, Austria, 2016, pp. 515-518, doi: 10.1109/TSP.2016.7760933. keywords: Speech;Feature extraction;Testing;Genetic algorithms;Support vector machines;Databases;Training;Dysarthric speech;diagnosis;assessment;speech disorder;prosodic features;support vector machines.

[9] A. A. Joshy and R. Rajan, "Automated Dysarthria Severity Classification Using Deep Learning Frameworks," 2020 28th European Signal Processing Conference (EUSIPCO), Amsterdam, Netherlands, 2021, pp. 116-120, doi: 10.23919/Eusipco47968.2020.9287741. keywords: Deep learning;Support vector machines;Neural networks;Speech recognition;Signal processing;Reliability;Mel frequency cepstral coefficient;dysarthria;intelligibility;automatic assessment;deep learning.

[10] M. Purohit, M. Parmar, M. Patel, H. Malaviya and H. A. Patii, "Weak Speech Supervision: A case study of Dysarthria Severity Classification," 2020 28th European Signal Processing Conference (EUSIPCO), Amsterdam, Netherlands, 2021, pp. 101-105, doi: 10.23919/Eusipco47968.2020.9287502. keywords: Training;Training data;Machine learning;Signal processing;Data models;Task analysis;Speech processing;Dysarthria;Severity-based Classification;Data Scarcity;Weak Supervision;CNN.

[11] https://www.kaggle.com/datasets/iamhungundji/dysarthria-detection