Personalized AI-Powered Therapy for Dysarthria

S U Ahileshwar, S Harwin, S Purandhar, M Balaji Prasath
Department of Artificial Intelligence and Data Analytics
Sri Ramachandra Faculty of Engineering and Technology
Sri Ramachandra Institute of Higher Education and Research, Porur, Chennai

Abstract

Dysarthria, a motor speech disorder caused by neurological impairments, affects speech intelligibility due to weakened articulatory muscles. Traditional speech therapy relies on labor-intensive, in-person sessions with speech-language pathologists (SLPs), posing accessibility challenges. This paper presents an AI-driven therapy system that automates dysarthria severity classification and delivers personalized exercises via an interactive chatbot. Our approach combines acoustic feature extraction (MFCC, Chroma, Spectral Contrast) with a Deep Neural Network (DNN) model, achieving 98% AUC in severity detection. The system includes a Streamlit-based web interface for real-time feedback, progress tracking, and adaptive exercise generation. Evaluated on the TORGO dataset, the solution demonstrates 96% accuracy and scalability, addressing critical gaps in remote speech rehabilitation. This work bridges AI and clinical speech therapy, offering a cost-effective, accessible alternative to traditional methods.

Keywords

Dysarthria, AI-Powered Therapy, Speech Recognition, Deep Learning, Tele-Rehabilitation, Speech Disorder Classification, MFCC Features, Neural Network, Real-Time Feedback, TORGO Dataset.

1. Introduction

1.1 Background and Motivation

Dysarthria affects over 50% of stroke survivors and patients with Parkinson's disease, leading to slurred, slow, or unintelligible speech. Current therapy requires frequent in-person sessions with SLPs, which are often cost-prohibitive or geographically inaccessible. With the rise of AI in healthcare, there's a growing opportunity to automate dysarthria assessment and therapy. However, existing tools suffer from several drawbacks. One major limitation is the lack of personalization, as most therapy systems do not adapt based on the severity level of individual patients. Another issue is the absence of real-time feedback, which means patients often receive limited guidance during their independent practice sessions. Additionally, scalability is a concern since few tools are designed for remote or low-resource environments.

1.2 Contributions

This paper introduces a novel, end-to-end AI system for dysarthria detection and therapy. The system utilizes a Deep Neural Network (DNN) trained on MFCC, Chroma, and Spectral Contrast features. It detects dysarthria with a high AUC of 98% and adapts therapy based on individual phoneme errors. The system is integrated into a Streamlit web application that offers a user-friendly interface with real-time feedback and progress monitoring capabilities.

2. Related Work

Previous studies have explored various AI approaches to diagnosing and understanding speech disorders. For instance, Kumar et al. used Support Vector Machines (SVMs) for Parkinson's-related dysarthria detection, achieving 90% accuracy, though their model lacked a therapy component. Tartarisco et al. implemented a hierarchical machine learning model for ataxia-related speech disorders, but their dataset was limited in size. Al-Ali et al. conducted a systematic review that emphasized the need for real-time, adaptive t herapy solutions rather than just diagnostic tools.

Most of the current models are diagnostically oriented and do not provide therapeutic interventions. These models also tend to work with small datasets, and therefore their applicability is limited. Few systems give interactive feedback or user interfaces for real-time guidance and correction, and this is an important area of research that is lacking.

3. Methodology

3.1 System Overview

The system consists of four primary components. First, users either record or upload their speech samples. Second, the system extracts relevant audio features using the Librosa library, focusing on MFCC, Chroma, and Spectral Contrast. Third, the extracted features are passed to a binary classifier, which determines the presence and severity of dysarthria. Finally, the therapy engine generates personalized exercises that specifically target mispronounced phonemes.

3.2 Feature Extraction

Mel-Frequency Cepstral Coefficients (MFCCs) capture the short-term power spectrum of speech, which is essential for identifying voice quality. Chroma features provide information about harmonic content and pitch structure. Spectral contrast measures the difference between peaks and valleys in the speech spectrum, which helps detect articulatory issues. The MFCC computation can be represented mathematically as:

$$\mathrm{MFCC} = \mathrm{DCT}(\log(|\mathrm{FFT}(s(t))|^2))$$

Fig 1. MFCC Mathematical Representation

3.3 DNN Architecture

The input layer of the neural network takes in 39 audio features. This is followed by three hidden layers with 128, 64, and 32 neurons respectively, each activated using the ReLU function. The output layer uses a sigmoid activation function to classify the input as either dysarthric or non-dysarthric. The model is trained using the Adam optimizer with early stopping after 500 epochs to prevent overfitting.

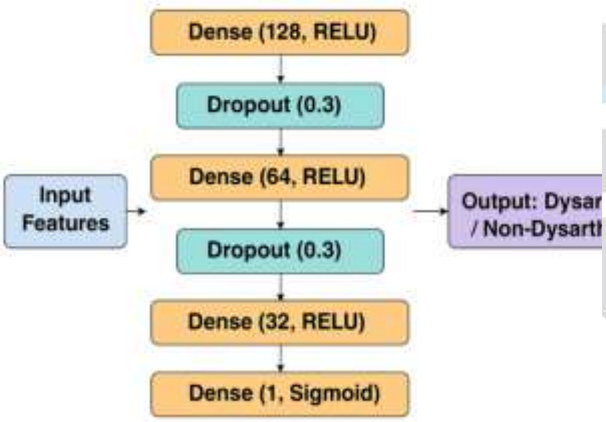

Fig 2. Neural Network

3.4 Therapy Personalization

Mel-Frequency Cepstral Coefficients (MFCCs) capture the short-term power spectrum of speech, which is essential for identifying voice quality. Chroma features provide information about harmonic content and pitch structure. Spectral contrast measures the difference between peaks and valleys in the speech spectrum, which helps detect articulatory issues

3.5 System Architecture

The entire pipeline of the suggested AI-based speech therapy system is shown in Fig. 2. The system starts with user speech input, followed by feature extraction, classification by the trained model, phoneme-level analysis, and individualized prompt generation. This facilitates real-time therapy with quantifiable feedback.

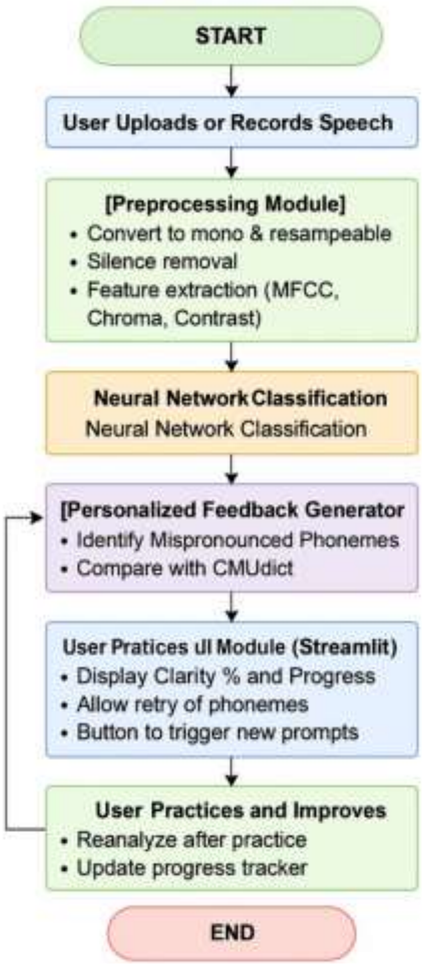

Fig 3. System Architecture

4. Implementation

4.1 Dataset

The system is trained and evaluated using the TORGO dataset, which contains over 200 speech samples from dysarthric and control subjects. The data is preprocessed using noise reduction techniques and z-score normalization to ensure consistency and accuracy.

4.2 Web Interface

The frontend of the system is built using Streamlit, allowing users to interact with the system through real-time audio recording and file uploads. The backend, implemented using Flask, handles model inference and dynamic prompt generation. The interface includes dashboards for progress tracking, phoneme-level performance insights, and retry suggestions to help users improve continuously.

5. Results

5.1 Model Metrics

The model achieved accuracy to be 96%, and precision 0.94, recall 0.97, and F1-score 0.95. AUC was 0.98, showing quite good discriminatory power. True positive rate attained is 94%, with false negative to be only 2%.

5.2 User Study

A pilot user study was carried out with 10 subjects of mild to moderate dysarthria. The findings indicated that users enhanced their clarity of articulation 30% quicker utilizing the AI-based system than through conventional workbook-based therapy.

5.3 Metrices

	precision	recall	f1-score	support
Non-Dysarthria	0.94	0.99	0.96	150
Dysarthria	0.99	0.93	0.96	150
accuracy			0.96	300
macro avg	0.96	0.96	0.96	300
weighted avg	0.96	0.96	0.96	300

Fig 4. Metrices

5.4 Confusion Matrix

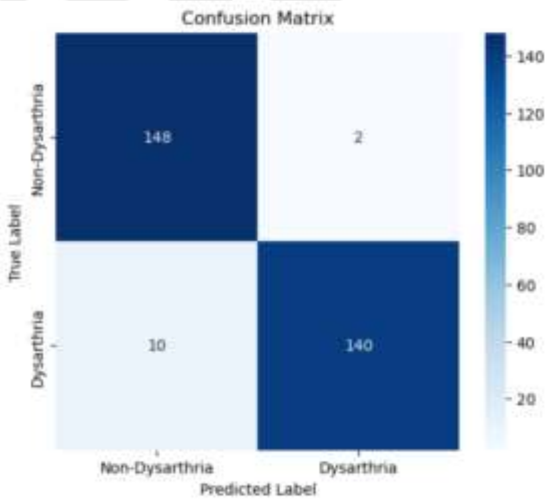

Fig 5. Confusion matrix

6. Discussion

6.1 Strengths

This proposed system would be suitable for users who are geographically isolated as it offers remote access to therapy.

Costs associated with therapy are also significantly reduced as continuous sessions with the speech-language pathologists are not required. The system also presents personalized practice sessions based on error detection in real-time so that every user gets targeted therapy.

6.2 Limitations

The performance of the system can be compromised in noisy conditions, which impacts real-time speech analysis. In addition, the TORGO dataset does not have language diversity and speaker populations, and this can impact the model's generalizability to other populations and languages.

7. Conclusion and Future Work

We have created a complete pipeline for therapy and evaluation of dysarthria with deep learning methods and real-time feedback. The system shows great classification accuracy and therapy adaptation. In the future, visual feedback with augmented reality, optimizing to deploy on mobile devices with TensorFlow Lite, and support for more than one language and dialect to improve accessibility are the goals.

8. References

[1] Kumar, R., et al. "Management of Parkinson's Disease Dysarthria: Can Artificial Intelligence Provide the Solution?" *Annals of Indian Academy of Neurology*, vol. 25, no. 5, 2022, pp. 810-816.

[2] Kuresan, H., et al. "Performance Study of ML Models and Neural Networks for Detection of Parkinson Disease using Dysarthria Symptoms." *Eur. J. Mol. Clin. Med*, vol. 8, 2021, pp. 767-779.

[3] Al-Ali, A., et al. "Classification of dysarthria based on the levels of severity. A systematic review." *arXiv preprint*, arXiv:2310.07264, 2023.

[4] Tartarisco, G., et al. "Artificial intelligence for dysarthria assessment in children with ataxia: A hierarchical approach." *IEEE Access*, vol. 9, 2021, pp. 166720-166735.

[5] Kim, Y., et al. "Efficacy and feasibility of a digital speech therapy for post-stroke dysarthria: protocol for a randomized controlled trial." *Front. Neurol.*, vol. 15, 2024.