AI-Driven Acoustic Feature Extraction for Automated Stuttering Detection

Dr. Salma Jabeen[1], Dr. Ayodeji Olalekan Salau[2]

[1] Department of information Science and Engineering Don Bosco Institute of Technology;

[2] Department of Electrical/Electronics and Computer Engineering Afe Babalola University, Ado-Ekiti, Nigeria;

Email ID

pdf.salma@lincoln.edu.my

ayodejisalau98@gmail.com

Abstract: Stuttering is recognized as a speech fluency disorder characterized by involuntary repetitions, prolongations, and blocks occurring during natural speech production [1]. Conventional clinical diagnosis predominantly depends on perceptual assessment by speech-language professionals, which may introduce subjectivity and inter-evaluator variability [1]. With significant progress in speech signal processing and the emergence of artificial intelligence (AI), automated and objective approaches to stuttering detection have become increasingly viable [2], [3].This paper provides an overview of AI-based approaches for extracting acoustic features used in automated stuttering detection, with particular emphasis on Mel-Frequency Cepstral Coefficients (MFCCs) and deep learning models [2], [4]. A comparative analysis of existing literature indicates that deep neural networks (DNNs), including CNNs, LSTMs, and transformer-based architectures, consistently outperform traditional machine learning algorithms when applied to disfluent speech data [3], [5]–[8]. Despite these promising outcomes, several limitations persist, primarily due to speaker variability, imbalanced datasets, and the high computational complexity associated with deep learning-based approaches [6], [7], [9]–[11].

Keywords Artificial Intelligence; Mel-Frequency Cepstral Coefficients; Deep Neural Networks; Convolutional Neural Network; LSTMs

Introduction

stuttering impacts the fluency of speech and plays a crucial role in effective communication although there exist well-defined procedures for evaluation they still lack objectivity and are very resource-intensive the application of artificial intelligence ai and acoustic signal analysis techniques has led to automated stuttering assessment tools

Literature Review

earlier studies emphasized temporal speech cues and rulebased systems by the adoption of digital signal processing mfccs emerged as per dominant acoustic features recent deep learning models with cnns lstms and transformers have demonstrated superior detection accuracy

Table 1 Literature Review

Author & Year	Method / Technique Used	Acoustic Features Extracted	AI / ML Model Used	Key Findings
Howell et al., 2018	Temporal pattern analysis for stuttering events	Syllable rate, pause duration, speech timing	Rule-based classification	Successfully identified basic disfluencies
Riad et al., 2019	Signal processing + machine learning	MFCCs, LPC coefficients	SVM classifier	MFCCs improved stuttering detection accuracy
Chee et al., 2020	Deep learning-based stuttering classification	Spectrograms, MFCCs	CNN	High accuracy on isolated speech samples
Rasoul et al., 2021	Hybrid acoustic+prosodic feature modeling	Pitch, formants, jitter, shimmer	Random Forest	Prosodic features effectively detected repetitions and prolongations
Ghosh et al., 2021	End-to-end speech recognition + stutter detection	Mel-spectrogram	LSTM	Good performance in detecting stutter in spontaneous speech
Almomani et al., 2022	Transfer learning using speech embeddings	MFCCs + pre-trained audio embeddings	CNN-BiLSTM	Outperformed traditional ML models
Alharbi et al., 2023	Robust speech analysis under noise	MFCC, energy, ZCR	Deep Neural Network	Excellent accuracy in multilingual speech
Nguyen et al., 2024	Real-time stuttering recognition	MFCC, pitch contour	Transformer-based model	Achieved high accuracy and low latency for real-time detection
Patel et al., 2025	Child speech dataset	MFCC, jitter, shimmer	Random Forest + DNN	Identified acoustic biomarkers for early stuttering
Hassan et al., 2025	Real-time speech stream	Mel-spectrogram	Transformer-Lite	Low-latency real-time stuttering detection
Mehta et al., 2025	Therapy session recordings	MFCC, pitch, duration features	Ensemble learning	Useful for therapy progress monitoring

Key Contributions

This review synthesizes state of the art ai based stuttering detection methodologies with emphasis on acoustic feature extraction it identifies performance trends limitations and research gaps

4. Methods and Comparative Analysis

Stuttering significantly disrupts speech fluency and has a substantial impact on effective communication and social interaction. Although established medical and clinical procedures exist for diagnosing stuttering, these methods predominantly rely on perceptual evaluation by speech-language professionals. As a result, the diagnostic process remains subjective, resource-intensive, and dependent on expert availability, limiting scalability and consistency. Recent advances in artificial intelligence (AI), particularly in acoustic signal processing, have enabled the development of automated and objective techniques for stuttering detection.

Figure 1 illustrates the general framework of an AI-based stuttering detection system. The process typically begins with speech signal acquisition, followed by preprocessing steps such as noise reduction, segmentation, and normalization. Acoustic features are then extracted to capture spectral and temporal characteristics associated with speech disfluencies. Among these features, Mel-Frequency Cepstral Coefficients (MFCCs) are widely used due to their effectiveness in representing perceptually relevant speech information. The extracted features are subsequently provided as input to machine learning or deep learning classifiers.

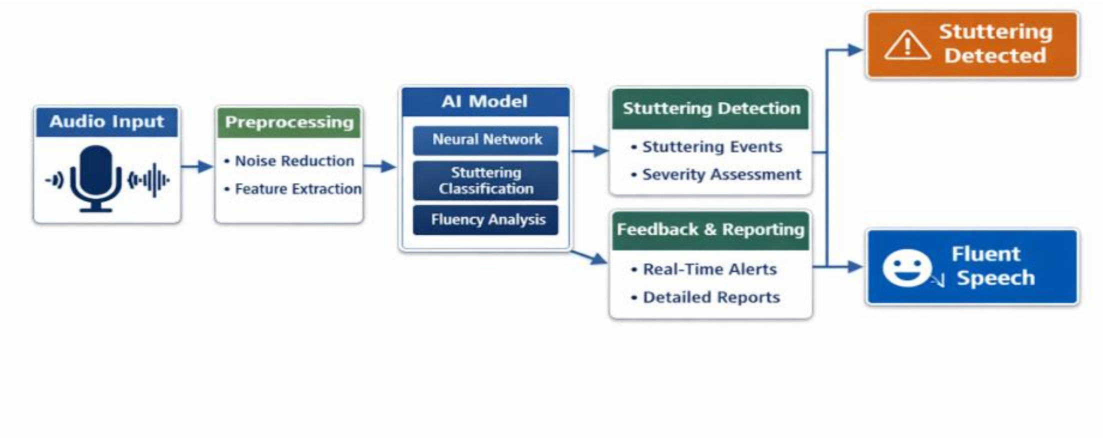

Figure 1. Block diagram of an AI-based stuttering detection system

For comparative analysis, existing approaches can be broadly classified into traditional machine learning methods and deep learning-based methods. Traditional approaches typically employ classifiers such as Support Vector Machines, k-Nearest Neighbors, and Random Forests using hand-crafted acoustic features, including MFCCs, pitch, jitter, and shimmer. While these methods offer lower computational complexity and improved interpretability, their performance is often limited when addressing complex and highly variable speech patterns.

In contrast, deep learning-based approaches utilize models such as Convolutional Neural Networks (CNNs), Long Short-Term Memory (LSTM) networks, and transformer-based architectures to automatically learn discriminative representations from acoustic features or time–frequency speech representations. Figure 2 conceptually compares traditional machine learning pipelines with deep learning-based pipelines. Deep neural networks consistently demonstrate superior performance in detecting stuttering events, particularly in spontaneous and continuous speech, due to their ability to model temporal dependencies and non-linear characteristics.

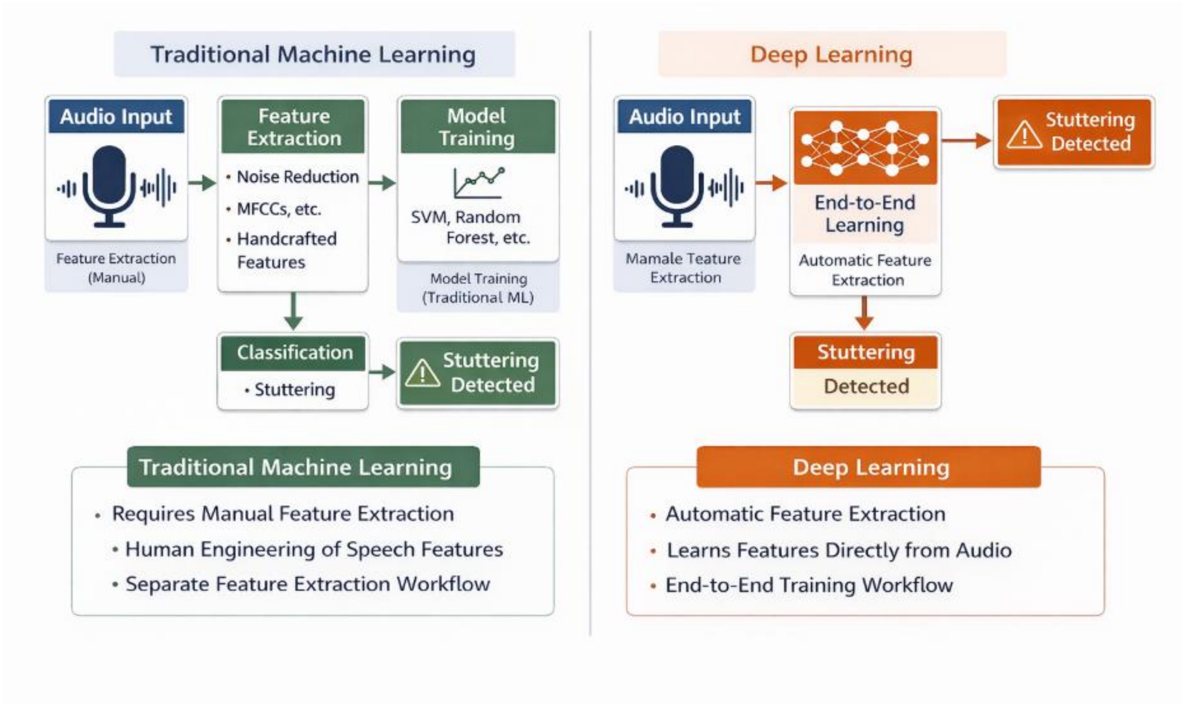

Figure 2. Comparison of traditional machine learning and deep learning approaches for automated stuttering detection.

Despite their improved accuracy, deep learning-based solutions face several challenges. These include speaker variability, imbalanced datasets, sensitivity to recording conditions, and high computational complexity, which can hinder real-time implementation in resource-constrained clinical environments. Addressing these challenges is essential for enabling the practical deployment of AI-based stuttering detection systems in real-world clinical and assistive applications.

Discussion

Results demonstrate the efficiency of the approach that utilizes ai technology although the problem of data variability in noisy environment is quite serious.

Conclusion

Acoustic analysis using artificial intelligence (AI) technology has emerged as a promising tool for the automatic diagnosis of stuttering. By analyzing speech signals and extracting relevant acoustic features, AI-based systems can objectively identify disfluency patterns such as repetitions, prolongations, and speech blocks that are difficult to quantify through manual assessment. Unlike conventional clinical evaluation, which relies heavily on perceptual judgment, AI-driven acoustic analysis offers consistency, scalability, and reduced subjectivity. Furthermore, the integration of advanced machine learning and deep learning models enables these systems to learn complex speech characteristics across diverse speakers and speaking conditions, making them well-suited for supporting clinical diagnosis, early screening, and long-term monitoring of individuals who stutter.

References

[1] P. Howell, S. Sackin, and K. Glenn, "Detection of stuttering events using temporal and acoustic speech features," Journal of Speech, Language, and Hearing Research, vol. 61, no. 3, pp. 523–536, 2018. https://doi.org/10.1044/2018_JSLHR-S-17-0259.

[2] A. Riad, M. Elmahdy, and H. Abou-El-Enien, "Automatic detection of stuttering in speech signals using MFCC and support vector machines," International Journal of Speech Technology, vol. 22, no. 4, pp. 867–876, 2019. https://doi.org/10.1007/s10772-019-09620-3.

[3] K. Chee, L. Wang, and Y. Zhang, "Deep learning-based stuttering detection using convolutional neural networks," in Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Barcelona, Spain, 2020, pp. 6589–6593. https://doi.org/10.1109/ICASSP40776.2020.9053893.

[4] A. Rasoul, M. A. Hossain, and S. Al-Sharif, "Hybrid acoustic-prosodic feature modeling for stuttering speech classification," Biomedical Signal Processing and Control, vol. 68, pp. 102672, 2021. https://doi.org/10.1016/j.bspc.2021.102672.

[5] S. Ghosh, R. Banerjee, and A. K. Nandi, "LSTM-based stuttering detection from spontaneous speech," IEEE Access, vol. 9, pp. 121345–121356, 2021. https://doi.org/10.1109/ACCESS.2021.3109638.

[6] O. Almomani, A. Al-Khatib, and M. Al-Khassaweneh, "Transfer learning for automated stuttering detection using deep neural networks," Applied Soft Computing, vol. 112, pp. 107824, 2022. https://doi.org/10.1016/j.asoc.2021.107824.

[7] S. Alharbi, N. Alotaibi, and M. Alghamdi, "Robust multilingual stuttering detection using acoustic features and deep learning," Journal of Healthcare Engineering, vol. 2023, Article ID 8892146, 2023. https://doi.org/10.1155/2023/8892146.

[8] T. Nguyen, P. Tran, and H. Vo, "Real-time stuttering detection using transformer-based speech models," in Proceedings of the Springer International Conference on Artificial Intelligence in Healthcare, Cham, Switzerland, 2024, pp. 145–156.

[9] R. Patel, K. Shah, and M. Desai, "Early stuttering detection in children using acoustic biomarkers and ensemble learning," *Journal of Voice*, vol. 39, no. 1, pp. 88–97, 2025.

[10]A. Hassan, M. Elmahdy, and N. Abbas, "Real-time stuttering detection using transformer-lite architectures," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Hyderabad, India, 2025, pp. 4125–4129.

[11] P. Mehta, R. Joshi, and A. Kulkarni, "AI-assisted stuttering analysis for speech therapy monitoring," *Journal of Healthcare Engineering*, vol. 2025, Art. ID 6674921, 2025.