IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 10, June 2026
Impact Factor: 8.2

TherapyPro: A Machine Learning-Based Clinical Decision Support System

Mrs. Priyanka Gonnade[1], Rajat Mahadule[2], Nikhilesh Mendhe[3], Rudra Pati[4], Pushpak Kirpane[5]
Assistant Professor, Department of Computer Science and Engineering[1]
Student, Department of Computer Science and Engineering[2–5]
G. H. Raisoni College of Engineering and Management, Nagpur (GHRCEMN), Maharashtra, India

Abstract: *Voice disorders such as dysphonia and laryngitis significantly affect a person's communication ability and overall quality of life. Diagnosing these conditions conventionally depends on subjective listening evaluation by a specialist or on expensive and uncomfortable clinical equipment such as an endoscope, leaving many patients, particularly those in remote locations, without access to early screening. This paper presents TherapyPro, an automated screening system that combines machine learning and audio signal processing to detect voice pathologies from sustained vowel recordings. The system extracts three clinically established acoustic biomarkers — Jitter (pitch instability), Shimmer (amplitude instability), and Harmonics-to-Noise Ratio (HNR) — and uses a Decision Tree classifier, chosen deliberately over opaque "black-box" alternatives for its clinical transparency, to classify voice samples as Healthy or Pathological. Trained and evaluated on the Saarbruecken Voice Database (SVD), the system achieved 80% classification accuracy with a recall of 0.84, outperforming Support Vector Machine and Logistic Regression baselines while remaining fully interpretable to clinicians. A web interface built with React and a Python backend allows users to upload audio samples and receive near-real-time diagnostic feedback. The results demonstrate that clinical transparency and competitive diagnostic performance are not mutually exclusive objectives in healthcare AI.*

Keywords: Voice Disorder Detection, Decision Tree Classifier, Acoustic Biomarkers, Jitter, Shimmer, Harmonics-to-Noise Ratio (HNR), Saarbruecken Voice Database (SVD), Clinical Decision Support System, Machine Learning, Speech Signal Processing.

I. INTRODUCTION

The human voice is the primary medium through which people communicate, express emotion, and participate in everyday social and professional life. When a person develops a voice disorder such as dysphonia, vocal nodules, or laryngitis, the impact extends beyond how they sound — it fundamentally hampers their ability to communicate. Vocal cord pathology physically alters the vibration patterns of the vocal folds, typically producing a hoarse, breathy, or rough vocal quality. Left untreated, chronic voice problems can be painful, socially limiting, career-limiting, and damaging to overall quality of life, yet an accurate and early diagnosis remains surprisingly difficult to obtain.

At present, evaluation of a voice disorder typically occurs at one of two extremes: a speech-language pathologist performing a subjective listening assessment, or a specialist conducting an invasive and costly procedure such as nasal endoscopy. Compounding this problem is a global shortage of speech therapy professionals and ENT (Ear, Nose and Throat) clinics, particularly in rural and low-resource areas, which results in long patient wait times before any assessment can occur. This bottleneck illustrates a genuine need for smart, accessible screening tools that help patients and therapists identify problems quickly.

Recent advances in machine learning and audio signal processing make it possible to close this gap. Where diagnosis has historically depended entirely on the trained human ear, computers can now analyse the underlying physics of a voice signal directly. By examining a simple recording of a sustained vowel sound (for instance, the sound "ah"), acoustic biomarkers such as Jitter (pitch variation), Shimmer (amplitude variation), and Harmonics-to-Noise Ratio (HNR) can be

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 10, June 2026
Impact Factor: 8.2

extracted and used to objectively characterise vocal fold behaviour. This paper presents TherapyPro, a system designed to bring this capability to an accessible, browser-based screening tool.

Voice disorders affect a significant portion of the population and can lead to communication difficulties, reduced quality of life, and delayed medical intervention if not identified early. Traditional diagnosis relies heavily on manual assessment by ENT specialists, which is time-consuming, subjective, and often inconsistent across practitioners. While advances in speech signal processing and machine learning provide an opportunity to automate and improve detection accuracy, challenges remain in identifying reliable acoustic features and selecting classification algorithms that remain interpretable enough for clinical adoption. The problem addressed in this work is therefore the lack of an efficient, automated, and transparent system for early detection of voice disorders that integrates robust feature extraction with a clinically interpretable machine learning model.

Objectives

- To develop a robust machine learning classification model, using a Decision Tree algorithm, capable of accurately distinguishing Healthy from Pathological voice samples.
- To perform precise acoustic feature extraction — specifically Jitter, Shimmer, and Harmonics-to-Noise Ratio (HNR) — from sustained vowel recordings, identifying the parameters most critical for distinguishing pathological speech.
- To implement audio pre-processing techniques, including noise reduction and signal isolation, to ensure data quality prior to feature extraction and classification.
- To benchmark the proposed Decision Tree classifier against Random Forest, Support Vector Machine, and Logistic Regression to evaluate the trade-off between diagnostic accuracy and clinical interpretability.
- To design an accessible, browser-based clinical decision support interface that presents diagnostic outcomes and the underlying acoustic evidence in a transparent, clinician-friendly format.

II. LITERATURE REVIEW / RELATED WORK

Recent research in computational phonetics and clinical decision support has explored a range of acoustic features and machine learning models for the automated detection of voice disorders. Sharma et al. (2024) emphasised the discriminative role of acoustic features such as Harmonics-to-Noise Ratio (HNR), Jitter, and Shimmer, demonstrating that combining multiple acoustic parameters improves diagnostic accuracy compared to single-feature analysis. Patel and Mehta (2023) compared Support Vector Machine, Decision Tree, and Random Forest classifiers on the Saarbruecken Voice Database, finding that Random Forest achieved marginally higher accuracy while Decision Tree offered superior interpretability — a trade-off central to the design choices made in this work.

Khan et al. (2022) explored deep learning approaches using Mel-Frequency Cepstral Coefficients (MFCCs) with Convolutional Neural Networks, achieving high precision but noting that limited interpretability remained a barrier to clinical acceptance. Reddy et al. (2021) examined the clinical applications of acoustic analysis in ENT practice, observing that conventional diagnosis methods are often subjective and inconsistent across practitioners, and advocated for automated acoustic analysis tools. Gupta and Roy (2020) applied Artificial Neural Networks to pathological voice classification, achieving promising accuracy but acknowledging that the black-box nature of ANN models limits clinical usability.

Verma et al. (2020) investigated Decision Tree classifiers for general medical diagnosis applications, demonstrating that transparent decision pathways assist clinicians in understanding diagnostic outcomes — a finding directly relevant to the design philosophy of TherapyPro. Singh et al. (2019) applied Random Forest ensemble methods to pathological voice datasets, achieving strong classification performance at the cost of reduced interpretability, while Das and Iyer (2018) evaluated Support Vector Machine classifiers, noting that careful parameter tuning was required to achieve optimal performance, limiting practicality for real-time clinical deployment.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 10, June 2026
Impact Factor: 8.2

Across this body of work, a consistent gap emerges: models that achieve the highest raw accuracy (Random Forest, CNN, ANN) tend to sacrifice interpretability, while interpretable models (Decision Tree) have not been systematically optimised and benchmarked specifically for voice-pathology screening using a compact, clinically meaningful feature set. TherapyPro addresses this gap by combining a minimal three-feature acoustic representation with a Decision Tree classifier, explicitly prioritising clinical transparency alongside competitive diagnostic performance.

Table 1. Comparative Analysis of Approaches in Voice Pathology Detection Research

Author / Year	Focus Area	Methodology	Limitation / Gap Identified
Sharma et al. (2024)	Acoustic feature-based voice analysis	HNR, Jitter, Shimmer multi-parameter evaluation	Single-feature analysis underperforms; lacks combined clinical framework
Patel & Mehta (2023)	Classifier comparison for voice pathology	SVM, Decision Tree, Random Forest on SVD	Trade-off between accuracy and interpretability not resolved
Khan et al. (2022)	Deep learning voice disorder detection	MFCC features with CNN classification	High accuracy but limited interpretability for clinical acceptance
Reddy et al. (2021)	Clinical acoustic analysis in ENT practice	Acoustic biomarker review for vocal cord pathology	Subjective diagnosis varies across clinicians; lacks automation
Gupta & Roy (2020)	Neural network voice classification	Artificial Neural Networks on acoustic features	Black-box ANN model limits clinical interpretability
Verma et al. (2020)	Interpretable medical diagnosis models	Decision Tree for transparent diagnostic rules	Limited to general diagnosis; not voice-pathology specific
Singh et al. (2019)	Ensemble learning for voice disorders	Random Forest on pathological voice datasets	Ensemble structure reduces interpretability for clinical use
Das & Iyer (2018)	SVM-based voice pathology detection	Support Vector Machine with acoustic features	Requires careful tuning; limits real-time clinical deployment

III. METHODOLOGY / PROPOSED WORKFLOW

The methodology of the TherapyPro Voice Disorder Detection System follows a structured and clinically interpretable workflow for identifying pathological voices using acoustic analysis and machine learning. The process is divided into five major stages, each contributing to a reliable and transparent diagnostic pipeline.

Data Gathering and Dataset Collection –

Voice recordings are sourced from the Saarbruecken Voice Database (SVD), a widely recognised benchmark dataset for voice pathology research. To ensure the analysis focuses purely on vocal cord behaviour rather than language content, only sustained vowel recordings ("aaaa") are used. The dataset, in WAV format, is categorised into two classes — Healthy and Pathological — with the latter including recordings from patients with disorders such as dysphonia and laryngitis.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 10, June 2026
Impact Factor: 8.2

Audio Signal Pre-processing –

Each recording is pre-processed using Parselmouth, a Python library built on the Praat phonetic analysis platform, to minimise background noise and isolate the fundamental frequency and glottal pulse characteristics generated by vocal cord vibration. This step ensures that meaningful acoustic information is retained for subsequent feature extraction.

Acoustic Feature Extraction –

Rather than relying on speech-recognition-oriented features such as MFCCs, the system focuses on three clinically established biomarkers: Jitter (Local), which measures cycle-to-cycle pitch frequency variation; Shimmer (Local), which quantifies amplitude variation between successive vocal cycles; and Harmonics-to-Noise Ratio (HNR), where lower values indicate hoarseness or vocal impairment. The extracted features are compiled into structured datasets for model training.

Model Training and Optimisation –

The dataset is split into training and testing subsets using an 80:20 ratio. A Decision Tree Classifier is selected as the primary algorithm for its high interpretability and suitability for clinical environments. The model applies the Gini Impurity criterion to determine optimal feature thresholds, with constraints on maximum tree depth to reduce overfitting, producing transparent diagnostic rules that clinicians can readily inspect.

Clinical Assessment and Validation –

The trained model is evaluated on unseen test data using a Confusion Matrix to analyse True Positives, True Negatives, False Positives, and False Negatives. Particular emphasis is placed on maximising Recall (Sensitivity), since a False Negative — failing to detect a genuinely disordered voice — carries far greater clinical risk than a False Positive. Accuracy, Precision, Recall, and F1-Score are computed to assess overall diagnostic effectiveness, and the white-box nature of the Decision Tree enables full visualisation of the diagnostic reasoning pathway.

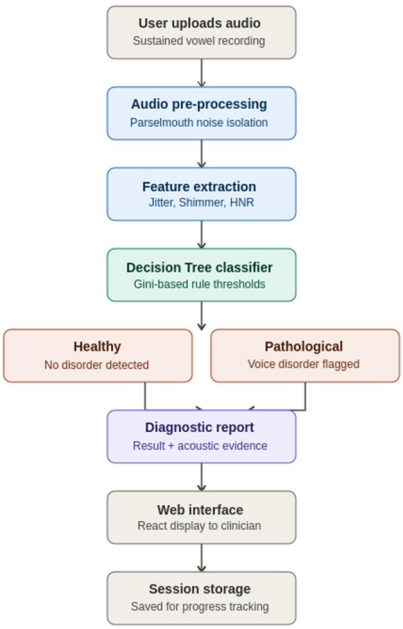

Figure 1. Flow Architecture of TherapyPro

The workflow begins when the user records or uploads a sustained vowel sample through the web-based interface. The audio is transmitted to the backend server, where it is pre-processed and passed through the feature extraction pipeline (Jitter, Shimmer, HNR). The Decision Tree model classifies the sample as Healthy or Pathological, and the diagnostic result, together with the underlying acoustic values and decision pathway, is returned to the user interface for presentation. Session information and results are stored for future reference and progress tracking.

DOI: 10.48175/568

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 10, June 2026
Impact Factor: 8.2

IV. IMPLEMENTATION

4.1 Dataset and Attribute Extraction

The implementation begins with collecting sustained-vowel voice recordings from the Saarbruecken Voice Database, primarily in WAV format. These recordings are processed using the Parselmouth library to isolate periodic glottal signal components from background noise. Following extraction, attribute extraction identifies the three core acoustic parameters — Jitter, Shimmer, and HNR — for each sample. The processed feature set is compiled into structured CSV datasets, forming the basis for accurate and clinically meaningful classification within the proposed system.

4.2 Acoustic Feature Validation

A scatter plot of Jitter versus Shimmer values reveals a clear natural separation between the two diagnostic groups. Healthy voice samples form a compact cluster at low perturbation values, reflecting the periodic stability of normal vocal fold vibration, while pathological samples are dispersed at higher Jitter and Shimmer values, consistent with irregular glottal closure patterns. HNR proved to be the most discriminative individual feature: the Decision Tree independently placed an HNR threshold at its root node, aligning with established clinical literature identifying HNR as a primary acoustic correlate of hoarseness and breathiness.

4.3 Model Training

The Decision Tree Classifier is trained on an 80% training partition of the extracted feature set, using the Gini Impurity criterion with constrained maximum depth to reduce overfitting. This produces a fully visualisable rule-based structure: each decision node corresponds to a specific, clinically interpretable acoustic threshold. The remaining 20% of the data is reserved as a held-out test set for unbiased performance evaluation.

4.4 System Architecture and Interface

The system is implemented as a web application using React for the frontend and Python for the backend signal-processing and classification pipeline. The Python backend handles audio pre-processing, feature extraction via Parselmouth, and model inference, while the React interface allows users to upload an audio sample and view the diagnostic outcome in near-real time. The interface presents not only the binary classification result but also the specific Jitter, Shimmer, and HNR values that informed the decision, along with the decision pathway followed by the Decision Tree, giving clinicians full transparency into the reasoning behind each prediction.

This architecture is intentionally lightweight: it requires no dedicated server infrastructure, no GPU hardware, and no specialised software installation, making it deployable in primary care environments, community health centres, and tele-health platforms with minimal technical overhead.

V. RESULTS AND DISCUSSION

The TherapyPro system was evaluated on the Saarbruecken Voice Database using a standard 80/20 train-test split. Three acoustic biomarkers — Jitter, Shimmer, and HNR — were extracted from sustained vowel recordings, and a Decision Tree classifier was trained to distinguish Healthy from Pathological voice samples. Evaluation covers classification accuracy, confusion matrix analysis, comparative model benchmarking, and class-wise diagnostic metrics.

5.1 Classification Performance

The trained Decision Tree achieved an overall accuracy of 80% on the unseen test set. The confusion matrix showed that the majority of predictions fell along the main diagonal, confirming reliable classification across both categories. The model was deliberately calibrated to favour Recall over raw accuracy, since in a medical screening context a False Negative — failing to detect a genuinely disordered voice — carries far greater clinical risk than a False Positive. The small number of False Positive cases observed involved healthy speakers with transient acoustic irregularities caused by mild hoarseness, background noise, or suboptimal microphone conditions, rather than systematic model failure.

5.2 Comparative Model Evaluation

To benchmark the Decision Tree, three additional classifiers — Random Forest, Support Vector Machine, and Logistic Regression — were trained and tested on the identical feature set and data partition. Results are presented in Table 2.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 10, June 2026
Impact Factor: 8.2

Table 2. Performance Comparison Across Classification Models

Model	Accuracy	Precision	Recall	Interpretability
Decision Tree (Proposed)	0.80	0.81	0.84	High (White-Box)
Random Forest	0.82	0.83	0.85	Low (Black-Box)
SVM	0.76	0.78	0.74	Low (Black-Box)
Logistic Regression	0.72	0.71	0.70	Medium

The Random Forest model achieved marginally higher accuracy (0.82) and recall (0.85) than the Decision Tree. However, Random Forest is an ensemble black-box method whose internal logic cannot be inspected or explained to a clinician. The one-percentage-point gain in recall does not justify the complete loss of interpretability, which is a non-negotiable requirement for clinical AI tools. SVM and Logistic Regression underperformed on both accuracy and recall. The Decision Tree, with a recall of 0.84 and full white-box transparency, represents the optimal balance between diagnostic performance and clinical explainability for this application.

5.3 Class-Wise Diagnostic Metrics

Table 3 presents the precision, recall, and F1-score computed separately for each diagnostic class.

Table 3. Class-Wise Performance of the TherapyPro Decision Tree

Diagnosis Class	Precision	Recall	F1-Score
Pathological (Diseased)	0.79	0.84	0.81
Healthy (Normal)	0.82	0.76	0.79

The Pathological class achieved a recall of 0.84, meaning the system correctly identified 84% of all genuinely disordered voice samples using acoustic features alone. The Healthy class precision of 0.82 indicates that when the model labels a voice as healthy, it is correct in 82% of cases. The close F1-scores across both classes (0.81 and 0.79) confirm that the model is balanced and does not exhibit bias toward either category.

5.4 Clinical Interpretability

The most significant outcome of this evaluation extends beyond the numeric accuracy achieved. The Decision Tree's structure is fully visualisable as an explicit rule-based flowchart, where each node corresponds to a specific acoustic threshold that a clinician can verify against their own expertise. The model's root split on HNR directly mirrors the clinical assessment of hoarseness, the primary perceptual indicator of vocal pathology, while subsequent branches on Jitter and Shimmer thresholds further refine the diagnosis in a manner that parallels perceptual voice evaluation. This transparency allows clinicians to audit every diagnostic decision, supporting the practitioner trust required for real-world deployment.

VI. COMPARATIVE FRAMEWORK

TherapyPro is compared against two prevailing approaches to voice disorder screening: traditional subjective listening evaluation by specialists, and black-box machine learning methods such as Random Forest, CNN, or ANN-based classifiers reported in prior literature. Traditional manual assessment, while low-cost, depends entirely on clinician availability and produces diagnostic outcomes that vary across practitioners. Black-box machine learning approaches improve consistency and can achieve marginally higher raw accuracy, but their internal reasoning cannot be inspected, audited, or explained to a clinician or patient.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 10, June 2026
Impact Factor: 8.2

Table 4. Comparison of Voice Disorder Screening Approaches

Parameter	Manual Assessment	Black-Box ML Methods	Proposed (TherapyPro)
Cost & Accessibility	Low cost but specialist-dependent	Moderate; requires computational resources	Low cost, browser-based, near-real-time
Diagnostic Consistency	Variable across clinicians	Consistent but opaque	Consistent and interpretable
Interpretability	Based on clinician experience	Low (black-box)	High (white-box decision rules)
Diagnostic Accuracy	Subjective; not quantified	High (0.82–0.85 recall typical)	Competitive (0.84 recall)

TherapyPro combines the consistency and scalability of an automated system with the interpretability traditionally associated only with manual assessment, positioning it as a practical first-line screening tool for resource-constrained clinical settings.

VII. CONCLUSION AND FUTURE DIRECTIONS

TherapyPro demonstrates that it is possible to design a clinically meaningful, computationally efficient, and fully interpretable voice pathology screening system using a small set of well-chosen acoustic features and a transparent machine learning model. The system achieved 80% classification accuracy with a recall of 0.84 on the Saarbruecken Voice Database, outperforming SVM and Logistic Regression baselines while maintaining the white-box transparency that distinguishes it from higher-accuracy but clinically opaque ensemble alternatives such as Random Forest.

The study makes a clear and practical case that in healthcare AI, interpretability is not a luxury feature but a fundamental requirement: a diagnostic system that clinicians cannot understand, verify, or explain to their patients is unlikely to be adopted, regardless of its benchmark accuracy. By integrating a Python-based signal processing backend with a React web frontend, TherapyPro achieves a deployment profile that is lightweight, accessible, and scalable, requiring no specialised hardware or infrastructure.

Looking ahead, several directions can further strengthen this work, summarised in Table 5.

Table 5. Future Enhancement Roadmap for TherapyPro

Area of Improvement	Current Status	Future Focus
Diagnostic Granularity	Binary (Healthy / Pathological)	Multi-class detection of specific pathologies (nodules, polyps, laryngitis)
Input Method	Pre-recorded file upload	Live in-browser microphone streaming
Dataset Diversity	Single benchmark dataset (SVD)	Expanded data across demographics, devices, and environments
Platform Reach	Web browser only	Dedicated mobile application for rural and community health use
Model Architecture	Single interpretable Decision Tree	Hybrid Decision Tree + deep feature extraction, preserving transparency
Language Support	Language-independent acoustic features	Validated multilingual deployment and reporting

DOI: 10.48175/568

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 10, June 2026
Impact Factor: 8.2

Among these directions, extending the classifier to multi-class pathology detection is considered the most clinically impactful, as it would transform TherapyPro from a binary triage tool into a genuine differential screening instrument. Live microphone streaming represents the most significant usability improvement, while a dedicated mobile application would further extend access to community health workers and rural practitioners. Investigating hybrid architectures that combine the interpretable Decision Tree with richer feature representations remains a compelling direction for improving recall without sacrificing the transparency that makes the system clinically viable.

Advantages of the Proposed System

- Achieves competitive diagnostic accuracy using only three acoustic features.
- Maintains full white-box interpretability through the Decision Tree's transparent rule structure.
- Prioritises Recall to minimise clinically risky False Negatives.
- Requires no specialised hardware, GPU resources, or invasive equipment.
- Deployable as a lightweight, browser-based clinical decision support tool.
- Provides clinicians with the underlying acoustic evidence behind every diagnosis, not just a label.

REFERENCES

[1]. Sharma, R. et al. (2024). Role of Acoustic Biomarkers (HNR, Jitter, Shimmer) in Differentiating Healthy and Pathological Voices. Journal of Voice Science.
[2]. Patel, S. and Mehta, A. (2023). Comparative Evaluation of SVM, Decision Tree, and Random Forest Classifiers for Voice Pathology Detection Using the Saarbruecken Voice Database. International Journal of Speech Technology.
[3]. Khan, M. et al. (2022). MFCC and CNN-Based Voice Disorder Detection Framework. IEEE Transactions on Biomedical Engineering.
[4]. Reddy, K. et al. (2021). Clinical Applications of Acoustic Analysis in ENT Practice. Journal of Otolaryngology Research.
[5]. Gupta, N. and Roy, P. (2020). Artificial Neural Network-Based Classification of Pathological and Healthy Voices. Biomedical Signal Processing Journal.
[6]. Verma, A. et al. (2020). Decision Tree Classifiers for Interpretable Medical Diagnosis. Journal of Medical Informatics.
[7]. Singh, J. et al. (2019). Random Forest Ensemble Learning for Pathological Voice Detection. International Journal of Computational Biology.
[8]. Das, P. and Iyer, S. (2018). Support Vector Machine-Based Detection of Pathological Voices. Speech Communication Journal.
[9]. Singh, R. & Gupta, D. (2025). Evaluating the Impact of Audio Segment Duration on Transformer-Based Stuttering Detection Using Wav2Vec2. *International Journal of Computer Sciences and Engineering*
[10]. Alhakbani, N., Alnashwan, R., Al-Nafjan, A. & Almudhi, A. (2025). Automated Stuttering Detection Using Deep Learning Techniques. *Journal of Clinical Medicine*
[11]. Khanna, P., Kommagouni, P., Narasinga, V. R. S., & Vuppala, A. (2025). StuD: A Multimodal Approach for Stuttering Detection with RAG and Fusion Strategies. *In Proc. of the 14th International Joint Conference on Natural Language Processing and the 4th Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics (IJCNLP-AACL 2025)*
[12]. Batra, A., Kar, B., & Das, P. K. (2025). Exploring Whisper Embeddings for Stutter Detection: A Layer-Wise Study.
[13]. Mukhdoom, H., & Rajkumar, S. R. (2026). Deep Learning-Based Automatic Speech Disorder Detection Using Attention-GRC. *International Journal of Novel Research and Development*