Automated Nasalance Analysis Using Machine Learning for Speech Disorder Diagnosis

Nikhil N., Sinchana R., Vanitha S.L., Girija Suchitra M.
P.E.S College of Engineering, Mandya

Abstract- Nasalance measurement is a fundamental component of speech evaluation, particularly for identifying resonance-related conditions including hypernasality and hyponasality, both of which are widely encountered in patients with cleft palate, neurological conditions, and structural speech impairments. Traditional clinical assessment approaches largely depend on manual expert evaluation and perceptual interpretation, which can introduce variability, inconsistency, and delays in arriving at diagnostic conclusions. To overcome these shortcomings, this study introduces the NasalanceAI Clinical Analyzer, an AI-powered framework that performs automated nasalance evaluation using machine learning. The system takes clinical and demographic inputs—mean nasalance score, age, gender, and language—and computes a Z-score representing the degree of deviation from typical resonance norms. Speech is then classified as hyponasal, normal, or hypernasal, and further graded by severity as mild, moderate, or severe. The system additionally incorporates automated report generation and patient data management features to streamline clinical operations. Experimental results show that the Logistic Regression classifier achieved an accuracy of 80.5%, providing a strong balance among predictive capability, model transparency, and generalizability for medical applications. By facilitating objective, repeatable, and efficient evaluation, the system lessens reliance on manual processes and strengthens the dependability of speech disorder diagnosis. The system overall highlights the capacity of AI to enable scalable and accessible clinical speech evaluation tools.

Keywords: Nasalance, Speech Analysis, Machine Learning, Clinical AI, Regression Model, Healthcare Automation

I. INTRODUCTION

Effective verbal communication is central to human interaction, and disruptions in speech production can have far-reaching consequences on an individual's social participation, academic development, and overall quality of life. Resonance-related speech conditions, particularly hypernasality and hyponasality, are of notable clinical concern as they alter the proportional distribution of acoustic energy between the oral and nasal passages during articulation. Such conditions are frequently documented in people diagnosed with cleft palate, neurological disorders, structural anomalies, or other forms of speech impairment. Precise evaluation of nasalance is therefore a critical requirement for effective diagnosis, treatment formulation, and ongoing therapeutic monitoring.

Conventional nasalance assessment relies on specialized clinical equipment and expert-based interpretation. While these approaches remain widely used in practice, they are inherently dependent on manual observation and professional judgment, often yielding inconsistent results across different evaluators. Furthermore, such manual approaches tend to be time-intensive, potentially slowing the diagnostic process in high-demand clinical environments. These limitations underscore the necessity for intelligent automated tools capable of delivering objective, dependable, and timely assessments.

This paper addresses the above challenges through the NasalanceAI Clinical Analyzer, an AI-driven framework built to process and interpret speech nasalance data via machine learning techniques. The system ingests key acoustic parameters, including mean nasalance values, alongside patient demographic attributes such as age, gender, and linguistic background. From these inputs, the framework derives a Z-score that reflects the extent of deviation from reference resonance patterns and assigns speech samples to categories such as hyponasal, normal, or hypernasal. The system also quantifies the severity of any identified condition.

A robust and well-structured dataset is essential for constructing reliable predictive models. In the context of clinical speech analysis, raw measurements may be affected by environmental interference, inter-speaker variability, recording inconsistencies, and cross-language differences. Failure to appropriately address these factors can degrade model performance and reduce diagnostic trustworthiness. Accordingly, preprocessing, normalization, and feature engineering constitute indispensable stages of the proposed workflow.

The incorporation of artificial intelligence into clinical speech assessment offers meaningful practical advantages. Automated systems have the potential to support clinicians by alleviating diagnostic burden, enhancing assessment consistency, and accelerating result generation. These tools also enable longitudinal therapy monitoring through repeated evaluations. Moreover, telemedicine and remote healthcare services stand to benefit greatly, as AI-powered speech tools can extend diagnostic access to patients in geographically disadvantaged regions who may otherwise lack specialized clinical support.

Nasalance analysis was selected as the primary domain of this research because it presents both significant clinical relevance and an inherent complexity well-suited to machine learning solutions. Physiological variation, diverse pronunciation styles, and multilingual speech patterns contribute to the difficulty of the classification problem. Given the prevalence of resonance disorders in speech-language therapy and ENT medicine, an intelligent automated nasalance evaluation tool holds strong practical applicability.

The remainder of this paper is organized as follows. Section 2 surveys existing literature on automated speech disorder detection and machine learning methodologies. Section 3 describes the proposed approach in detail, encompassing data preparation, feature design, and model construction. Section 4 presents and discusses the experimental findings. Sections 5 and 6 offer the conclusion and outline directions for future research.

II. LITERATURE

Advancements in artificial intelligence have substantially transformed the landscape of speech disorder evaluation. Traditional diagnostic workflows typically rely on clinician experience, auditory-perceptual judgment, and manual review of recorded speech samples. While these methods remain in widespread clinical use, their subjective nature and inter-rater variability have prompted researchers to explore machine learning and deep learning alternatives capable of delivering more consistent, scalable, and objective outcomes.

Seoyengo Lee, Sunghan Lee, and Goeun Park explored AI-enabled speech monitoring systems for dysarthria detection in consumer devices. Their work demonstrated that speech-based healthcare tools can be integrated into everyday electronics, enabling continuous monitoring and improving accessibility for patients outside clinical settings.

Kanuri N. V. S. H. and Duggiraju S. L. proposed an automated method for detecting hypernasality in cleft palate speech using phase-based acoustic features. Their framework combined vowel segmentation, feature extraction, and Support Vector Machine classification, showing that carefully designed speech features can improve the identification of resonance disorders.

Kothadia K. investigated cross-lingual hypernasality evaluation using Wav2Vec2 representations. The study highlighted the value of transformer-based self-supervised learning models in extracting robust speech representations that generalize across multiple languages. This is particularly relevant for multilingual clinical environments.

Rohini S. M., Aditya P., and Hemant A. P. focused on classifying dysarthria severity through phase-based features of LP residual signals. Their findings demonstrated that machine learning applications in speech pathology extend beyond mere disorder detection to encompass severity grading, which plays an essential role in guiding treatment planning and evaluating therapeutic progress.

Keerthika M., Abinaya N., Santhiya S., Arvindh C. B., Dhanush T., and Nithika M. examined deep learning approaches for dysarthria diagnosis, employing convolutional neural networks with frequency-domain inputs including Mel-frequency cepstral coefficients and spectrograms. Their results demonstrated accuracy improvements over conventional classification methods.

Tuan Nguyen and Corinne Fredouille studied ASR-based Wav2Vec2 systems for automated speech disorder assessment. Their research emphasized the importance of explainable and interpretable AI models in healthcare applications, where transparency is essential for clinician trust and adoption.

Shaik M. S., Mohan B., and Kooli R. presented a review of machine learning methods for dysarthric speech analysis. Their survey covered disorders such as dysarthria and Parkinsonian speech, highlighting the growing demand for automated diagnostic tools that are accurate, non-invasive, and scalable.

Rui F., Yu-Ang C., Yin-Long L., Jia-Hong Y., and Zhen-Hua L. introduced Wav2Nas, a deep learning approach for estimating nasalance directly from raw speech signals. This work minimized dependence on costly specialized hardware, proposing a more broadly accessible pathway for resonance-based speech evaluation.

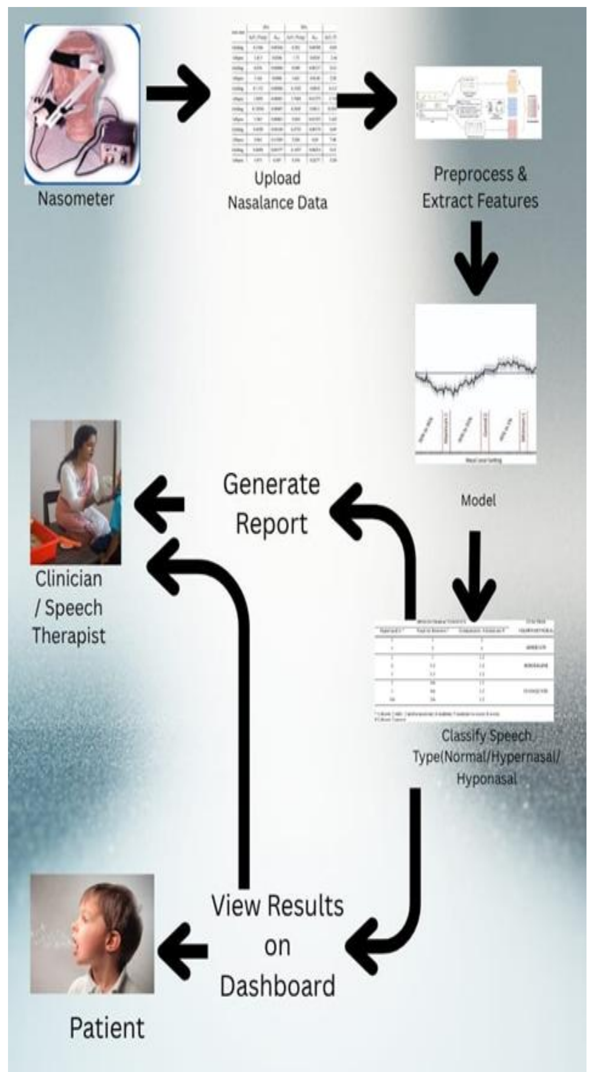
Akhilesh Kumar Dubey and Mahadeva Prasanna S. discussed spectral analysis techniques for objective hypernasality assessment in cleft palate speech. Their study established that signal-based acoustic evaluation can serve as a viable complement—or even alternative—to purely perceptual clinical methods.

Anushree Raj and Pallavi M. O. explored voice disorder detection using machine learning algorithms such as Support Vector Machine, Random Forest, and k-Nearest Neighbors in combination with feature selection strategies. Their findings highlighted that targeted feature selection meaningfully enhances the predictive accuracy of classification models.

In summary, the reviewed body of literature confirms that machine learning, deep learning, and advanced signal processing have substantially elevated the precision, speed, and reliability of speech disorder diagnosis. Nevertheless, persistent challenges include restricted dataset sizes, cross-speaker and cross-language variation, overfitting risks, and the need for clinically interpretable models. The NasalanceAI Clinical Analyzer proposed in this work builds on these findings by unifying statistical analysis, machine learning modeling, and automated reporting into a cohesive, practically deployable solution for nasalance-based speech assessment.

III. METHODOLOGY

The NasalanceAI Clinical Analyzer is conceived as an intelligent clinical decision-support system for automated nasalance evaluation through machine learning. The framework's primary goal is to analyze speech resonance parameters and assign each speech sample to one of three categories—hyponasal, normal, or hypernasal—while simultaneously estimating condition severity. The end-to-end pipeline encompasses data collection, preprocessing, feature construction, model training, prediction, performance evaluation, and structured report generation.



3.1 Data Acquisition

The initial stage of the pipeline focuses on gathering speech measurement data using a nasometer device. Core acoustic parameters—mean nasalance, minimum nasalance, and maximum nasalance—are captured for every subject. Alongside these acoustic metrics, patient contextual information such as age, gender, and spoken language is also recorded. Integrating both acoustic and demographic data yields a more comprehensive input representation for subsequent predictive modeling.

3.2 Data Preprocessing

Raw clinical data frequently contains inconsistencies arising from varying recording environments, individual speaker differences, or missing entries. A dedicated preprocessing stage is therefore applied to enhance data quality ahead of model training. This step covers dataset cleaning, structural organization, and preparation of values in a format appropriate for machine learning algorithms.

All numerical features undergo normalization and standardization so that each variable contributes on a proportional scale during the learning process. Categorical attributes such as gender and language are mapped to numeric representations through appropriate encoding schemes.

nasalance_min	nasalance_max	nasalance_mean	gender_encoded	language_encoded	age
26	37	31.5	0	0	21
27	39	33.0	1	1	22
28	41	34.5	0	0	20
29	43	36.0	1	1	22
30	40	35.0	0	0	22
31	42	36.5	1	1	19
32	44	38.0	0	0	20
33	46	39.5	1	1	20
34	48	41.0	0	0	20
25	35	30.0	1	1	22
26	37	31.5	0	0	21
27	39	33.0	1	1	20
28	41	34.5	0	0	23
29	43	36.0	1	1	22
30	40	35.0	0	0	19
31	42	36.5	1	1	21
32	44	38.0	0	0	23
33	46	39.5	1	1	23
34	48	41.0	0	0	19
25	35	30.0	1	1	21
26	37	31.5	0	0	22
27	39	33.0	1	1	18
28	41	34.5	0	0	21
29	43	36.0	1	1	19
30	40	35.0	0	0	23

3.3 Feature Engineering

To strengthen predictive capacity, a set of informative derived features is constructed from the raw collected inputs. A key engineered feature is the nasalance range, computed as the arithmetic difference between the maximum and minimum nasalance values recorded per subject.

The final set of input features supplied to the model during training comprises the following: Nasalance Range, Gender Encoded (0 = Female, 1 = Male), Language Encoded (0 = Kannada, 1 = Malayalam), and Age. These engineered features enhance the model's ability to capture subtle variation in speech resonance patterns.

3.4 Z-Score Computation

To enable comparison of individual speech samples against language-specific normative reference data, a statistical Z-score transformation is applied. This metric quantifies the number of standard deviations by which a given observation diverges from the expected mean of normal speech.

Z score: $Z = (X - \mu) / SD$

Based on the computed Z-score, speech samples are assigned diagnostic categories as follows:

- If $Z < -1$ → Hyponasal
- If $Z > 1$ → Hypernasal
- Otherwise → Normal

This statistically grounded approach introduces objectivity and reproducibility into the diagnostic classification process.

3.5 Machine Learning Models

The system employs both classification and regression modelling strategies.

Logistic Regression (Classification)

Logistic Regression was chosen as the primary classification algorithm due to its transparency, ease of interpretation, and reliable generalization behavior in healthcare contexts. The model estimates the conditional probability of each speech category and assigns the label with the highest predicted likelihood. Experimental validation yielded an accuracy of 80.5%, confirming its practical applicability in clinical deployment.

$\hat{y} = \text{argmax } P(y|X)$

Comparative Models

Complementary classifiers including Support Vector Machine (SVM) and k-Nearest Neighbors (KNN) were also assessed. Despite some models attaining greater training accuracy, they were ultimately deemed less appropriate for clinical deployment due to their reduced interpretability and susceptibility to overfitting on limited datasets.

Random Forest Regressor

For the task of continuous Z-score estimation, a Random Forest Regressor was employed. By aggregating predictions from an ensemble of decision trees, this model effectively captures nonlinear feature interactions and delivers robust, stable numerical outputs.

$\hat{y} = (1/N) \Sigma \hat{y}_i$

3.6 Severity Classification

The Z-score derived from the regression model is subsequently leveraged to characterize the clinical intensity of the detected speech disorder according to its absolute magnitude:

- Mild if $|Z| < 2$
- Moderate if $2 \leq |Z| < 3$
- Severe if $|Z| \geq 3$

This grading layer equips clinicians with actionable information about the degree of speech abnormality, facilitating more targeted and effective treatment planning.

3.7 Report Generation and Visualization

Upon completion of the prediction process, the system assembles a comprehensive structured clinical report that includes patient demographic details, the input feature values supplied, the predicted Z-score, the derived speech classification, and the estimated severity grade.

All outputs are presented through an intuitive visual dashboard, enabling both clinical staff and patients to readily interpret the diagnostic findings. Reports are also stored for longitudinal tracking and future reference, supporting ongoing patient monitoring.

In summary, the proposed methodology fuses statistical signal processing with machine learning to build a dependable and efficient diagnostic framework for speech resonance disorders. Through full automation of data processing, classification, and clinical documentation, the system substantially reduces manual workload while elevating diagnostic consistency and supporting evidence-based clinical decision-making.

IV. RESULTS AND DISCUSSION

The NasalanceAI Clinical Analyzer was evaluated using real patient input data processed through the trained machine learning pipeline. The integrated system handles data entry, automated prediction, classification, and report generation within a unified workflow. This section presents findings related to feature contribution, system usability, model performance, and observed behavioral characteristics. An analysis of the trained model's feature importance revealed that mean nasalance is by far the most dominant predictor in speech classification decisions. This confirms the primacy of acoustic resonance measurements in detecting nasality-related disorders—a finding consistent with established clinical understanding. Encoded language and gender variables contributed at moderate levels, suggesting that linguistic and demographic attributes exert some influence on resonance patterns. Age, by contrast, exhibited relatively minimal predictive weight in the current dataset. The feature importance analysis additionally contributes to model interpretability by making the decision process transparent, which is a valuable property in clinically sensitive AI applications.

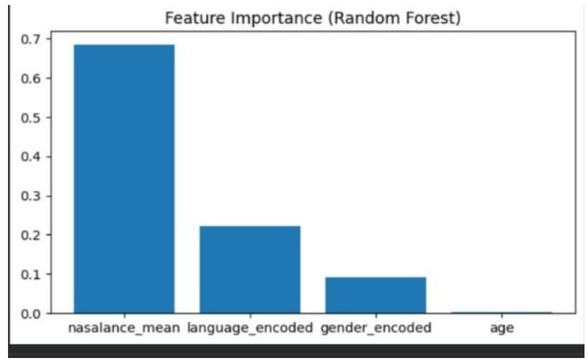


Fig. 4.1: Feature Importance Analysis of Input Variables in Nasalance Prediction Model

The developed application includes a structured and user-friendly interface for entering patient information and nasalance measurements. The input form accepts patient name, age, gender, language, and core nasalance parameters (mean, minimum, and maximum). The interface uses dropdown selectors, numeric spinners, and a well-organized layout to reduce input errors and improve usability. Because no advanced technical expertise is required to operate the system, it can be deployed effectively across a range of healthcare settings and used by clinicians, students, and support staff alike.

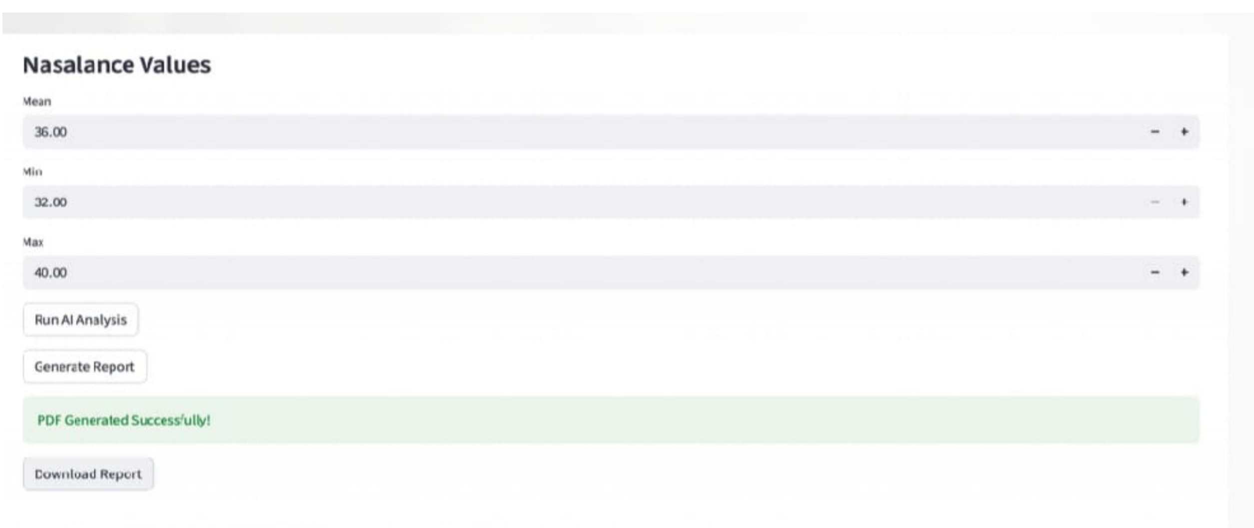
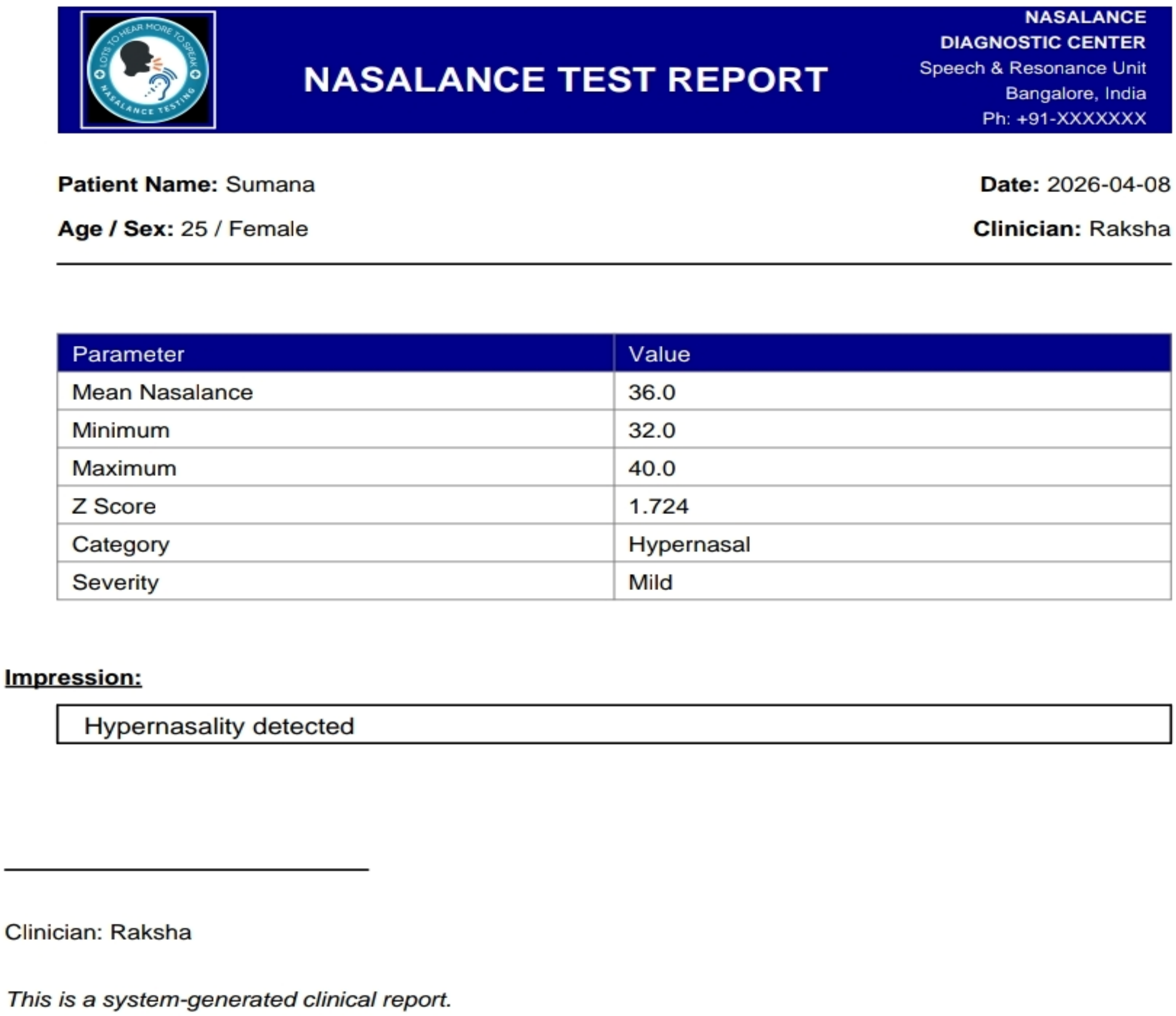


Fig. 4.2: Nasalance Data Input Interface and AI Analysis Execution Screen

After processing the input values, the system automatically predicts a Z-score and classifies the speech sample into one of the diagnostic categories: Normal, Hyponasal, or Hypernasal. In addition, the severity level is identified as Mild, Moderate, or Severe based on the predicted deviation. A structured clinical report is then generated containing patient details, input nasalance values, predicted Z-score, diagnostic category, and severity level. The report can also be exported as a PDF file, which supports documentation, record maintenance, and communication between clinicians and patients. Automated report generation reduces administrative effort and saves time during clinical workflows.


NASALANCE TEST REPORT

NASALANCE DIAGNOSTIC CENTER
Speech & Resonance Unit
Bangalore, India
Ph: +91-XXXXXXX

Patient Name: Sumana
Date: 2026-04-08
Age / Sex: 25 / Female
Clinician: Raksha

Parameter	Value
Mean Nasalance	36.0
Minimum	32.0
Maximum	40.0
Z Score	1.724
Category	Hypernasal
Severity	Mild

Impression:

Hypernasality detected

Clinician: Raksha

This is a system-generated clinical report.

Fig. 4.3: Generated Clinical Report Output with Z-Score, Diagnosis, and Severity Classification

Among the evaluated classification algorithms, Logistic Regression achieved an accuracy of 80.5%. While certain alternative models demonstrated higher performance on the training set, Logistic Regression was selected for deployment on account of its superior interpretability, reduced model complexity, and stronger capacity to generalize to previously unseen cases. In clinical settings, model reliability and consistency tend to take precedence over maximal training accuracy, since overfitted models may fail to perform well in real-world conditions.

A notable observation during evaluation was that many predicted Z-score values clustered within a constrained range, resulting in a predominance of Normal classifications. This pattern suggests that expanding dataset diversity and incorporating richer feature representations—such as direct acoustic signal features including MFCCs, spectral descriptors, or deep learning-derived embeddings—could enable finer discrimination between subtle speech conditions in future iterations.

The experimental outcomes collectively demonstrate that the proposed system effectively integrates machine learning with real-time clinical data to deliver consistent, rapid, and automated speech disorder support. Despite current dataset limitations, the framework exhibits strong potential for expanding diagnostic accuracy, reducing clinician workload, and enabling evidence-based decision support. The study affirms that AI-powered speech evaluation platforms can serve as valuable instruments in contemporary healthcare delivery, particularly in settings where responsiveness, accuracy, and equitable access are critical priorities.

V. CONCLUSION

The NasalanceAI Clinical Analyzer presents a robust and efficient AI-driven approach to automated speech nasalance evaluation. The system effectively processes patient clinical and demographic data to compute Z-scores, categorize speech resonance conditions as normal, hyponasal, or hypernasal, and assign severity ratings that support comprehensive clinical assessment. Feature importance analysis substantiates the dominant role of acoustic nasalance measurements in driving accurate predictions, ensuring that the model's output aligns with established clinical expectations. The framework further enhances clinical utility through automated documentation and patient record management capabilities. While the current implementation reflects certain constraints—particularly regarding dataset diversity and granularity of prediction—the overall system demonstrates substantial promise in minimizing manual effort, reducing evaluator-dependent bias, and improving diagnostic reproducibility. The study affirms that artificial intelligence holds genuine value in modern speech disorder management by providing clinicians with objective, scalable, and dependable decision-support instruments.

VI. FUTURE SCOPE

The NasalanceAI Clinical Analyzer provides a strong foundation for continued advancement in automated speech disorder assessment, and a number of enhancements are envisioned to further elevate its clinical utility. A primary avenue for future development is the direct integration of real-time speech signal processing from nasometer hardware or microphone inputs, eliminating the current reliance on manual data entry. Broadening the training dataset to encompass a wider diversity of patient demographics, spoken languages, and clinical conditions would enhance model generalizability and robustness. Future iterations may also integrate advanced deep learning architectures together with richer acoustic descriptors such as MFCCs, spectrogram-based representations, and transformer-derived speech embeddings to achieve improved classification accuracy. The development of a mobile or web-accessible platform would further facilitate remote diagnostics, continuous patient monitoring, and telemedicine integration—particularly beneficial for underserved populations. Connection with hospital electronic health record systems and cloud infrastructure could additionally streamline data management, improve accessibility, and support deployment at scale. Collectively, the future trajectory of this work points toward a more intelligent, inclusive, and clinically validated platform for speech resonance analysis and healthcare support.

REFERENCES

[1] N. V. S. H. Kanuri and S. L. Duggiraju, "Automatic Detection of Hypernasality in Cleft Palate Speech Using Phase Feature," Proc. IEEE Int. Conf., 2024.

[2] M. S. Shaik, B. Mohan, and R. Kooli, "Machine Learning Assisted Diagnosis of Speech Disorders: A Review of Dysarthric Speech," Proc. IEEE Int. Conf., 2023.

[3] F. Rui, C. Yu-Ang, L. Yin-Long, Y. Jia-Hong, and L. Zhen-Hua, "Wav2Nas: An Exploring Approach to Nasalance Estimation in Speech," Proc. IEEE Int. Conf., 2024.

[4] S. M. Rohini, P. Aditya, and A. P. Hemant, "Dysarthria Severity Classification Using Phase-Based Features of LP Residual," Proc. APSIPA ASC, 2024.

[5] M. Keerthika, N. Abinaya, S. Santhiya, C. B. Arvindh, T. Dhanush, and M. Nithika, "Enhancing Dysarthria Diagnosis with Deep Learning Techniques," Proc. IEEE Int. Conf., 2024.

[6] I. Baumann, D. Wagner, M. Schuster, K. Riedhammer, E. Nöth, and T. Bocklet, "Towards Interpretability of Automatic Phoneme Analysis in Cleft Lip and Palate Speech," Proc. ICASSP, 2024.
[7] A. S. Shakeel and M. Sahidullah, "Overview of Automatic Speech Analysis and Technologies for Neurodegenerative Disorders: Diagnosis and Assistive Applications," IEEE Journal, 2024.
[8] S. Lee, S. Lee, and G. Park, "AI-Enabled Speech Monitoring for Dysarthria Detection in Consumer Electronics," IEEE Int. Conf. on Consumer Electronics (ICCE), 2025.
[9] K. Kothadia, "Cross-lingual Evaluation of Hypernasality Using Wav2Vec2 Features," Proc. IEEE Int. Conf., 2025.
[10] T. Nguyen and C. Fredouille, "Exploring ASR-Based Wav2Vec2 for Automated Speech Disorder Assessment: Insights and Analysis," IEEE Spoken Language Technology Workshop (SLT), 2024.
[11] A. K. Dubey and M. Prasanna S. R., "Spectral Analysis Based Objective Assessment of Hypernasality in Cleft Palate Speech: A Review," Circuits, Systems, and Signal Processing, 2025.
[12] A. Raj and P. M. O., "Voice Disorder Detection and Classification Using Machine Learning Techniques and Feature Selection Methods," IEEE Int. Conf. on Distributed Computing, VLSI, Electrical Circuits and Robotics, 2024.