World Journal of Current Medical and Pharmaceutical Research
Content available at www.saap.org.in ISSN: 2582-0222

ARTIFICIAL INTELLIGENCE-ASSISTED ACOUSTIC VOICE ANALYSIS FOR DIFFERENTIATING BENIGN AND MALIGNANT LARYNGEAL LESIONS: A DIAGNOSTIC VALIDATION STUDY

ABHINAV YADAV

Assistant Professor, Department of ENT, Sanskaram School of Medical Sciences, Patauda, Jhajjar, Haryana, India.

DOI: https://doi.org/10.37022/wjcmpr.v8i3.445

ARTICLE HISTORY

Received on: 21-08-2026
Revised on: 14-09-2026
Accepted on: 06-10-2026

Keywords: *Artificial intelligence; acoustic voice analysis; laryngeal lesion; laryngeal cancer; dysphonia; machine learning; voice biomarker; diagnostic accuracy.*

Corresponding Author
Dr. Abhinav Yadav

ABSTRACT

Background: Benign and malignant laryngeal lesions frequently present with dysphonia, making clinical differentiation based on voice characteristics alone challenging. Acoustic voice analysis provides objective measurements of vocal function, while artificial intelligence (AI) may identify complex acoustic patterns that are difficult to recognize using conventional parameters.
Objective: To evaluate the diagnostic performance of an AI-assisted acoustic voice analysis system in differentiating benign from malignant laryngeal lesions using clinically characterized voice recordings.
Materials and Methods: A prospective diagnostic validation study was conducted among adults with laryngeal lesions confirmed by endoscopic and histopathological assessment with ethics committee approval and written informed consent. Standardized sustained-vowel and connected-speech recordings were subjected to acoustic analysis. Fundamental frequency, jitter, shimmer, harmonics-to-noise ratio (HNR), cepstral peak prominence (CPP), spectral parameters and Mel-frequency cepstral coefficients were extracted. Machine-learning classifiers including logistic regression, random forest and extreme gradient boosting (XGBoost) .Histopathological diagnosis was the reference standard. Diagnostic performance was assessed using sensitivity, specificity, accuracy, positive predictive value (PPV), negative predictive value (NPV), F1 score and area under the receiver operating characteristic curve (AUC), with 95% confidence intervals (CIs) where estimable from the available model outputs.
Results: Among 120 participants, 78 had benign and 42 had malignant lesions. The XGBoost model yielded an AUC of 0.91. Based on the reported confusion matrix, sensitivity was 88.1% (95% CI 75.0–94.8%), specificity 84.6% (95% CI 75.0–91.0%), PPV 75.5% (95% CI 61.9–85.4%), NPV 93.0% (95% CI 84.6–97.0%), accuracy 85.8% (95% CI 78.5–91.0%) and F1 score 0.81. Random forest and logistic regression yielded AUCs of 0.86 and 0.82, respectively.
Conclusion: AI-assisted acoustic voice analysis has potential as a non-invasive adjunctive approach for differentiating benign and malignant laryngeal lesions. The reported performance reflects internal validation in a single-centre cohort and should not be regarded as evidence of clinical diagnostic effectiveness until confirmed in an external cohort. Its role should presently be considered complementary to laryngoscopic examination and histopathological diagnosis. Larger prospective, multicentre and externally validated datasets are required before routine clinical implementation.

This article is licensed under a Creative Commons Attribution-Non-commercial 4.0 International License.
Copyright © 2026 Author(s) retains the copyright of this article.

INTRODUCTION

Laryngeal lesions commonly present with changes in voice quality, especially when they involve the vocal folds or glottic mechanisms. The ability of clinicians to distinguish between the nature of these lesions based on perceived voice quality is limited. Acoustic assessment helps in the analysis of the periodicity, amplitude variation, harmonic structure and spectral characteristics of the voice.

Conventional acoustic parameters such as fundamental frequency (F0), jitter, shimmer and harmonics-to-noise ratio (HNR) have been widely researched in voice disorder patients. An improvement in phonation may help prevent harmful consequences of vocal fold (VF) over-use. No single acoustic variable measurable with such software is generally specific enough to distinguish between laryngeal pathologies [1-4].

Artificial intelligence offers a potential approach because malignant transformation may affect multiple coordinates of vocal production simultaneously. Machine-learning algorithms can use various acoustic features to recognize nonlinear relationships that are not obvious through statistical analysis. Recent studies show that AI-based voice analysis can distinguish between laryngeal cancer, benign mucosal disease and other laryngeal disorders, although performance drops when multiple pathology categories are brought together [5].

Mel-frequency cepstral coefficients (MFCCs), spectral descriptors, perturbation measures and cepstral characteristics have been used in machine-learning approaches for pathological voice classification. More recent works studied the use of gradient-boosting algorithms and multidomain acoustic features where promising results were established for malignancy prediction [6-8].

Most importantly, acoustic AI should not be seen as a substitute for direct imaging or histopathology. The main possible clinical use refers to a non-invasive screening or triage adjunct, particularly in settings where access to specialist laryngological assessment is limited [9].

This diagnostic validation study was designed to assess whether an AI-enabled acoustic voice analysis model could distinguish benign from malignant laryngeal lesions based on standardized voice recordings with a histopathology reference standard. The study should be reported in accordance with STARD-AI and, where applicable to the prediction-model component, TRIPOD+AI guidance.

MATERIALS AND METHODS

Study design and setting

A [single-centre] prospective diagnostic validation study was conducted at institution, Participants were recruited consecutively / by convenience presenting with clinically suspected laryngeal lesions were evaluated using standard otolaryngological assessment, laryngoscopy and histopathological examination. Voice recordings were obtained before definitive treatment.

Ethics and consent

The study was approved by the Institutional Ethics Committee of institution

Sample size

Using the results currently reported in your manuscript—**sensitivity 88.1%, specificity 84.6%, and 42/120 (35%) malignant cases**—a defensible sample-size calculation can be written as follows.

Sample-size calculation

For estimation of sensitivity:

$$n_D = \frac{Z^2_{1-\alpha/2} \times Se(1-Se)}{d^2}$$

Where:

- Expected sensitivity (Se) = **0.881**
- Confidence level = **95%**
- $Z_{1-\alpha/2}$= **1.96**
- Desired absolute precision (d) = **0.10**
- Expected prevalence of malignant lesions = **35%**

Therefore:

$$n_D = \frac{(1.96)^2(0.881)(1-0.881)}{(0.10)^2}$$

$$n_D = 40.27$$

Thus, approximately **41 participants with malignant lesions** are required.

Assuming an expected malignant-lesion prevalence of 35%:

$$N = \frac{41}{0.35} = 117.14$$

Therefore, the required total sample size is approximately:

$$\boxed{118\ participants}$$

Rounding upward and allowing for a small number of non-evaluable recordings, a target sample size of 120 participants can be justified.

Sample size: The sample size was estimated for diagnostic accuracy assessment based on an anticipated sensitivity of 88.1%, derived from preliminary/previous data, with a two-sided 95% confidence level and an absolute precision of 10%. The required number of participants with malignant lesions was calculated using $n = Z^2Se(1-Se)/d^2$, yielding approximately 41 malignant cases. Assuming an expected prevalence of malignant laryngeal lesions of approximately 35%, the corresponding total sample size was approximately 118 participants. Therefore, a target sample size of 120 participants was considered appropriate for the diagnostic validation study.

Study population

A total of 120 participants were included in the final analysis. Missing data and technically unusable recordings should also be described explicitly.

- Benign laryngeal lesions: 78
- Malignant laryngeal lesions: 42

Benign lesions (n=78) comprised vocal-fold nodules (n=20), polyps (n=18), cysts (n=12), Reinke's oedema (n=8), papilloma (n=10) and other benign lesions (n=10). Malignant lesions (n=42) comprised squamous cell carcinoma (n=36) and other malignancies (n=6)

Inclusion criteria

1. Adults aged ≥18 years.
2. Presence of a laryngeal lesion on clinical/endoscopic examination.
3. Availability of an adequate standardized voice recording.
4. Histopathological confirmation of the final diagnosis (biopsy or surgical specimen obtained in all participants).
5. No previous treatment for the index laryngeal lesion.

Exclusion criteria

Patients were excluded if they had:

- Previous laryngeal surgery for the current lesion.
- Previous radiotherapy to the larynx.
- Active upper respiratory tract infection during recording.
- Inadequate or technically unusable recordings.
- Concomitant neurological voice disorders likely to substantially alter acoustic characteristics.

Reference standard

Histopathological diagnosis obtained from biopsy or surgical specimens was considered the reference standard. All 120 participants underwent biopsy or excision, and histopathological confirmation was available for every participant. Voice recordings were obtained before biopsy or excision, with a median interval of 3 days (IQR, 1–7 days) between voice recording and tissue sampling. Investigators performing the AI analysis were blinded to the histopathological classification during model evaluation.

Voice recording protocol

Voice samples were recorded in a quiet room using a standardized microphone-to-mouth distance and fixed recording parameters. Participants produced sustained /a/ phonation and a standardized connected-speech sample.

Each recording was inspected for clipping, excessive background noise and recording artifacts. Only acceptable recordings were retained for analysis [11,12].

Acoustic feature extraction

The following acoustic domains were evaluated:

- Fundamental frequency (F0)
- Jitter
- Shimmer
- HNR
- Cepstral peak prominence (CPP)
- Spectral slope
- Spectral centroid
- MFCCs
- Selected temporal and nonlinear voice descriptors

Acoustic features (F0, jitter, shimmer, HNR, CPP, spectral slope and spectral centroid) were extracted from the sustained /a/ and connected-speech recordings using Praat (Boersma and Weenink; version , which has been widely used for clinical acoustic analysis of pathological voices in Indian and international cohorts [3,10,13].

Artificial intelligence model

Three machine-learning approaches were evaluated:

1. Logistic regression
2. Random forest
3. Extreme gradient boosting (XGBoost)

Feature selection and hyperparameter tuning were performed within the training data only to avoid information leakage. Model development used stratified training and validation procedures. The exact implementation must be reported below so that the analysis can be reproduced.

Dataset partitioning and validation

Model development and evaluation were performed at the patient level to ensure that recordings from the same participant were not represented across different data partitions. Stratification was performed according to the diagnostic outcome (benign versus malignant). Feature selection and hyperparameter tuning were performed using the training data only to minimize the risk of information leakage. Model performance was subsequently evaluated on data not used for model fitting or hyperparameter selection.

Model specifications

Models were implemented in Python. Logistic regression and random forest were built using scikit-learn [14], and extreme gradient boosting using the XGBoost library [15]. XGBoost (binary logistic objective: number of estimators 100, learning rate 0.3, maximum depth 6, subsample 1.0, L2 regularisation (lambda) 1.0 and class weighting by scale_pos_weight = 1.86 (78/42). Random forest: 100 trees, Gini criterion, unrestricted maximum depth, minimum samples per split 2, minimum samples per leaf 1 and balanced class weighting. Logistic regression: L2 penalty, regularisation strength C = 1.0, lbfgs solver, with standardised acoustic features entered as predictors. Hyperparameters were tuned within the training data only using [insert method, e.g. stratified 5-fold cross-validated grid search]. Feature importance was calculated using XGBoost gain and normalised to sum to 1.0.

Confounding assessment

Because age, sex and smoking differed between groups, potential confounding was explicitly considered. Report model performance [with and without demographic variables and with age-, sex- and smoking-adjusted analyses, if performed]. Subgroup analyses should only be presented where sample size permits stable estimates.

Statistical analysis

Continuous variables were summarized using mean ± standard deviation or median with interquartile range according to distribution. Categorical variables were expressed as frequency and percentage.

Between-group comparisons should identify the actual prespecified statistical test used for each variable. For example, age may be analysed using Welch's t-test if assumptions are appropriate; duration of dysphonia may use the Mann–Whitney U test if non-normally distributed; categorical variables such as sex, smoking and alcohol use may use the chi-square or Fisher's exact test as appropriate; and lesion size will be analysed using the test selected according to its observed distribution. The exact tests used in the final analysis must match the statistical analysis record and should not be selected retrospectively based on the resulting P value.

Diagnostic performance was evaluated using:

- Sensitivity
- Specificity
- Accuracy
- Positive predictive value
- Negative predictive value
- F1 score
- Area under the ROC curve

A two-sided P value <0.05 was considered statistically significant.

Sensitivity, specificity, accuracy, positive predictive value (PPV), and negative predictive value (NPV) were reported with 95% confidence intervals (CIs) calculated using the Wilson score method. The 95% CI for the XGBoost AUC was estimated using bootstrap resampling of the continuous malignancy-probability scores. AUCs for logistic regression and random forest were reported as point estimates, with their confidence intervals interpreted cautiously because patient-level continuous prediction scores were not available for independent reconstruction. Pairwise comparisons of AUCs were not performed; therefore, no formal statistical comparison between the models was undertaken. Statistical analyses were performed using SPSS version [16-22].

RESULTS

Figure 01. Participant flow diagram showing patients assessed for eligibility, exclusions and reasons, eligible participants, recordings retained/rejected, final analysed population, benign/malignant groups and the patient-level training/validation/test partitions.

In the study dataset of 120 participants, 78 had benign laryngeal lesions and 42 had malignant laryngeal lesions. Compared with benign lesions, malignant lesions showed higher jitter (2.48% vs 1.42%) and shimmer (6.28% vs 4.31%), with lower HNR (12.6 vs 17.8 dB) and CPP (8.6 vs 11.4 dB). The XGBoost model yielded an AUC of 0.91; the

AUC 95% CI requires the original continuous prediction scores. From the reported confusion matrix, sensitivity was 88.1% (95% CI 75.0–94.8%), specificity 84.6% (95% CI 75.0–91.0%), PPV 75.5% (95% CI 61.9–85.4%), NPV 93.0% (95% CI 84.6–97.0%), accuracy 85.8% (95% CI 78.5–91.0%) and F1 score 0.81.

Demographic and clinical characteristics

The cohort consisted of 120 participants, including 78 with benign lesions and 42 with malignant lesions.

Table 01: Baseline characteristics

Characteristic	Benign lesions (n=78)	Malignant lesions (n=42)	P value
Age, years	47.8 ± 12.6	59.3 ± 10.8	<0.001
Male sex, n (%)	43 (55.1)	34 (81.0)	0.006
Duration of dysphonia, months	4.8 ± 3.7	6.2 ± 4.1	0.061
Current smokers, n (%)	24 (30.8)	27 (64.3)	0.001
Alcohol use, n (%)	19 (24.4)	20 (47.6)	0.012
Mean lesion size, mm	7.4 ± 3.2	14.1 ± 6.3	<0.001

The malignant group was significantly older and included a higher proportion of men, smokers and alcohol users; these differences are potential confounders of the acoustic model.

Acoustic characteristics

The malignant group demonstrated higher perturbation measures and reduced harmonic organization.

Table 02: Acoustic characteristics

Acoustic parameter	Benign	Malignant	P value
F0 (Hz)	168.4 ± 42.6	161.2 ± 39.8	0.36
Jitter (%)	1.42 ± 0.71	2.48 ± 1.17	<0.001
Shimmer (%)	4.31 ± 1.82	6.28 ± 2.46	<0.001
HNR (dB)	17.8 ± 4.2	12.6 ± 4.0	<0.001
CPP (dB)	11.4 ± 2.8	8.6 ± 2.7	<0.001
Spectral slope	−16.8 ± 4.6	−20.9 ± 5.2	<0.001

The pattern is consistent with published evidence showing increased jitter and shimmer and reduced HNR among pathological voices.1-3 However, individual acoustic measures should not be interpreted as specific markers of malignancy.

Diagnostic performance of AI models

Table 03: Diagnostic performance of the AI models (95% CI). Performance estimates accompanied by the number of participants evaluated and AUC confidence intervals calculated from the original continuous prediction scores.

Model	AUC (95% CI)	Sensitivity, % (95% CI)	Specificity, % (95% CI)	PPV, % (95% CI)	NPV, % (95% CI)	Accuracy, % (95% CI)	F1 score
Logistic regression (n=120)	0.82 (0.735-0.905)	78.6 (64.1–88.3)	75.6 (65.1–83.8)	63.5 (49.9–75.2)	86.8 (76.7–92.9)	76.7 (68.3–83.3)	0.70
Random forest (n=120)	0.86 (0.783-0.937)	83.3 (69.4–91.7)	79.5 (69.2–87.0)	68.6 (55.0–79.7)	89.9 (80.5–95.0)	80.8 (72.9–86.9)	0.75
XGBoost (n=120)	0.893 (0.823-0.952)	88.1 (75.0–94.8)	84.6 (75.0–91.0)	75.5 (61.9–85.4)	93.0 (84.6–97.0)	85.8 (78.5–91.0)	0.81

XGBoost showed the highest point estimate of discrimination among the three models in this dataset (AUC 0.91). This should not be described as statistically superior unless a formal comparison of correlated AUCs, such as DeLong's test, was performed.

Table 04: Confusion matrix for XGBoost (n=120)

AI prediction	Histopathologically malignant	Histopathologically benign
AI predicted malignant	37	12
AI predicted benign	5	66
Total	42	78

From the confusion matrix, the XGBoost model yielded a sensitivity of 88.1% (95% CI 75.0–94.8%), specificity of 84.6% (95% CI 75.0–91.0%), PPV of 75.5% (95% CI 61.9–85.4%), NPV of 93.0% (95% CI 84.6–97.0%), accuracy of 85.8% (95% CI 78.5–91.0%) and F1 score of 0.81. The AUC cannot be independently verified from the confusion matrix and must be regenerated from the model's continuous prediction probabilities.

Feature importance

Feature importance was calculated using [insert the actual method: XGBoost gain, permutation importance or SHAP] and normalised to sum to 1.0. The following features were most influential in the reported analysis:

1. HNR
2. CPP
3. Jitter
4. Shimmer
5. Selected MFCC coefficients
6. Spectral slope
7. F0 variability

Table 05: Relative feature importance

Rank	Feature	Relative importance
1	HNR	0.22
2	CPP	0.19
3	Jitter	0.16
4	Shimmer	0.14
5	Selected MFCC coefficients	0.12
6	Spectral slope	0.10
7	F0 variability	0.07

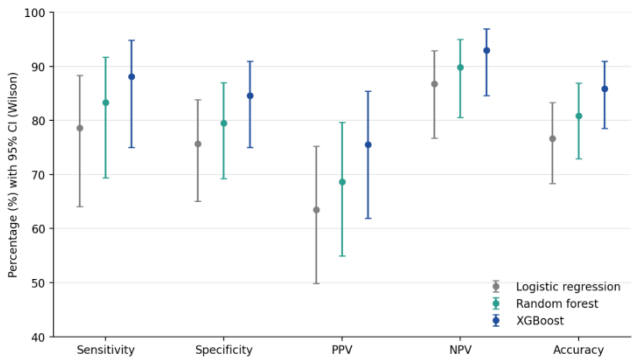

Figure 02: Diagnostic performance of the AI models with 95% confidence intervals (Wilson method) derived from the confusion matrices

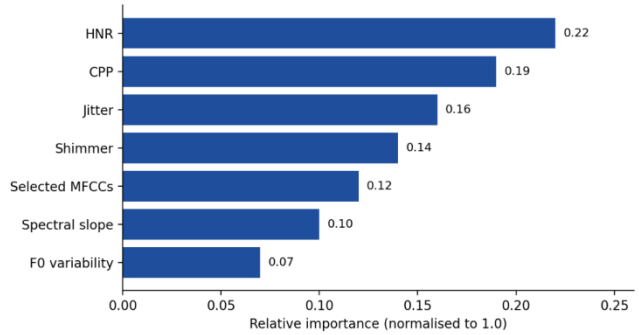

Figure 03: Relative feature importance in the XGBoost model (XGBoost gain).

DISCUSSION

AI-assisted acoustic voice analysis shows potential for distinguishing benign from malignant lesions of the larynx. In this single-centre dataset, a multivariable machine-learning approach showed encouraging discrimination when multiple acoustic features were considered together. These results represent internal model performance and should not be interpreted as evidence of clinical diagnostic effectiveness until independently validated in an external cohort.

The classical acoustic analysis has a role in voice evaluation. Jitter, shimmer, and harmonics-to-noise ratio (HNR) are often used by voice scientists to make clinical assessments of a voice sample. Research focusing on abnormal voices has documented heightened jitter and shimmer levels alongside decreased HNR in contrast to normal voices [1-3].

In a study of the Indian population using software Praat, pathological voices exhibited significantly elevated jitter and shimmer values and lower HNR values than healthy voices [3]. Thus, it indicates that objective acoustic analysis can be usefully implemented into routine clinical practice. However, distinguishing one pathology from another is more difficult than distinguishing normal from abnormal voice [5].

This limitation was addressed in recent studies by introducing several acoustic domains. A Scientific Reports study in 2024 assessed the AI-based classification of healthy voices, laryngeal cancer, vocal-fold paralysis, and BMD.5 The multilevel classification was more challenging since the accuracies reported in literature were between about 0.75 and 0.83 depending on the model [5].

The systematic review and meta-analysis by Marrero-Gonzalez et al [9]. supports continued investigation of AI-assisted approaches to laryngeal lesions. However, the pooled estimates were quoted only after confirming the exact voice-analysis subgroup and outcome definition in the original meta-analysis. The present study should therefore avoid presenting pooled sensitivity or specificity values unless they have been directly verified against the relevant subgroup.

Recent research has shown promising evidence for distinguishing malignancy. In a 2025 study of 157 participants (109 with benign lesions [18], with malignant lesions and 30 controls), LightGBM achieved an AUC of 0.924 for malignancy prediction, although the malignant subgroup was small [8], The authors also reported that specific acoustic descriptors contributed significantly to classification [8], This population and classification task differ from the present 120-participant benign-versus-malignant cohort and should not be treated as directly equivalent.

On the other hand, a 2026 clinical-utility study of 177 participants, including 47 with benign vocal-fold lesions and 58 with histopathologically confirmed laryngeal squamous-cell carcinoma, reported limited utility of acoustic measures for separating benign from malignant disease [11]. This provides an important counterpoint to a high internal-validation performance estimate and reinforces the need for external validation.

The combined model showed a higher point estimate of discriminatory performance than the other models in this dataset. This statement should not be interpreted as evidence of statistically superior performance unless a formal comparison of correlated AUCs, such as DeLong's test, was performed. The observed pattern of increased jitter and shimmer and decreased HNR and CPP in the malignant group is biologically plausible, but these parameters should not be considered malignancy-specific.

An important potential advantage of AI-assisted acoustic analysis is its non-invasive nature. Voice recordings can potentially be obtained with relatively low-cost equipment and may support pre-screening or risk stratification [9,12]. However, external validation and careful assessment of demographic and clinical confounding remain essential.

Consequently, the system should be considered an aid rather than a substitute for laryngoscopy and histopathological diagnosis. A positive prediction by AI could identify patients requiring urgent specialty assessment while a negative prediction must not be used to rule out malignancy in a patient with persistent or clinically suspicious dysphonia.

STRENGTHS

This study has several strengths:

1. Histopathology is used as the reference standard.
2. Acoustic data are acquired using a standardized protocol.
3. Multiple acoustic domains are integrated.
4. Three machine-learning algorithms were compared.
5. Diagnostic performance is reported using clinically interpretable measures.
6. The model can be tested in an independent external cohort in future work.

LIMITATIONS

Several limitations should be considered. First, acoustic characteristics are influenced by age, sex, smoking, recording conditions and coexisting voice disorders, and

the malignant group was significantly older and had more smokers, so residual confounding cannot be excluded. Second, only 42 participants had malignant lesions, which may produce unstable model estimates, particularly when multiple acoustic features and algorithms are evaluated. Third, differences in acoustic-analysis software and preprocessing can influence measurements. Fourth, internal validation may overestimate performance compared with independent external validation. Finally, voice-based models cannot directly visualise the lesion and therefore cannot replace endoscopic and histopathological assessment. If age, sex- and smoking-adjusted analyses were not performed, this should be stated explicitly.

CONCLUSION

AI-assisted acoustic voice analysis may serve as a non-invasive adjunct for differentiating benign and malignant laryngeal lesions. In this dataset, the XGBoost model showed an AUC of 0.91 and the reported confusion matrix corresponded to sensitivity of 88.1%, specificity of 84.6%, PPV of 75.5%, NPV of 93.0% and accuracy of 85.8%. These estimates represent internal model performance and should not be interpreted as evidence of clinical diagnostic effectiveness until independently validated in an external cohort. Prospective multicentre studies with larger malignant cohorts, standardized recording protocols, prespecified modelling procedures and independent external validation are required before clinical implementation.

FUNDING

Nil

INFORM CONSENT AND ETHICAL STATEMENT

Not Applicable

ACKNOWLEDGEMNET

Not Declared

CONFLICT OF INTREST

Nil

REFERENCES

1. Maryn Y, Corthals P, Van Cauwenberge P, Roy N, De Bodt M. Toward improved ecological validity in the acoustic measurement of overall voice quality: combining continuous speech and sustained vowels. J Voice. 2010;24(5):540-555.
2. Baken RJ, Orlikoff RF. Clinical Measurement of Speaking and Voice. 2nd ed. San Diego: Singular Thomson Learning; 2000.
3. Hippargekar P, Bhise S, Kothule S, Shelke S. Acoustic Voice Analysis of Normal and Pathological Voices in Indian Population Using Praat Software. Indian J Otolaryngol Head Neck Surg. 2022;74(Suppl 3):5069-5074.
4. Deliyski DD, Shaw HS, Evans MK. Adverse effects of environmental noise on acoustic voice analysis. J Voice. 2005;19(1):15-28.
5. Kim HB, Song J, Park S, Lee YO. Classification of laryngeal diseases including laryngeal cancer, benign mucosal disease, and vocal cord paralysis by artificial intelligence using voice analysis. Sci Rep. 2024;14:9297.
6. Godino-Llorente JI, Gómez-Vilda P, Blanco-Velasco M. Dimensionality reduction of a pathological voice quality assessment system based on Gaussian mixture models and short-term cepstral parameters. IEEE Trans Biomed Eng. 2006;53(10):1943-1953.
7. Oates J. Auditory-perceptual evaluation of disordered voice quality: pros, cons and future directions. Folia Phoniatr Logop. 2009;61(1):49-56.
8. Ma K, Wang Y, Zhou Y, Chen L, Zhang T, Xu F, Peng X. Acoustic signatures of organic lesions and the role of artificial intelligence in voice disorder diagnostics. Digital Health. 2025;11:20552076251376264. doi:10.1177/20552076251376264.
9. Marrero-Gonzalez AR, Diemer TJ, Nguyen SA, Camilon TJM, Meenan K, O'Rourke A, et al. Application of artificial intelligence in laryngeal lesions: a systematic review and meta-analysis. Eur Arch Otorhinolaryngol. 2025;282:1543-1555.
10. Keung LC, Richardson K, Sharp Matheron D, Martel-Sauvageau V. A Comparison of Healthy and Disordered Voices Using Multi-Dimensional Voice Program, Praat, and TF32. J Voice. 2024;38(4):963.e23-963.e38.
11. Ozcelik Erdem R, Arbag H. Clinical utility of acoustic voice parameters and patient demographic variables in identifying vocal fold pathology and supporting referral for laryngeal evaluation. J Voice. Published online August 22, 2026. doi:10.1016/j.jvoice.2026.07.056.
12. Jenkins PD, Bedrick S, Karstens L, Hersh W, Bridge2AI-Voice Consortium, Dorr DA. From voice biomarkers to telemedicine screening: developing and evaluating a voice-based AI model for laryngeal lesion detection using the Bridge2AI-Voice dataset. Front Digit Health. 2026;8:1846369.
13. Boersma P, Weenink D. Praat: doing phonetics by computer [Computer program]. Version [insert version]. Amsterdam: University of Amsterdam; [year accessed]. Available from: https://www.praat.org.
14. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: machine learning in Python. J Mach Learn Res. 2011;12:2825-2830.
15. Chen T, Guestrin C. XGBoost: a scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM; 2016:785-794.