Review Article Pol J Public Health, Vol. 136 (2026): 71-75

Artificial intelligence-assisted acoustic voice analysis as a public health tool for screening and triage of laryngeal diseases: a systematic review with qualitative synthesis

Emilia Wiśniewska[1,A-B,D-I], Michał Azierski[1,A-I,K-L]

[1] *Student Scientific Association of MedTech at the Center for Remote Learning and Educational Effects Analysis, Faculty of Medical Sciences in Katowice, Silesian Medical University, Katowice, Poland*

A – Development of the concept and methodology of the study; B – Query – a review and analysis of the literature; C – Submission of the application to the appropriate Bioethics Committee; D – Collection of research material; E – Analysis of the research material; F – Preparation of draft version of manuscript; G – Critical analysis of manuscript draft version; H – Statistical analysis of the research material; I – Interpretation of the performed statistical analysis; K – Technical preparation of manuscript in accordance with the journal regulations; L – Supervision of the research and preparation of the manuscript

Abstract

Introduction and aim. Voice disorders and laryngeal diseases impair communication, quality of life and work ability. Persistent dysphonia may also indicate glottic or laryngeal cancer. This systematic review evaluates artificial intelligence (AI)-assisted acoustic voice analysis as a non-invasive tool for screening support, triage and risk stratification of laryngeal diseases and voice disorders.

Methods. The review was guided by the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 statement. PubMed/MEDLINE, Scopus, Web of Science, IEEE Xplore and Google Scholar were searched using predefined Boolean strategies. The search identified 412 records; after removal of 118 duplicates, 294 titles and abstracts were screened. Eighty-two full texts were assessed, and 30 publications were included in qualitative synthesis. Meta-analysis was not performed because of heterogeneity in target conditions, voice tasks, acoustic features, artificial intelligence (AI) models and outcome metrics.

Brief summary of current knowledge. AI-assisted acoustic voice analysis shows high performance in binary classification of healthy and pathological voice, especially using mel-frequency cepstral coefficients (MFCCs), perturbation measures, harmonic-to-noise ratio (HNR) and spectrogram-based features. However, multiclass classification and differentiation between benign and malignant laryngeal disease remain challenging. Multimodal models using voice, demographic and clinical data appear more promising than voice-only approaches.

Summary. AI-assisted acoustic voice analysis may support scalable screening, triage and telemonitoring, particularly where specialist laryngological care is limited. Implementation requires standardized recordings, external validation, explainable models, safety-netting and prospective clinical studies.

Keywords: Artificial Intelligence; Voice; Voice Disorders; Laryngeal Diseases; Telemedicine; Mass Screening.

DOI: 10.12923/2083-4829/2026-0015

INTRODUCTION AND AIM

Voice disorders and laryngeal diseases are relevant clinical and public health problems as they impair communication, social participation and professional functioning. Dysphonia may result from benign vocal fold lesions, vocal fold paralysis, neurological voice disorders and malignant laryngeal diseases. Specialist examination with laryngoscopy or videolaryngostroboscopy remains essential, but access to such examination may be limited by geographic location, equipment availability and waiting times [1-4].

Artificial intelligence (AI)-assisted acoustic voice analysis has been therefore proposed as a scalable digital health approach. Voice can be recorded remotely, repeatedly and at low cost, making it an attractive support for screening and referral prioritisation. At the same time, the voice is a complex biomarker influenced by anatomy, phonation, respiratory function, language, age, sex and recording conditions, which means that performance claims require careful interpretation [5-10].

The aim of this systematic review was to evaluate the evidence on artificial intelligence-assisted acoustic voice analysis as a non-invasive tool for screening support, triage and risk stratification of laryngeal diseases and voice disorders, with emphasis on public health implementation, safety and limitations.

METHODS OF REVIEW

Design and reporting

This systematic review with qualitative synthesis was guided by the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 statement. The protocol was not prospectively registered. Because the field is heterogeneous and includes engineering, clinical and public health studies, the review focused on qualitative synthesis rather than statistical pooling.

Search strategy

A search of the PubMed/MEDLINE, Scopus, Web of Science, IEEE Xplore and Google Scholar was completed on 10 May 2026. Search strings combined terms related to artificial intelligence, acoustic voice or speech analysis, dysphonia, laryngeal diseases, vocal fold pathology, cancer and screening. Reference lists of key reviews and eligible studies were also screened.

The detailed database-specific search strategies, dates and results are presented in Table 1.

TABLE 1. Search strategy used for the systematic review.

Database	Search string	Date searched	Filters/ results
PubMed /MEDLINE	(“voice analysis” OR “speech analysis” OR dysphonia OR “voice disorder” OR “vocal biomarker”) AND (“artificial intelligence” OR “machine learning” OR “deep learning” OR “neural network” OR CNN OR SVM) AND (laryngeal OR “vocal fold” OR “laryngeal cancer”)	20.04.2026	English; humans; n = 126
Scopus	TITLE-ABS-KEY((“voice analysis” OR “speech analysis” OR dysphonia) AND (“artificial intelligence” OR “machine learning” OR “deep learning”) AND (laryngeal OR “vocal fold” OR cancer))	20.04.2026	English; n = 94
Web of Science	TS=((“voice analysis” OR “speech analysis” OR dysphonia) AND (AI OR “machine learning” OR “deep learning”) AND (laryngeal OR “vocal fold” OR “voice disorder”))	20.04.2026	English; n = 71
IEEE Xplore	(“voice pathology” OR “voice analysis” OR “speech analysis”) AND (“machine learning” OR “deep learning” OR “artificial intelligence”)	20.04.2026	Conference /journal records; n = 58
Google Scholar	Targeted search using combinations of: AI voice pathology; machine learning dysphonia; laryngeal cancer voice analysis; MFCC voice disorder. First 100 results and citation links were screened.	20.04.2026	Additional records n = 63

Eligibility criteria

The inclusion and exclusion criteria used to determine eligibility are summarized in Table 2.

TABLE 2. Eligibility criteria.

Criterion	Inclusion	Exclusion
Population /condition	Laryngeal diseases, voice disorders, vocal fold lesions, laryngeal cancer, dysphonia, post-laryngectomy voice and methodologically relevant neurological voice disorders	Studies unrelated to voice production, laryngeal pathology or acoustic voice/speech assessment
Index test /method	AI/ML or deep learning applied to acoustic voice/speech recordings or extracted acoustic features	Image-only laryngoscopy studies without voice component; subjective voice assessment only
Outcomes	Screening, triage, pathology detection/classification, risk stratification, performance metrics, validation, implementation barriers	Speech recognition studies without clinical diagnostic relevance
Publication type	Original studies, systematic/scoping reviews, meta-analyses and high-relevance methodological papers in English	Editorials, letters, book chapters, non-English publications, abstracts without full text

Study selection and data extraction

Two reviewers independently screened titles and abstracts, followed by full-text assessment. Disagreements were resolved by discussion and consensus. Reasons for full-text exclusion were recorded. Data extracted from eligible publications included study design, population or target condition, voice task, acoustic features, AI/ML model, validation strategy, performance metrics, clinical application and reported limitations.

The search identified 412 records. After removal of 118 duplicates, 294 titles and abstracts were screened. Eighty-two full-text reports were assessed for eligibility, and 30 publications were included in the qualitative synthesis. No study was eligible for quantitative meta-analysis.

The study selection process is illustrated in Figure 1, which shows record identification, screening, eligibility assessment and final inclusion.

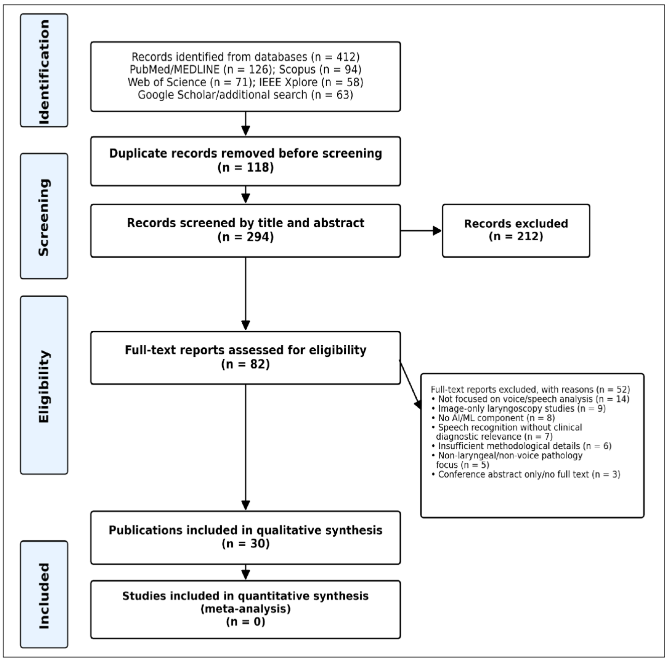

FIGURE 1. PRISMA flow diagram of study selection.

Quality assessment and synthesis

The Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2)-style assessment was applied to primary diagnostic or model-development studies, considering patient selection, the index test, the reference standard, the flow and the study timing, and overall risk of bias. Reviews and methodological papers were used to contextualise the field and identify additional primary studies, but they were not treated as primary clinical performance evidence. Due to the heterogeneity in target conditions, recording protocols, voice tasks, acoustic features, model architectures, validation methods and outcome metrics, statistical pooling would have been methodologically inappropriate.

Current knowledge

Study selection and evidence map

The included publications covered several evidence domains: general AI detection of voice pathology, acoustic feature extraction, deep learning and hybrid models, laryngeal cancer risk stratification, telemedicine and adjacent voice biomarker applications. Primary studies were prioritised when interpreting model performance, whereas systematic and scoping reviews were used to summarise broader methodological trends [1,11-22].

The main evidence domains and their public health interpretations are summarized in Table 3.

TABLE 3. Evidence domains and public health interpretation.

Evidence domain	Main findings	Public health interpretation
Binary detection of pathological voice	Often high reported performance when healthy and pathological voices are compared; MFCCs, perturbation measures, HNR and spectrograms are frequently used [11-17].	Potential screening support, but performance may be inflated by simple case-control designs.
Multiclass laryngeal disease classification	Performance decreases when models distinguish specific laryngeal conditions rather than healthy vs pathological voice [2,4,14,15,30].	Clinically relevant but not yet robust enough for standalone diagnostic use.
Cancer-oriented risk stratification	Voice-only models may be insufficient; multimodal approaches combining voice, demographics and medical records perform better [20-22].	Useful for triage and referral prioritisation, not for excluding cancer.
Telemedicine and primary care	Voice recordings can be collected remotely and may support decision-making before specialist evaluation [16-18,28].	May improve access where specialist laryngological services are limited.
Adjacent voice biomarker fields	Parkinson's disease, post-laryngectomy voice, COVID-19 and wearable sensing provide transferable methodological lessons [23-29].	Useful for method development, but not direct evidence for laryngeal disease screening.

Acoustic features and artificial intelligence methods

Acoustic analysis typically relies on parameters reflecting phonation stability and signal structure, including fundamental frequency, jitter, shimmer, harmonic-to-noise ratio, cepstral measures and spectral features. Mel-frequency cepstral coefficients (MFCCs) are among the most widely used representations because they approximate aspects of human auditory perception and can be combined with classical machine learning or deep learning models [11-14].

Classical algorithms such as support vector machines, Gaussian mixture models, the k-nearest neighbours and random forests remain common, particularly in smaller datasets. Deep learning models, including convolutional neural networks and recurrent architectures, can analyse spectrograms and temporal patterns but may require larger datasets and careful validation. Hybrid models combining acoustic parameters, cepstral features and ensemble classifiers have reported high accuracy, although external validation is often limited [15-19].

Screening, triage and laryngeal cancer risk stratification

The clearest distinction emerging from the review is between the screening support and clinical diagnosis. AI-assisted acoustic voice analysis can identify abnormal voice patterns and may help prioritise referrals, but it does not replace laryngoscopy, histopathology or specialist assessment. Binary classification has the strongest evidence, whereas clinically relevant multiclass classification remains more difficult [1,2,4,15-17,30].

Laryngeal cancer is the most safety-sensitive application. Meta-analytic evidence suggests promising diagnostic performance, but many studies compare cancer voices with healthy controls rather than with cases of benign dysphonia, which may overestimate real-world performance [20]. In a clinically relevant glottic neoplasm setting, the addition of demographic and structured clinical data improved performance compared with voice-only analysis, and acoustic signatures have also been explored for distinguishing organic lesions and malignancy [21,22].

False negative results represent the most important safety concern. A low-risk AI output should not delay referral in patients with persistent hoarseness, progressive symptoms, dysphagia, dyspnoea, haemoptysis, neck mass or relevant risk factors such as smoking. Therefore, any screening pathway should include safety-netting advice and predefined referral criteria independent of the algorithmic score.

Telemedicine, multimodal data and implementation pathway

From a public health perspective, the strongest value of AI-assisted acoustic voice analysis lies in scalability. A practical pathway could include a short clinical questionnaire, standardised voice recording, automated recording-quality assessment, risk scoring and referral recommendation. Patients could then be classified into urgent specialist consultation, standard consultation, monitoring with safety-netting or repeat recording due to poor signal quality.

Future models should incorporate multimodal variables, rather than voice alone. Relevant variables include age, sex, smoking history, duration of hoarseness, dysphagia, dyspnoea, throat pain, reflux symptoms, occupational exposure, prior laryngeal disease and patient-reported voice handicap. Such information may improve calibration and reduce false reassurance in high-risk patients.

The proposed public health implementation pathway, including clinical triage stages and safeguards, is presented in Table 4.

TABLE 4. Proposed public health implementation pathway.

Stage	Action	Safety safeguard
Patient entry	Persistent hoarseness or voice complaint reported in primary care or telemedicine	Alarm symptoms trigger referral regardless of AI score
Data collection	Standardised sustained vowel/ reading task plus short clinical questionnaire	Recording quality check and repeat recording if signal is inadequate
AI analysis	Feature extraction and risk score generation	Model output presented as risk support, not diagnosis
Clinical triage	Urgent referral, routine referral, monitoring, or repeat recording	Referral criteria independent of the algorithm remain active
Follow-up	Outcome confirmation after specialist evaluation	Continuous audit of false negatives, bias and calibration

Risk of bias, governance and implementation barriers

The risk of bias was driven mainly by patient selection, small or single-centre datasets, inconsistent reference standards, incomplete reporting of validation procedures and limited external validation. the generalisability is further affected by language, sex, age, microphone type, recording environment and clinical spectrum. These issues are especially important for public health deployment, where models must perform safely in heterogeneous populations [1,5,10,26,27].

The main risk-of-bias domains and recommended implementation safeguards are summarized in Table 5.

TABLE 5. Main risk-of-bias and implementation safeguards.

Domain	Common concern	Recommended safeguard
Patient selection	Small samples, enriched case-control datasets and limited demographic diversity	Prospective multicentre cohorts representing primary care and specialist settings
Index test	Variable microphones, tasks and preprocessing pipelines	Standardised recording protocol and automated signal-quality assessment
Reference standard	Different diagnostic labels and incomplete clinical confirmation	Uniform laryngological reference standard and transparent ground truth
Validation	Internal validation dominates; external validation uncommon	External validation, calibration reporting and subgroup performance analysis
Governance	Voice data are potentially identifiable and may reveal sensitive information	Consent, encryption, privacy-by-design and audit trails

Data governance is also central. The voice has potential to identify an individual and may contain information beyond the target disorder. Implementation should therefore include explicit consent, data minimisation, secure storage, model monitoring, transparent communication of limitations and clinician oversight. Wearable acoustic sensors and mobile recording tools may expand monitoring possibilities, but they also increase the need for privacy-preserving design [7,28,29].

SUMMARY

AI-assisted acoustic voice analysis is a promising public health-oriented tool for screening support, triage and telemonitoring of laryngeal diseases and voice disorders. The evidence is strongest for the binary classification of healthy and pathological voice. More clinically useful tasks, such as differentiating benign from malignant disease or distinguishing multiple laryngeal pathologies, remain underdeveloped.

The key conclusion is that voice-based AI should be implemented as an additional decision-support layer before specialist assessment, not as a replacement for laryngoscopy or oncological diagnostics. The safest pathway is multimodal process: voice recordings should be interpreted together with demographic data, symptoms, risk factors and patient-reported outcomes. Prospective clinical studies, external validation, explainable models, standardised recording protocols and active safety-netting are required before routine deployment.

REFERENCES

1. Liu GS, Jovanovic N, Sung CK, et al. A scoping review of artificial intelligence detection of voice pathology: challenges and opportunities. Otolaryngol Head Neck Surg. 2024;171:658-666. https://doi.org/10.1002/ohn.809
2. Hu H, Chang S, Wang CH, et al. Deep learning application for vocal fold disease prediction through voice recognition: preliminary development study. J Med Internet Res. 2021;23:e25247. https://doi.org/10.2196/25247
3. Wang TV, Song PC. Neurological voice disorders: a review. Int J Head Neck Surg. 2022;13:32-40. https://doi.org/10.5005/jp-journals-10001-1521
4. Marchese MR, Sensoli F, Campagnini S, et al. Artificial intelligence for the recognition of benign lesions of vocal folds from audio recordings. Acta Otorhinolaryngol Ital. 2023;43:317-323. https://doi.org/10.14639/0392-100x-n2309
5. Torborg S, Kim AY, Rameau A. New developments in the application of artificial intelligence to laryngology. Curr Opin Otolaryngol Head Neck Surg. 2024;32:391-397. https://doi.org/10.1097/MOO.0000000000000999
6. Bensoussan Y, Vanstrum EB, Johns MM, et al. Artificial intelligence and laryngeal cancer: from screening to prognosis: a state of the art review. Otolaryngol Head Neck Surg. 2022;168:319-329. https://doi.org/10.1177/01945998221110839
7. Muddaloor P, Baraskar B, Shah H, et al. The human voice as a digital health solution leveraging artificial intelligence. Sensors (Basel). 2025;25:3424. https://doi.org/10.3390/s25113424
8. Rogers HP, Hseu A, Kim JW, et al. Voice as a biomarker of pediatric health: a scoping review. Children (Basel). 2024;11:684. https://doi.org/10.3390/children11060684
9. Jenkins PM, Harrison R, Bedrick S, et al. Voice as a biomarker: exploratory analysis for benign and malignant vocal fold lesions. Front Digit Health. 2025;7:1609811. https://doi.org/10.3389/fdgth.2025.1609811
10. Wilson BS, Tucci DL, Moses DA, et al. Harnessing the power of artificial intelligence in otolaryngology and the communication sciences. JARO. 2022;23:319-349. https://doi.org/10.1007/s10162-022-00846-2
11. Islam R, Tarique M, Abdel-Raheem E. A survey on signal processing based pathological voice detection techniques. IEEE Access. 2020;8:66749-66776. https://doi.org/10.1109/ACCESS.2020.2985280
12. Tracey B, Volfson D, Glass J, et al. Towards interpretable speech biomarkers: exploring MFCCs. Sci Rep. 2023;13:22787. https://doi.org/10.1038/s41598-023-49352-2
13. Reddy MK, Alku P. A comparison of cepstral features in the detection of pathological voices by varying the input and filterbank of the cepstrum computation. IEEE Access. 2021;9:135953-135963. https://doi.org/10.1109/ACCESS.2021.3117665
14. Song J, Kim H, Lee YO. Laryngeal disease classification using voice data: octave-band vs. mel-frequency filters. Heliyon. 2024;10:e40748. https://doi.org/10.1016/j.heliyon.2024.e40748
15. Barlow J, Sragi Z, Rivera-Rivera G, et al. The use of deep learning software in the detection of voice disorders: a systematic review. Otolaryngol Head Neck Surg. 2024;170:1531-1543. https://doi.org/10.1002/ohn.636
16. Verma V, Benjwal A, Chhabra A, et al. A novel hybrid model integrating MFCC and acoustic parameters for voice disorder detection. Sci Rep. 2023;13:22719. https://doi.org/10.1038/s41598-023-49869-6
17. Compton EC, Cruz T, Andreassen M, et al. Developing an artificial intelligence tool to predict vocal cord pathology in primary care settings. Laryngoscope. 2023;133:1952-1960. https://doi.org/10.1002/lary.30432
18. Suleman A, Rutt AL. Global utilization of artificial intelligence in the diagnosis and management of voice disorders over the past five years. Eye ENT Res. 2025;2:88-95. https://doi.org/10.1002/eer3.70006
19. Oliver JR, Alapati R, Lee JT, et al. Artificial intelligence in head and neck surgery. Otolaryngol Clin North Am. 2024;57:803-820. https://doi.org/10.1016/j.otc.2024.05.001

20. Alabdalhussein A, Al-Khafaji M, Al-Busairi R, et al. Artificial intelligence in laryngeal cancer detection: a systematic review and meta-analysis. Curr Oncol. 2025;32:338. https://doi.org/10.3390/curroncol32060338
21. Wang C, Chen T, Lee N, et al. AI detection of glottic neoplasm using voice signals, demographics, and structured medical records. Laryngoscope. 2024;134:4585-4592. https://doi.org/10.1002/lary.31563
22. Ma K, Wang Y, Zhou Y, et al. Acoustic signatures of organic lesions and the role of artificial intelligence in voice disorder diagnostics. Digit Health. 2025;11:20552076251376264. https://doi.org/10.1177/20552076251376264
23. Bose D, Mukherjee A, Acharya M, et al. Artificial intelligence for detection of Parkinson's disease from speech signals–a comprehensive review. Biofactors. 2025;51:e70065. https://doi.org/10.1002/biof.70065
24. Liu Y, Reddy MK, Penttila N, et al. Automatic assessment of Parkinson's disease using speech representations of phonation and articulation. IEEE/ACM Trans Audio Speech Lang Process. 2023;31:242-255. https://doi.org/10.1109/TASLP.2022.3212829
25. Maskeliunas R, Kulikajevas A, Damasevicius R, et al. Lightweight deep learning model for assessment of substitution voicing and speech after laryngeal carcinoma surgery. Cancers (Basel). 2022;14:2366. https://doi.org/10.3390/cancers14102366
26. Syamal MN. The effects of COVID-19 on voice. World J Otorhinolaryngol Head Neck Surg. 2025;11:524-529. https://doi.org/10.1002/wjo2.70012
27. Gomez-Vilda P, Gomez-Rodellar A, Palacios-Alonso D, et al. The role of data analytics in the assessment of pathological speech–a critical appraisal. Appl Sci (Basel). 2022;12:11095. https://doi.org/10.3390/app122111095
28. Muhammad G, Alhussein M. Convergence of artificial intelligence and internet of things in smart healthcare: a case study of voice pathology detection. IEEE Access. 2021;9:89198-89209. https://doi.org/10.1109/ACCESS.2021.3090317
29. Liu T, Mao YJ, Dou H, et al. Emerging wearable acoustic sensing technologies. Adv Sci. 2025;12:2408653. https://doi.org/10.1002/advs.202408653
30. Kim HB, Song J, Park S, et al. 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. https://doi.org/10.1038/s41598-024-58817-x

ORCID

Emilia Wiśniewska https://orcid.org/0009-0005-8796-4682
Michał Azierski https://orcid.org/0009-0009-7247-2086

Corresponding author
Michał Azierski
Student Scientific Association of MedTech at the Center for Remote Learning and Educational Effects Analysis, Faculty of Medical Sciences in Katowice, Silesian Medical University, Katowice, Poland
Medyków 18, Katowice, Poland
e-mail: michal.azierski@gmail.com