International Journal of Innovative Technologies in Social Science

e-ISSN: 2544-9435

Operating Publisher
SciFormat Publishing Inc.
ISNI: 0000 0005 1449 8214

2734 17 Avenue SW,
Calgary, Alberta, T3E0A7,
Canada
+15878858911
editorial-office@sciformat.ca

ARTICLE TITLE	VOICE AND COUGH AS DIGITAL BIOMARKERS (2025–2026): LINGUISTIC EQUITY, PRIVACY, AND THE PROMISE OF ACOUSTIC DIAGNOSTICS IN GLOBAL HEALTH
DOI	https://doi.org/10.31435/ijitss.3(51).2026.5732
RECEIVED	04 May 2026
ACCEPTED	04 August 2026
PUBLISHED	10 August 2026
LICENSE	The article is licensed under a **Creative Commons Attribution 4.0 International License.**

VOICE AND COUGH AS DIGITAL BIOMARKERS (2025–2026): LINGUISTIC EQUITY, PRIVACY, AND THE PROMISE OF ACOUSTIC DIAGNOSTICS IN GLOBAL HEALTH

Wiktor Adamiec *(Corresponding Author, Email: wiktor.adamiec97@gmail.com)*
John Paul II Municipal Hospital in Rzeszów, Rycerska 4, 35-241 Rzeszów, Poland
ORCID ID: 0009-0002-3322-5479

Damian Danilczuk
John Paul II Municipal Hospital in Rzeszów, Rycerska 4, 35-241 Rzeszów, Poland
ORCID ID: 0009-0005-5755-1115

Jerzy Buszko
Medical Care Center in Jarosław, 3-go Maja 70, 37-500 Jarosław, Poland
ORCID ID: 0009-0008-2708-1222

Michał Dyś
Health Care Complex in Dębica, Krakowska 91, 39-200 Dębica, Poland
ORCID ID: 0009-0000-2644-7626

Nicol Baran
Medical Center in Łańcut, Ignacego Paderewskiego 5, 37-100 Łańcut, Poland
ORCID ID: 0009-0003-3607-6744

Marta Bajkowska-Piterak
Medical Center in Łańcut, Ignacego Paderewskiego 5, 37-100 Łańcut, Poland
ORCID ID: 0009-0000-6689-0788

Przemysław Piterak
Hospital of the Ministry of Interior and Administration in Rzeszów, Krakowska 16, 35-111 Rzeszów, Poland
ORCID ID: 0009-0002-3553-7353

Monika Szlachta-Gubernat
Hospital of the Ministry of Interior and Administration in Rzeszów, Krakowska 16, 35-111 Rzeszów, Poland
ORCID ID: 0009-0005-6994-9761

Adrianna Buż
John Paul II Municipal Hospital in Rzeszów, Rycerska 4, 35-241 Rzeszów, Poland
ORCID ID: 0009-0001-8202-5628

ABSTRACT

Recent advances in artificial intelligence and mobile sensing have renewed interest in the human voice and cough as non-invasive, low-cost biomarkers of disease. While respiratory, neurological, and psychiatric applications have shown moderate-to-high diagnostic accuracies in classification studies, the translation of these tools from proof-of-concept into routine clinical practice remains incomplete. This narrative review synthesizes the 2025–2026 literature across five domains: (a) the biophysiological foundations that make voice and cough acoustically informative, (b) clinical evidence spanning respiratory, neurodegenerative, mental-health, cardiometabolic, pediatric, and laryngeal use cases, (c) the growing body of work on linguistic equity and structural bias in speech-based systems, (d) privacy and adversarial security concerns specific to voice as biometric data, and (e) deployment prospects in low- and middle-income countries, with tuberculosis screening as the most advanced use case. The analysis reveals that no single disease domain has reached the level of large-scale, independently validated deployment, though respiratory and laryngeal applications are closest. Structural language bias—particularly against speakers of tonal languages—remains a first-order obstacle, and current speaker de-identification systems demonstrably leak identity information. These findings point to three conditions for responsible scale-up: methodological standardization, inclusive multilingual datasets, and governance frameworks that treat voice data as sensitive biometric medical information.

KEYWORDS

Voice Biomarkers, Cough Analysis, Digital Health, Linguistic Equity, Privacy, Global Health

CITATION

Wiktor Adamiec, Damian Danilczuk, Jerzy Buszko, Michał Dyś, Nicol Baran, Marta Bajkowska-Piterak, Przemysław Piterak, Monika Szlachta-Gubernat, Adrianna Buż. (2026) Voice and Cough as Digital Biomarkers (2025–2026): Linguistic Equity, Privacy, and the Promise of Acoustic Diagnostics in Global Health. *International Journal of Innovative Technologies in Social Science*. 3(51). doi: 10.31435/ijitss.3(51).2026.5732

1. Introduction

The idea that a person's voice or cough might carry medically actionable information is not new. Clinicians have long listened for hoarseness, stridor, or the distinctive bark of croup; what has changed is the availability of cheap, high-quality microphones embedded in billions of smartphones and the maturation of machine-learning architectures capable of extracting subtle acoustic patterns from short audio clips. Together, these developments have turned voice and cough into candidate digital biomarkers—measurable features of a biological signal that can be captured passively, analyzed computationally, and, in principle, deployed at population scale without specialized laboratory equipment.

The promise is considerable. A vocal biomarker collected through a smartphone application could, in theory, flag early Parkinson's disease in a rural clinic that has never seen a neurologist, screen for tuberculosis at a community health post in sub-Saharan Africa, or monitor depressive episodes between psychiatric appointments. Yet the 2025–2026 literature, when read carefully, suggests a field that is still negotiating the distance between promising classification results and trustworthy clinical deployment. Most published work remains at the level of review-level synthesis and proof-of-concept modeling rather than mature clinical validation (Landry et al., 2025). Several of the largest systematic reviews note persistent heterogeneity in study design, limited external validation, and a recurring failure to test models in populations that differ linguistically, ethnically, or socioeconomically from the training cohorts.

At the same time, a parallel literature has emerged that treats equity and privacy not as afterthoughts but as constitutive problems. Research on automatic speech recognition (ASR) and adjacent speech technologies has documented structural language bias that penalizes speakers of tonal languages such as Mandarin and Vietnamese (Roll & Graham, 2025), and a scoping review of fairness in ASR for African American English speakers has argued that technical fairness fixes, while important, are insufficient without governance structures that center community agency and linguistic justice (Cunningham et al., 2025). On the privacy side, a 2025 benchmark found that every state-of-the-art speaker de-identification system tested leaked identity

information (Seo et al., 2025), and an adversarial attack study demonstrated that synthetic multilingual pathological speech could deceive existing diagnostic models (Zhu et al., 2025).

This review integrates both streams—the clinical and the sociotechnical—into a single analytical framework. The aim is not to catalogue every voice-biomarker study published in the period, but to characterize the maturity of clinical evidence, map the most consequential equity and privacy risks, and identify what conditions would need to be met before acoustic diagnostics could be responsibly scaled, especially in low- and middle-income countries (LMICs). The analysis is organized around five thematic areas: biophysiological foundations, clinical applications, linguistic equity, privacy and regulation, and global-health deployment.

It is worth noting at the outset what the COVID-19 pandemic contributed to this field—and what it distorted. The global urgency of 2020–2022 prompted a surge in cough-based and voice-based COVID-19 detection studies, many of which reported impressively high accuracies on datasets collected under controlled conditions. As the immediate crisis receded, however, much of this work proved difficult to replicate in independent external validations, a lesson that has made the post-pandemic literature more cautious about claiming deployment readiness on the basis of single-dataset classification performance. The studies reviewed here inherit that caution, and readers will notice that the most credible clinical claims tend to come from multi-center designs, real-world recording conditions, or systematic reviews that explicitly assess risk of bias.

2. Methodology

This study adopted a narrative review design, appropriate for synthesizing evidence across a heterogeneous and rapidly evolving interdisciplinary field. The authors conducted a comprehensive manual literature search across multiple bibliographic databases, including PubMed/MEDLINE, Embase, Scopus, Web of Science, IEEE Xplore, and Google Scholar. Six thematic searches were performed independently by the authors, targeting (a) voice and cough biophysiology and acoustic taxonomy, (b) clinical applications of audio biomarkers across disease domains, (c) continuous cough monitoring and respiratory disease classification, (d) linguistic bias and equity in ASR and speech-biomarker systems, (e) privacy, de-identification, and adversarial attacks on voice data, and (f) tuberculosis screening and acoustic diagnostics in LMICs. Reference lists of key articles were also screened by hand to identify additional relevant publications (snowball sampling).

Search terms included combinations of “voice biomarker,” “cough analysis,” “digital biomarker,” “speech recognition bias,” “linguistic equity,” “speaker de-identification,” “tuberculosis cough screening,” and related variants, with appropriate Boolean operators (AND, OR, NOT). Eligible studies were published between 2019 and 2026, with a strong emphasis on work from 2024–2026. Both peer-reviewed publications and preprints with clear institutional affiliation were included. Regulatory and policy documents from the World Health Organization (WHO) and the European Commission were retrieved directly from the respective institutional websites and consulted for the privacy and global-health sections. Records were exported and deduplicated manually by cross-checking titles, author lists, and digital object identifiers (DOIs). Each candidate study was then screened by at least two authors against pre-specified inclusion criteria; disagreements were resolved through discussion. In total, 43 unique sources were identified and verified for accuracy. Sources were organised thematically, and a narrative synthesis was conducted with attention to the level of evidence, study design, and explicit limitations reported by each study’s authors.

3. Results

3.1 Biophysiological Foundations of Voice and Cough

The acoustic features extracted from voice and cough recordings derive their diagnostic informativeness from well-understood, if complex, physiology. Voice production is still most parsimoniously described by source–filter theory: the vocal folds generate a quasi-periodic source signal, the vocal tract shapes it through resonance, and the interaction between the two—traditionally treated as negligible—is increasingly recognized as clinically relevant (Calvache et al., 2021). A separate line of perceptual research argues that the recurring descriptive terms used across disciplines, languages, and even species track a small number of acoustic dimensions related to body size and physiological arousal, which may explain why voice taxonomies keep converging on a limited set of perceptual–acoustic axes (Kreiman, 2023).

In clinical practice, the relevant feature families are relatively stable. A tutorial review on clinical acoustic markers identified fundamental frequency, pitch and amplitude variation, harmonic-to-noise ratio (HNR), cepstral peak prominence (CPP), maximum phonation time, and pause-related measures as the core

analytic toolkit (Schultz & Vogel, 2022). A 2025 review of dysphonia assessment reached a similar list, adding relative noise level and the acoustic voice quality index as informative parameters (Timerbulatov et al., 2025). In essence, most published "voice biomarker" studies measure one of three properties: the periodicity of the laryngeal source, the degree of noise or aperiodicity in the signal, or the way the vocal-tract filter reshapes the harmonic spectrum.

Cough occupies a different biophysical niche. The CHEST Guideline and Expert Panel report frames cough not as a simple burst of sound but as a phenomenon with characteristic patterns, morphologic features, triggers, and a hypersensitivity phenotype (Lee et al., 2021). A 2023 acoustic study in healthy adults demonstrated that voluntary cough, throat clearing, and induced reflexive cough separate acoustically, with distinct onset-pulse and frication profiles across maneuvers (Mootassim-Billah et al., 2023). A scoping review of 108 studies further divided the field into acquisition, detection, and classification stages, noting the increasing use of smartphones for non-contact recording since 2016 (Hegde et al., 2024). The diagnostically useful acoustic taxonomy for cough thus includes temporal envelope features, spectral-band energy distribution, and features that discriminate voluntary from reflexive, wet from dry, and productive from nonproductive cough.

The overlap between voice and cough biomarkers deserves emphasis. A study of cough airflow and acoustics after injection laryngoplasty showed that restoring glottic competence changes both cough airflow and cough acoustics (Rameau et al., 2022). This finding serves as a reminder that the cough signal is not simply "lung noise"; it is a laryngeal–respiratory event shaped jointly by expiratory airflow, glottal closure and release, secretion burden, and airway mechanics.

This biophysiological picture carries direct implications for the linguistic-equity question addressed later in this review. Because fundamental frequency (F0) is both a core diagnostic feature for conditions like depression and Parkinson's disease and a lexical feature in tonal languages such as Mandarin, Vietnamese, Yoruba, and Thai, the diagnostic meaning of F0-based biomarkers cannot be assumed to transfer across language families without explicit validation. Cough, by contrast, is largely involuntary and language-independent, which makes cough-derived features inherently more portable across populations—though they remain susceptible to confounding by non-pathological factors such as smoking, ambient humidity, and recording equipment.

3.2 Clinical Evidence Across Disease Domains

The clinical evidence base as of 2025–2026 is uneven. Respiratory and laryngeal use cases appear closest to routine deployment, mental-health and neurodegenerative applications are promising but remain validation-heavy, and cardiometabolic applications are narrowest in scope. Figure 1, compiled from Landry et al. (2025), Abdelhalim et al. (2025), Briganti and Lechien (2025), Yang Y. et al. (2025), and Ma et al. (2025), visualises these performance ranges across disease domains.

Respiratory disease. This is the most mature domain. A 2025 systematic review of 81 studies found that audio biomarkers generally achieved moderate (60–79%) to high (80–100%) diagnostic accuracies for asthma, COPD, and pneumonia, though the authors emphasized persistent methodological heterogeneity and the need to test models in clinically diverse populations (Landry et al., 2025). A 2026 device-invariant multimodal cough model, trained on a real-world multi-center dataset exceeding 10,000 cases, reported improved cross-device generalization—precisely the kind of engineering advance required before deployment (Yang et al., 2026). A narrative review of continuous cough monitoring argued that analytically and clinically validated monitoring is already informing management of refractory chronic cough, COPD, bronchiectasis, congestive heart failure, and gastroesophageal reflux (Galvosas & Small, 2025).

Neurodegenerative disease. Voice disorders affect up to 89% of Parkinson's disease patients, making this the flagship neurological use case. A 2025 systematic review of machine-learning applications for Parkinson's diagnosis via speech found that free speech and reading were the most commonly used elicitation tasks, but the field remains limited by comparability problems and uncertain clinical integration pathways (Hossain et al., 2025). A FAIRness audit of 27 publicly available voice-biomarker datasets reported high findability but substantial weaknesses in accessibility, interoperability, and reusability (Mahapatra & Mahapatra, 2026)—exactly the bottleneck that prevents classification models from crossing into clinical practice.

Mental health. A 2025 systematic review of speech and voice quality as biomarkers in depression reported areas under the curve (AUCs) ranging from 0.71 to 0.93, but six of the included studies carried high risk of bias, primarily from patient selection and validation design choices (Briganti & Lechien, 2025). That

positions voice as a plausible screening or monitoring signal for depression, not a stand-alone diagnostic tool. A perspective article on integrating vocal biomarkers into digital health proposed singing as a potentially more accessible elicitation task than speech, particularly for populations with linguistic or cognitive impairments (Rodrigo & Duñabeitia, 2025).

Cardiometabolic disease. The evidence base here is the narrowest. A scoping review of 19 studies on voice biomarkers in heart failure found that nine focused on diagnosis or differential diagnosis and ten on disease-progression monitoring, while still calling for standardized collection and analysis procedures (Yang, Chen, et al., 2025). Direct voice or cough evidence for diabetes or metabolic syndrome remains sparse.

Pediatrics. Pediatric cough work is encouraging for respiratory triage in low-resource settings. A 2025 systematic review reported sensitivities of 82–94% and specificities of 71–91% across six cough-AI studies, though most had small sample sizes and lacked external validation (Abdelhalim et al., 2025). A separate study suggested that standard deviation and variance of mel-frequency cepstral coefficients (MFCCs) may help screen pediatric vocal cord nodules, but this remains a single cross-sectional observation (Jiang et al., 2025).

Laryngeal applications. These are perhaps the most anatomically grounded of the voice use cases. The standard clinical acoustic families—fundamental frequency, HNR, CPP, maximum phonation time—map directly onto laryngeal source mechanics (Timerbulatov et al., 2025). The laryngoplasty study cited above (Rameau et al., 2022) reinforces the point that glottic competence is readable in both voice and cough acoustics, making these two biomarker classes partially overlapping rather than strictly independent. Table 1 synthesises the diagnostic-accuracy ranges and evidence levels reported across these domains, drawing on Landry et al. (2025), Hossain et al. (2025), Briganti and Lechien (2025), Yang Y. et al. (2025), Abdelhalim et al. (2025), and Ma et al. (2025).

Table 1. Summary of Clinical Evidence for Voice and Cough Biomarkers Across Disease Domains (2025–2026)

Disease Domain	Signal Type	Accuracy Range	Evidence Level	Key Limitation
Asthma	Voice / Cough	Sens. 80–100% Spec. 80–100%	Systematic review (81 studies)	Heterogeneous methods; limited diverse-population testing
COPD	Voice / Cough	Sens. 78–100% Spec. 56–100%	Systematic review	Specificity variable; cross-device drift
Pneumonia	Cough / Lung sounds	Sens. 64–100% Spec. 64–100%	Systematic review	Small pediatric samples; limited external validation
TB screening	Cough	Sens. 90.3% Spec. 73.1%	Prospective, 2-site (n=500)	Single-country; urban hospitals only
Parkinson's disease	Voice	Variable (review-level)	Systematic review	Comparability across tasks; no consensus protocol
Depression	Voice	AUC 0.71–0.93	Systematic review	High risk of bias in 6/N studies; patient selection issues
Heart failure	Voice	Not standardized	Scoping review (19 studies)	No standardized procedures; exploratory stage
Pediatric respiratory	Cough	Sens. 82–94% Spec. 71–91%	Systematic review (6 studies)	Small samples; no external validation

Accuracy ranges are compiled from the systematic reviews and primary studies cited in Section 3.2. "Sens." = sensitivity; "Spec." = specificity. AUC = area under the receiver operating characteristic curve.

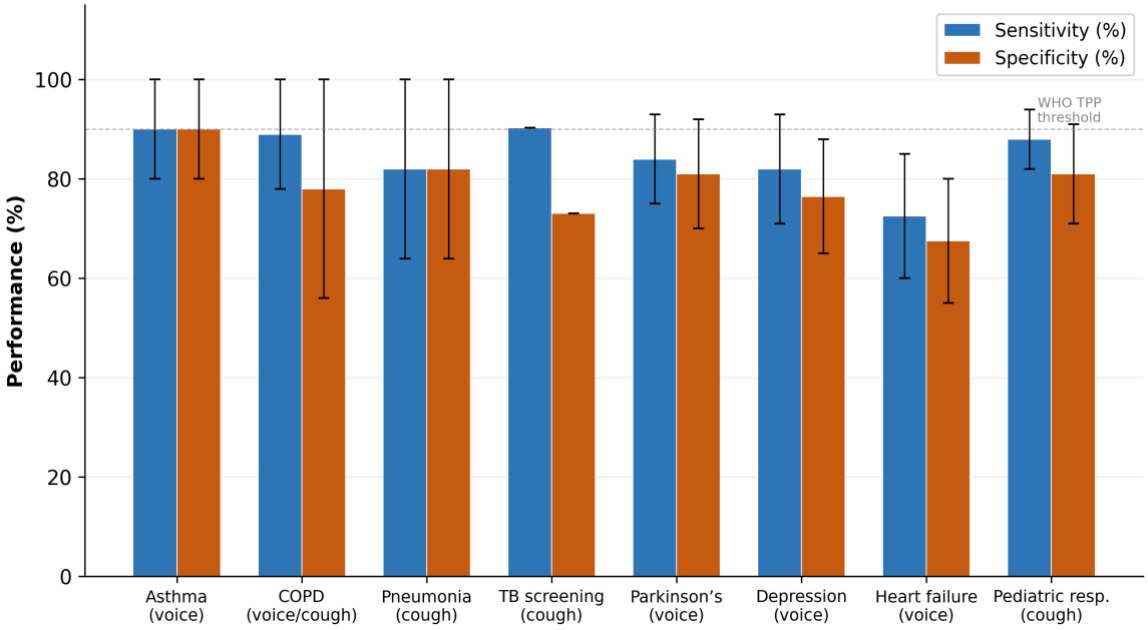

***Fig. 1**. Reported Diagnostic Sensitivity and Specificity Ranges for Voice and Cough Biomarkers Across Disease Domains*

Error bars represent the range of reported values across included studies. The dashed line indicates the WHO target product profile sensitivity threshold (90%) for TB screening. Single-point estimates (TB) reflect the best-performing multimodal model from Ma et al. (2025).

3.3 Linguistic Equity in Acoustic Diagnostic Systems

The literature on linguistic equity has converged, by 2025–2026, on a conclusion that should concern anyone building voice-based clinical tools: ASR and speech-biomarker systems do not merely vary by accent; they carry structural language bias, and speakers of tonal languages are among the groups most penalized even when a model looks strong on aggregate metrics.

Three ASR studies from 2024–2026 document the scale of the problem. A study using the Lottery Ticket Hypothesis to probe XLS-R, a large multilingual self-supervised model, found that during fine-tuning the model bypasses traditional linguistic knowledge and instead builds on weights learned primarily from the languages with the largest share of pretraining data (Storey et al., 2025). A comprehensive evaluation of ten ASR models across more than 300 English varieties found that as models increase in size and performance on native-speaker English, the relative disparity for second-language speakers worsens, with a significant bias against speakers whose first language is tonal (Roll & Graham, 2025). A 2026 study that is the first systematic investigation of speech bias in multimodal large language models found high sensitivity to language variation and option order, with architectural design and reasoning strategy substantially affecting robustness across languages (Wei et al., 2026).

For voice biomarkers, the implication is direct: “fairness” is not only about demographic parity in classification outcomes; it is about whether a model’s performance survives contact with the linguistic diversity of real clinical populations. Two technical approaches have shown promise. FairASR demonstrated that contrastive learning can reduce demographic disparities in ASR while maintaining competitive overall accuracy (Kim et al., 2025), and a privacy-preserving data-selection method showed that bias can be mitigated without requiring access to sensitive demographic attributes at inference time (Koudounas et al., 2025). Federated learning represents a complementary strategy: it keeps raw audio on-device, and parameter-efficient adapters can narrow performance gaps without centralizing voice data (Kan et al., 2024; Jain et al., 2025). The important caveat is that federated learning is a mitigation strategy, not a cure; the underlying data imbalance and language-coverage deficit must still be addressed at the dataset level.

Singing has emerged as the most intriguing proposal for a language-light biomarker task. A perspective article argued that singing engages motor, emotional, and cognitive systems in a highly integrated manner and can be standardized across cultures more readily than scripted speech tasks (Rodrigo & Duñabeitia, 2025). A

separate paper proposed minimum measurement standards for voice-related biomarkers in both speech and singing (Pedersen, 2025). A scoping-review protocol on group singing and health biomarkers suggests that the field is actively mapping this evidence base (Bouguettaya et al., 2025). These contributions support singing as a promising, relatively language-agnostic probe; they do not yet demonstrate that it is universal or that it bypasses linguistic inequity in a validated way. Table 2 brings together the principal findings from this literature, drawing on Storey et al. (2025), Roll and Graham (2025), Wei et al. (2026), Kim et al. (2025), Koudounas et al. (2025), and Cunningham et al. (2025).

Table 2. Evidence of Structural Language Bias in Speech and Audio AI Systems (2024–2026)

Study	System Tested	Key Finding	Affected Group	Proposed Mitigation
Storey et al. (2025)	XLS-R (SSL)	Model builds on high-resource language weights; bypasses linguistic structure	Low-resource languages	Language-aware pretraining data balancing
Roll & Graham (2025)	10 ASR models, 300+ varieties	Larger models increase L2 disparity; tonal-language speakers penalized	Tonal L1 speakers (Mandarin, Vietnamese)	Integration of L2 speech into training pipelines
Wei et al. (2026)	9 multimodal LLMs	High sensitivity to language; speech amplifies structural biases	Non-English speakers	Architectural redesign; stronger reasoning strategies
Kim et al. (2025)	FairASR	Contrastive learning reduces demographic disparities with minimal accuracy loss	Demographic subgroups	Fair contrastive learning framework
Koudounas et al. (2025)	Speech models (ACL industry)	Bias mitigated without sensitive attributes at inference	Multiple groups	Privacy-preserving data selection
Cunningham et al. (2025)	44 ASR / SLT papers	Technical fixes insufficient; governance gap identified	AAE speakers, diverse communities	Governance-centered ASR lifecycle

SSL = self-supervised learning; L1 = first language; L2 = second language; AAE = African American English; SLT = speech and language technology.

3.4 Privacy, Adversarial Security, and the Regulatory Landscape

The privacy challenge posed by voice biomarkers is more fundamental than the phrase “anonymized audio” typically implies. Voice data can be used to infer identity, sex, age, ethnic origin, health status, and emotional state, placing it squarely within the category of sensitive biometric information. Under the EU General Data Protection Regulation (GDPR) and the EU AI Act, biometric data occupies a special regulatory tier, and one legal analysis frames the resulting framework as a continuing negotiation between privacy protection and technological innovation (Kalodanis et al., 2025).

On the technical side, de-identification looks increasingly like risk reduction rather than true anonymization. A 2025 benchmark evaluated all major state-of-the-art speaker de-identification systems and found that every one of them leaked identity information; the best-performing system was only slightly better than random guessing (Seo et al., 2025). A multimodal attack study showed that combining anonymized speech with textual transcription exposes critical vulnerabilities in current anonymization methods (Aloradi et al., 2025). A 2024 perturbation study demonstrated that anonymized speech can be partially restored when the perturbation mechanism is known (Guo et al., 2024). DeepSick extends this line of concern into the diagnostic

domain itself: current synthetic-voice detectors, typically trained on healthy speech, perform poorly when confronted with generated pathological samples (Zhu et al., 2025). Table 3 consolidates these vulnerabilities, drawing on Seo et al. (2025), Aloradi et al. (2025), Guo et al. (2024), and Zhu et al. (2025).

Table 3. Demonstrated Vulnerabilities of Voice De-Identification and Diagnostic Systems

Study	Attack / Evaluation	Finding	Implication for Voice Biomarkers
Seo et al. (2025)	Identity-leakage benchmark on state-of-the-art de-ID systems	All systems leak identity; best system only slightly above random guessing	De-identification cannot be assumed to protect speakers
Aloradi et al. (2025)	VoxATtack: multimodal (audio + text) de-anonymization	ECAPA-TDNN + BERT fusion outperforms top attackers on 5/7 benchmarks	Transcript availability amplifies re-identification risk
Guo et al. (2024)	Adversarial perturbation removal from anonymized speech	Anonymized speech partially restorable when perturbation mechanism is known	Security-by-obscurity insufficient for privacy
Zhu et al. (2025)	DeepSick: synthetic pathological speech generation	Diagnostic models deceived; synthetic voice detectors fail on pathological speech	Clinical voice AI must be red-teamed against deepfakes

De-ID = de-identification. All findings are from peer-reviewed or preprint studies published 2024–2025.

The regulatory picture is fragmented. A cross-jurisdictional review compared the GDPR, the Medical Devices Regulation (MDR), the EU AI Act, HIPAA and FDA guidance in the United States, China's Personal Information Protection Law (PIPL), Japan's Act on the Protection of Personal Information (APPI) and Pharmaceutical and Medical Devices Act, and India's Digital Personal Data Protection Act (Jayasree et al., 2025). A separate EU-focused analysis described the legal and regulatory environment for multi-country medical AI as "complex, evolving, and presenting some conflicting principles" (Niemiec et al., 2025). The EU AI Act classifies AI systems used for medical purposes as high-risk, imposing obligations for data governance, transparency, and human oversight that operate in parallel with existing medical-device regulations (Kalodanis et al., 2025). Table 4 sets these jurisdictions side by side, synthesising Jayasree et al. (2025), Kalodanis et al. (2025), and Niemiec et al. (2025).

The practical implication of these findings is straightforward: voice biomarker systems should be treated as handling sensitive biometric medical data, built with data minimization and local or federated processing wherever feasible, and subjected to adversarial red-teaming against both re-identification and pathological-speech spoofing before they can be considered clinically trustworthy.

Table 4. Comparative Regulatory Landscape for Voice-Based Medical AI (2025–2026)

Jurisdiction	Primary Frameworks	Voice Data Classification	AI-Specific Rules	Key Challenge
European Union	GDPR + MDR + AI Act (Reg. 2024/1689)	Biometric (special category); high-risk AI	Mandatory conformity assessment; data governance under Art. 10	Overlapping MDR and AI Act obligations
United States	HIPAA + FDA SaMD + PCCP guidance	PHI under HIPAA; no unified biometric category	Predetermined change control plans for ML models	Fragmented state-level biometric laws
China	PIPL + medical device regulations	Personal sensitive information	Algorithmic regulation; data localization mandates	Cross-border data transfer restrictions
Japan	APPI + PMD Act	Personal information requiring careful handling	Sector-specific AI guidelines	Harmonization with international standards
India	DPDP Act 2023 + CDSCO	Digital personal data	Evolving; implementation rules pending	Regulatory infrastructure gaps for AI-SaMD
Many LMICs	Variable; often no dedicated AI framework	Often unspecified	Limited or absent	No local regulatory pathway for voice AI

Compiled from Jayasree et al. (2025), Kalodanis et al. (2025), and Niemiec et al. (2025). PIPL = Personal Information Protection Law; APPI = Act on Protection of Personal Information; PMD = Pharmaceutical and Medical Devices; DPDP = Digital Personal Data Protection; CDSCO = Central Drugs Standard Control Organisation; SaMD = Software as a Medical Device; PCCP = Predetermined Change Control Plan.

3.5 Acoustic Biomarkers in Low- and Middle-Income Countries

The strongest evidence for LMIC deployment comes from tuberculosis screening. A 2025 study conducted in Zambia enrolled 512 participants at two hospitals, obtained 500 usable cough recordings, and fine-tuned deep-learning classifiers based on pre-trained speech foundation models on three-second audio clips. The best multimodal model, which incorporated demographic and clinical features alongside cough acoustics, achieved 90.3% sensitivity and 73.1% specificity for distinguishing bacteriologically confirmed TB from all other participants—performance the authors reported as meeting the WHO target product profile (TPP) benchmarks for TB screening tools (Ma et al., 2025). The WHO's 2025 update to its TB screening TPPs is explicitly technology-agnostic and envisions any tool that can meet specified sensitivity and specificity thresholds in diverse operational settings (World Health Organization, 2025). The broader TB biomarker literature, however, has historically been heterogeneous and difficult to replicate independently (MacLean et al., 2019), so the Zambia result should be read as encouraging evidence rather than a deployment-ready product.

For childhood pneumonia, the evidence is promising but thin. A study in Malawi tested an AI digital stethoscope algorithm in hospitalized children and reported 80.3% sensitivity and 87.2% specificity at the chest-position level for detecting abnormal lung sounds (Hoekstra et al., 2025). Combined with the 2025 pediatric cough-AI review noted above (Abdelhalim et al., 2025), pneumonia emerges as one of the most plausible use cases for audio biomarkers in resource-limited pediatric care, though it remains a validation problem rather than a solved clinical workflow.

A cross-domain adaptation study used CycleGAN to translate cough audio features across heterogeneous recording conditions within a TB dataset, reporting improved generalizability after harmonization (Pande et al., 2025). A separate fairness-aware COVID-19 cough-detection study found that fairness constraints could be incorporated with minimal loss in predictive accuracy (Kostavasili et al., 2025).

Both results point in the same direction: these systems are most credible when treated as triage or decision-support tools that have been tested across languages, devices, and sites—not as universal diagnostic engines.

The risk of what some scholars call "data colonialism" is not hypothetical. One analysis frames the colonial legacy in global health governance as a structural pattern in which data flows from vulnerable populations in the Global South to institutions in the Global North without adequate benefit-sharing (Latif & Otieno, 2025). Another applies a hermeneutical-injustice framework to voice AI at national borders, arguing that such systems center dominant Western knowledge structures in ways that can marginalize the very communities they claim to serve (Maitra, 2025). For LMIC deployment of voice and cough biomarkers, these critiques translate into three non-negotiable conditions: local governance over data, adequate language coverage, and data sovereignty mechanisms that ensure value returns to the communities contributing recordings.

The practical architecture for equitable LMIC deployment is beginning to take shape, even if no single project has yet realized it fully. On-device inference using lightweight models avoids the need to transmit raw audio to remote servers, reducing both privacy risk and connectivity dependence. Federated learning allows model improvement across sites without centralizing sensitive recordings. Open-source acoustic toolkits—integrated with existing health-information platforms like DHIS2 or OpenMRS—could lower the barrier to adoption in settings where proprietary software is neither affordable nor appropriate. None of these technical solutions, however, substitutes for the institutional work of establishing local ethics review, community consent processes, and benefit-sharing agreements that treat voice data as a collective resource rather than a freely extractable input to commercial AI pipelines.

4. Discussion

Read together, the five thematic areas outlined above converge on a field that is rich in proof-of-concept work, growing in sociotechnical self-awareness, but still some distance from responsible large-scale deployment. Several cross-cutting observations deserve emphasis.

First, the equity and privacy signal is, at present, stronger than the clinical-deployment signal. The studies documenting structural language bias, adversarial vulnerability of de-identification systems, and identity leakage from anonymized speech are methodologically tighter and more immediately actionable than the clinical classification studies, many of which still suffer from small samples, single-center designs, and limited external validation. This asymmetry matters because it means that the sociotechnical risks of premature deployment are better characterized than the clinical benefits of timely deployment.

Second, the biophysiological foundations reviewed here suggest that voice and cough signals are genuinely informative—but for reasons that also constrain their diagnostic specificity. Voice measures like HNR, CPP, and F0 variability respond to any pathology affecting laryngeal mechanics, vocal-tract configuration, or respiratory drive. That gives them broad sensitivity but also makes them vulnerable to confounding by age, sex, smoking status, recording conditions, and emotional state. Similarly, cough acoustics encode information about secretion, airway geometry, and glottic function simultaneously, which is clinically rich but analytically challenging to disentangle.

Third, singing-based biomarkers and language-agnostic acoustic features (sustained vowels, cough) represent the most promising path toward equitable cross-linguistic use, but they are currently at the hypothesis stage. The 2025–2026 literature supports singing as a standardizable, culturally adaptable task that may probe cognitive and emotional function through motor integration; it does not yet demonstrate that singing-derived features bypass linguistic inequity in externally validated clinical studies. Until such validation exists, the field should avoid treating singing as a solved cross-cultural solution.

Fourth, the regulatory landscape in 2025–2026 is fragmented in ways that matter for multi-country deployment. The EU AI Act classifies medical AI as high risk and imposes data-governance and transparency requirements that operate alongside existing medical-device regulations. The United States relies on the FDA's Software-as-a-Medical-Device framework and predetermined change-control plans. Many LMICs lack dedicated regulatory infrastructure for medical AI altogether. For voice biomarker developers operating across jurisdictions, the practical consequence is that compliance must be designed in from the beginning, not retrofitted after a model has been validated in a single regulatory environment.

Fifth, the tuberculosis use case in Zambia is the single most compelling LMIC result in the current literature, but it also illustrates the limitations of extrapolation. The study enrolled approximately 500 participants at two urban hospitals, which is enough to demonstrate feasibility but not enough to claim generalizability across the geographic, linguistic, and epidemiological diversity of high-burden TB settings.

Future work will need to benchmark cough-based AI against chest X-ray with computer-aided detection and molecular diagnostics in head-to-head comparisons, and to assess model robustness under truly varied field conditions.

Sixth, the overlap between voice and cough biomarkers is underappreciated in the clinical literature, where these two signal types are typically studied by separate research communities. As the laryngoplasty study demonstrates, glottic competence shapes both voice quality and cough acoustics. Neurodegenerative conditions like Parkinson's disease affect both phonation and the cough reflex. Respiratory conditions alter both breathing patterns—audible in sustained vowels—and cough morphology. A more integrated research agenda that treats voice and cough as complementary windows onto shared physiology, rather than as independent biomarker modalities, could improve diagnostic specificity and reduce the feature-redundancy problem that plagues current multi-signal classification systems.

Seventh, the dataset-quality findings reported by Mahapatra and Mahapatra (2026) should be read as a systemic warning. If the publicly available datasets that underpin voice-biomarker research are findable but poorly accessible, interoperable, and reusable, then even well-designed classification studies risk building on fragile foundations. The FAIR principles—originally articulated for genomic data—need to be operationalized for audio health data in ways that account for the unique re-identification risks of voice. This means not only publishing datasheets and model cards, but also developing data-access governance structures that balance open science with speaker privacy.

Finally, the scoping review by Cunningham et al. (2025) on ASR for African American English speakers makes a point that extends well beyond that specific population: the deficit is not just in data or models, but in who gets to define what counts as "correct" speech. When speech-biomarker systems are built on a normative assumption about how language sounds, they risk pathologizing variation. The same pitch-flattening that signals depression in one cultural context may reflect perfectly healthy communication norms in another. Acoustic features that carry diagnostic information in English may carry lexical information in Mandarin. Resolving this requires not only diverse training data, but interdisciplinary collaboration between clinicians, computational linguists, and the communities being assessed.

Several limitations of the present review should be noted. Although the search spanned six major bibliographic databases and was supplemented by manual reference-list screening, relevant work indexed only in regional or non-English databases may have been missed. As a narrative review, the study selection was not governed by a pre-registered protocol, and a systematic review or meta-analysis would provide stronger evidence on diagnostic accuracy. The field is evolving rapidly, and findings reported here reflect the literature available through early 2026.

5. Conclusions

Voice and cough are credible, low-cost, and potentially scalable biomarkers of disease, but the 2025–2026 literature makes clear that their clinical translation depends on three conditions being met in parallel. First, methodological standardization—of recording protocols, feature-extraction pipelines, and validation benchmarks—is essential to move the field beyond the current patchwork of incomparable classification studies. Second, linguistic and demographic equity must be treated as a design requirement rather than a post-hoc evaluation metric; the documented bias against tonal-language speakers and the failure of current de-identification systems demonstrate that fairness cannot be assumed. Third, governance frameworks must recognize voice as sensitive biometric medical data, subject to the most protective tier of privacy regulation and to adversarial security testing before deployment.

The most advanced LMIC use case—cough-based tuberculosis screening—has reached WHO target-product-profile thresholds in a single-country study, but its extension to diverse settings will require the same three conditions: local data, local governance, and robust cross-site validation. Researchers, clinicians, regulators, and the communities whose voices are being recorded should jointly shape the standards now, before the field's trajectory is determined solely by commercial actors or by the inertia of data already collected under imperfect consent regimes.

Two observations for future research may be useful. First, the most promising path toward linguistically equitable biomarkers runs through tasks that minimize dependence on language-specific phonetics: sustained vowels, cough, and singing appear better suited to cross-cultural standardization than connected speech, though none has yet been validated as genuinely universal. Second, the adversarial security literature reviewed here suggests that clinical voice-AI systems will need to be red-teamed as rigorously as any other biometric authentication system—and that the threat model should include not only identity theft but also the generation

of synthetic pathological speech designed to manipulate diagnostic outcomes. How the field responds to these challenges in the next two to three years will determine whether voice and cough biomarkers fulfill their considerable promise or remain permanently stalled between laboratory performance and clinical reality.

Author Contributions:

Conceptualization: Wiktor Adamiec.

Methodology: Wiktor Adamiec, Damian Danilczuk, Michał Dyś, Jerzy Buszko, Przemysław Piterak, Marta Bajkowska-Piterak, Monika Szlachta-Gubernat, Nicol Baran, Adrianna Buż.

Formal analysis: Wiktor Adamiec, Damian Danilczuk, Michał Dyś, Jerzy Buszko, Przemysław Piterak, Marta Bajkowska-Piterak, Monika Szlachta-Gubernat, Nicol Baran, Adrianna Buż.

Investigation: Wiktor Adamiec, Damian Danilczuk, Michał Dyś, Jerzy Buszko, Przemysław Piterak, Marta Bajkowska-Piterak, Monika Szlachta-Gubernat, Nicol Baran, Adrianna Buż.

Resources: Wiktor Adamiec, Damian Danilczuk, Michał Dyś, Jerzy Buszko, Przemysław Piterak, Marta Bajkowska-Piterak, Monika Szlachta-Gubernat, Nicol Baran, Adrianna Buż.

Writing — original draft preparation: Wiktor Adamiec, Damian Danilczuk, Michał Dyś.

Writing — review and editing: Damian Danilczuk, Michał Dyś.

Visualization: Wiktor Adamiec.

Supervision: Damian Danilczuk, Michał Dyś.

Project administration: Wiktor Adamiec.

All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

Declaration of Generative AI and AI-assisted Technologies in the Writing Process: In preparing this work, the authors used ChatGPT (OpenAI) exclusively for language editing and grammatical correction of the manuscript, and Claude 4.6 Sonnet (Anthropic) to format and generate the tables and figures based on the provided data. Tools were not involved in data collection, analysis, interpretation of results, or the formulation of conclusions. All substantive content was generated, reviewed, and approved by the human authors, who accept full responsibility for the integrity of the publication.

REFERENCES

1. Abdelhalim, A. A. I., Osman, H. M. E. M., Sadaka, S. I. H., Mohammed, M. A. Y., Eissa, E., Abdalla, R., & Hassan, D. H. M. (2025). Predictive ability of artificial intelligence algorithms in pediatric respiratory disease diagnosis using cough sounds: A systematic review. *Cureus, 17*. https://doi.org/10.7759/cureus.88457
2. Aloradi, A., Gaznepoglu, U. E., Habets, E., & Tenbrinck, D. (2025). VoxATtack: A multimodal attack on voice anonymization systems. In *Proceedings of WASPAA 2025*. IEEE. https://doi.org/10.1109/WASPAA66052.2025.11230920
3. Bouguettaya, A., Reeves, K., Tat, P., & Tan, J. Y. (2025). The effect of group singing interventions on health biomarkers in people with stigmatising conditions: A scoping review protocol. *Systematic Reviews, 14*, Article 171. https://doi.org/10.1186/s13643-025-02971-4
4. Briganti, G., & Lechien, J. R. (2025). Speech and voice quality as digital biomarkers in depression: A systematic review. *Journal of Voice*. https://doi.org/10.1016/j.jvoice.2025.05.002
5. Calvache, C., Soláque, L., Velasco, A., & Peñuela, L. (2021). Biomechanical models to represent vocal physiology: A systematic review. *Journal of Voice, 37*. https://doi.org/10.1016/j.jvoice.2021.02.014
6. Cunningham, J. L., Adjagbodjou, A. D., Basoah, J., Jawara, J., Kadoma, K., & Lewis, A. (2025). Toward responsible ASR for African American English speakers: A scoping review of bias and equity in speech technology. *arXiv*. https://doi.org/10.48550/arXiv.2508.18288
7. Galvosas, M., & Small, P. M. (2025). The value of continuous cough monitoring: A narrative review. *Journal of Thoracic Disease, 17*. https://doi.org/10.21037/jtd-2025-876
8. Guo, C., Chen, L., Li, Z., Lee, K.-A., Ling, Z., & Guo, W. (2024). On the generation and removal of speaker adversarial perturbation for voice-privacy protection. In *Proceedings of SLT 2024*. IEEE. https://doi.org/10.1109/SLT61566.2024.10832243
9. Hegde, S., Sreeram, S., Alter, I. L., Shor, C., Valdez, T. A., Meister, K. D., & Rameau, A. (2024). Cough sounds in screening and diagnostics: A scoping review. *Laryngoscope, 134*(2). https://doi.org/10.1002/lary.31042
10. Hoekstra, N. E., Chagomerana, M., Smith, Z. M., Kala, A., McLane, I., Verwey, C., Olson, D., Buck, W., Mulindwa, J., Gaudio, A., Kapoor, S., Schuh, H. B., Chiume, M., FitzGerald, E., Elhilali, M., Mvalo, T., Hosseinipour, M. C., & McCollum, E. D. (2025). Performance of an artificial intelligence algorithm for interpreting lung sounds from children hospitalised with pneumonia in Malawi. *Journal of Global Health, 15*, 04264. https://doi.org/10.7189/jogh.15.04264

11. Hossain, M. A., Traini, E., & Amenta, F. (2025). Machine learning applications for diagnosing Parkinson's disease via speech, language, and voice changes: A systematic review. *Inventions, 10*(4), 48. https://doi.org/10.3390/inventions10040048
12. Jain, N., Amate, G. A., Alapati, P. R., Bansode, G., Musunoori, S., & B, S. (2025). Real-time accent adaptation in English speech interfaces using federated deep learning across multilingual datasets. In *Proceedings of ResgenXAI 2025*. IEEE. https://doi.org/10.1109/ResgenXAI64788.2025.11343996
13. Jayasree, P., Gandhi, K. B., & Sunny, B. (2025). The data privacy protection laws impacting AI medical devices in the US, EU, China, Japan, and India. *Asia Pacific Journal of Medical Informatics, 2*(3). https://doi.org/10.70818/apjmi.v02i03.030
14. Jiang, Y., Li, Z., Hu, W., Kong, Y., Wang, X., Zhan, X., Lu, Y., Ye, P., Du, J., He, W., & Tai, J. (2025). Value of novel quantitative acoustic parameters based on mel frequency cepstral coefficients in pediatric vocal cord nodules and laryngopharyngitis. *Journal of Voice*. https://doi.org/10.1016/j.jvoice.2025.06.017
15. Kalodanis, K., Feretzakis, G., Rizomiliotis, P., Verykios, V., Papapavlou, C., Koutsikos, I., & Anagnostopoulos, D. (2025). Data governance in healthcare AI: Navigating the EU AI Act's requirements. *Studies in Health Technology and Informatics*. https://doi.org/10.3233/SHTI250050
16. Kan, X., Xiao, Y., Yang, T.-J., Chen, N., & Mathews, R. (2024). Parameter-efficient transfer learning under federated learning for automatic speech recognition. *arXiv*. https://doi.org/10.48550/arXiv.2408.11873
17. Kim, J., Yu, J., Kwon, M., & Kim, J. (2025). FairASR: Fair audio contrastive learning for automatic speech recognition. In *Proceedings of Interspeech 2025*. https://doi.org/10.21437/interspeech.2025-2590
18. Kostavasili, D., Ganitidis, T., Athanasiou, M., & Nikita, K. S. (2025). Fairness-aware deep learning model for COVID-19 detection from cough audio recordings. In *Proceedings of BIBE 2025*. IEEE. https://doi.org/10.1109/BIBE66822.2025.00045
19. Koudounas, A., Pastor, E., Mazzia, V., Giollo, M., Gueudre, T., Reale, E., Cagliero, L., Cumani, S., de Alfaro, L., Baralis, E., & Amberti, D. (2025). Privacy preserving data selection for bias mitigation in speech models. In *Proceedings of ACL 2025: Industry Track*. https://doi.org/10.18653/v1/2025.acl-industry.52
20. Kreiman, J. (2023). Why we talk about voices as we do. *Journal of the Acoustical Society of America, 153*(3). https://doi.org/10.1121/10.0018224
21. Landry, V., Matschek, J., Pang, R., Munipalle, M., Tan, K., Boruff, J., & Li-Jessen, N. Y. (2025). Audio-based digital biomarkers in diagnosing and managing respiratory diseases: A systematic review and bibliometric analysis. *European Respiratory Review, 34*. https://doi.org/10.1183/16000617.0246-2024
22. Latif, L., & Otieno, H. (2025). The colonial legacy of power, profit, and prejudice in global health governance. *Modern Africa: Politics, History and Society, 13*(2). https://doi.org/10.26806/modafr.v13i2.259
23. Lee, K. K., Davenport, P., Smith, J., Irwin, R., McGarvey, L. P., Mazzone, S., & Birring, S. (2021). Global physiology and pathophysiology of cough: Part 1. Cough phenomenology: CHEST guideline and expert panel report. *CHEST, 159*(1), 282–293. https://doi.org/10.1016/j.chest.2020.08.2086
24. Ma, N., Mirheidari, B., Brown, G. J., Sanjase, N., Maimbolwa, M. M., Chifwamba, S., Muzazu, S. G. Y., Muyoyeta, M., & Kagujje, M. (2025). Deep learning for tuberculosis screening in a high-burden setting using cough analysis and speech foundation models. *arXiv*. https://doi.org/10.48550/arXiv.2509.09746
25. MacLean, E., Broger, T., Yerlikaya, S., Fernández-Carballo, B., Pai, M., & Denkinger, C. (2019). A systematic review of biomarkers to detect active tuberculosis. *Nature Microbiology, 4*, 748–758. https://doi.org/10.1038/s41564-019-0380-2
26. Mahapatra, I., & Mahapatra, N. R. (2026). Systematic FAIRness assessment of open voice biomarker datasets for mental health and neurodegenerative diseases. In K. Ekštein et al. (Eds.), *Text, speech, and dialogue: TSD 2025* (Lecture Notes in Computer Science, Vol. 16029, pp. 356–368). Springer. https://doi.org/10.1007/978-3-032-02548-7_30
27. Maitra, S. (2025). Voice AI and hermeneutical injustice at the border. In *Proceedings of AIES 2025*. AAAI/ACM. https://doi.org/10.1609/aies.v8i3.36787
28. Mootassim-Billah, S., Schoentgen, J., De Bodt, M., Roper, N., Digonnet, A., Le Tensorer, M., Van Nuffelen, G., & Van Gestel, D. (2023). Acoustic analysis of voluntary coughs, throat clearings, and induced reflexive coughs in a healthy population. *Dysphagia, 39*, 195–209. https://doi.org/10.1007/s00455-023-10574-1
29. Niemiec, E., Minssen, T., Boland, P., McEntee, P., & Cahill, R. (2025). Legal and regulatory challenges in multi-country data-driven projects developing and validating medical AI systems in the EU. *Expert Review of Medical Devices, 22*. https://doi.org/10.1080/17434440.2025.2531295
30. Pande, M., Patange, A. P., Rastogi, S., Yadav, S., Yulduz, U., & Odamova, U. (2025). Cross-domain generative translation of cough audio biomarkers for tuberculosis screening across diverse global populations. *Indian Journal of Tuberculosis*. https://doi.org/10.1016/j.ijtb.2025.11.008
31. Pedersen, M. (2025). Minimum standards for voice-related biomarkers in speech and singing. *Journal of Clinical Case Reports, Medical Images and Health Sciences, 11*(5). https://doi.org/10.55920/JCRMHS.2025.11.001504

32. Rameau, A., Andreadis, K., German, A., Lachs, M., Rosen, T. E., Pitzrick, M. S., Symes, L., & Klinck, H. (2022). Changes in cough airflow and acoustics after injection laryngoplasty. *Laryngoscope, 133*(7). https://doi.org/10.1002/lary.30255
33. Rodrigo, I., & Duñabeitia, J. A. (2025). Listening to the mind: Integrating vocal biomarkers into digital health. *Brain Sciences, 15*(7), 762. https://doi.org/10.3390/brainsci15070762
34. Roll, N., & Graham, C. (2025). Scaling conformation bias in automatic speech recognition. In *Proceedings of SLaTE 2025*. https://doi.org/10.21437/slate.2025-27
35. Schultz, B., & Vogel, A. (2022). A tutorial review on clinical acoustic markers in speech science. *Journal of Speech, Language, and Hearing Research, 65*(10). https://doi.org/10.1044/2022_JSLHR-21-00647
36. Seo, S., Aulov, O., Godil, A., & Mangold, K. (2025). Evaluating identity leakage in speaker de-identification systems. *arXiv*. https://doi.org/10.48550/arXiv.2508.14012
37. Storey, E., Harte, N., & Bell, P. (2025). Language bias in self-supervised learning for automatic speech recognition. *arXiv*. https://doi.org/10.48550/arXiv.2501.19321
38. Timerbulatov, I. F., Savelieva, E. E., Pestova, R. M., Zagidullina, I. I., & Timerbulatov, R. S. (2025). Assessment of the possibility of acoustic voice analysis. *Meditsinskiy Sovet = Medical Council*, (7), 185–190. https://doi.org/10.21518/ms2025-029
39. Wei, S.-L., Liao, Y.-L., Chang, Y.-H., Huang, H.-H., & Chen, H.-H. (2026). Bias in the ear of the listener: Assessing sensitivity in audio language models across linguistic, demographic, and positional variations. *arXiv*. https://arxiv.org/abs/2602.01030
40. World Health Organization. (2025). *Target product profiles for tuberculosis screening tests* (ISBN 9789240113572). WHO Global Programme on Tuberculosis and Lung Health. https://www.who.int/publications/i/item/9789240113572
41. Yang, M., Liu, X., Du, W., Liu, Y., Zhu, W., Bu, Z., Mao, J., Wang, Q., Chen, S., Zhou, M., & Qu, J.-M. (2026). A device-invariant multi-modal learning framework for respiratory disease classification. *npj Digital Medicine*. https://doi.org/10.1038/s41746-026-02445-4
42. Yang, Y., Chen, M., Han, S., Yu, M., Yu, L., Wang, W., Zhang, W., Chen, S., Wang, X., Shan, S., & Wang, Z. (2025). How voice biomarkers have been used in heart failure management: A scoping review. *Current Medical Research and Opinion*. https://doi.org/10.1080/03007995.2025.2579378
43. Zhu, Y., Davoust, A., & Falk, T. H. (2025). DeepSick: Deceiving voice-based diagnostic models with synthetic multilingual pathological speech signals. In *Proceedings of SMC 2025*. IEEE. https://doi.org/10.1109/SMC58881.2025.11343240