

Artificial Intelligence in Hearing Care: From Algorithmic Performance to Clinical Readiness

Bilal Hasan^{1,*} 

¹Department of Otolaryngology, Yusuf Al-Azma Hospital, Damascus, Syria

*Correspondence: bilalhas78@gmail.com (Bilal Hasan)

Submitted: 28 July 2026 Revised: 14 August 2026 Accepted: 8 September 2026 Published: 23 October 2026

Artificial intelligence (AI) is increasingly applied to otoscopy, audiometry, imaging, phenotyping, and referral support in hearing care [1,2]. These developments address a real clinical problem: hearing assessment integrates heterogeneous information, and specialist interpretation is not equally available across clinical settings. Progress in image- and signal-based models has been substantial, yet technical performance on curated datasets does not by itself establish clinical usefulness.

This is a selective clinical commentary rather than a systematic review. The examples are used to illustrate differences in translational maturity across representative diagnostic and referral-oriented applications. The commentary does not attempt to cover all AI applications across hearing rehabilitation, hearing-device fitting, speech-in-noise assessment, audiological rehabilitation, remote hearing care, cochlear-implant programming, or patient-facing digital hearing care. “Clinical readiness” here refers to evidentiary and workflow readiness for clinician-supervised use, rather than regulatory approval or market availability.

This commentary evaluates the clinical readiness of selected AI applications in hearing care, with emphasis on otoscopic image analysis, audiogram interpretation, cochlear-implant referral, sensorineural hearing loss (SNHL) phenotyping, and temporal-bone imaging. For each application, it considers the relationship among the reported evidence, the clinical task proposed for the model, and the safeguards required before the output informs diagnosis, referral, or management. The commentary aims to distinguish promising clinician-supervised uses from applications for which validation, clinical relevance, or implementation evidence remains insufficient.

Otoscopy images represent a plausible early use case because they transform a familiar examination into a bounded image-classification task. Published syntheses describe a growing body of literature using tympanic-membrane photographs to classify otitis media and other middle-ear conditions [3,4,5].

The most defensible role is triage: flagging an image for expert review, supporting a nonspecialist’s referral decision, or facilitating tele-otology workflows. A stand-alone diagnostic role is harder to justify when image quality is poor, the tympanic membrane is obscured, the child is

uncooperative, or the clinical syndrome conflicts with the image. The decisive issue is the error pattern—especially false reassurance in ears that require further examination or treatment—and whether the reference standard reflects expert consensus, tympanometry, operative findings, or longitudinal clinical diagnosis.

Audiograms likewise contain structured clinical information that can be represented through rules, engineered features, or image-based models. AutoAudio provided an early proof of concept for deep-learning interpretation of diagnostic audiograms, demonstrating high performance on held-out-test data but also showing misclassification, particularly of mixed hearing losses [6]. Subsequent studies have explored classification, digitization, mobile support, and referral-oriented decision support [7,8,9].

The available studies support further evaluation of automated audiogram analysis for classification and referral-oriented decision support [6,7,8,9]. They do not establish that an algorithmic label can replace the broader clinical assessment of hearing loss. In routine use, the authors consider that an output should identify the audiometric pattern and its proposed action—for example, possible asymmetric sensorineural loss requiring review or referral—rather than present an unqualified diagnosis. Displaying the input trace and identifying incomplete, implausible, or out-of-domain records would allow the clinicians to judge whether the output is applicable. Symptoms, otoscopy, speech testing, history, and patient goals would still need to inform the final diagnosis and management plan.

Cochlear-implant referral illustrates why the clinical action should be specified before a model is built. Carducci et al. [10] developed models using routine hearing tests to identify adults likely to meet cochlear-implant candidacy thresholds. Framing the task around formal assessment, rather than automated candidacy determination, aligns the output with a defined action and preserves specialist evaluation as the final decision.

The evidentiary position is less mature for SNHL phenotyping and many imaging applications. Unsupervised learning may help disentangle the biological and audiological heterogeneity of SNHL, but mathematically separable clusters are not clinical phenotypes by default. A systematic assessment of adult-SNHL studies identified only seven el-

Table 1. Clinical-readiness matrix for AI in hearing care.

Use case	Near-term clinical role	Evidence signal	Main deployment risk	Minimum threshold before routine use
Otoscopic images/ middle-ear disease	Triage or second review	Multiple systematic reviews; pooled performance is promising [3,4,5].	Spectrum shift, image quality, unsafe false reassurance.	External, device-diverse validation; prespecified escalation pathway; clinician review.
Audiogram interpretation	Preliminary classification, quality check, referral prompt	Proof-of-concept and growing applied literature [6,8,9].	Loss of clinical context; mixed/asymmetric patterns; poor data capture.	Validation on native data and clinically meaningful referral endpoints; visible uncertainty.
Cochlear-implant referral	Identification for formal candidacy assessment	Clinically action-aligned development study [10].	Over- or under-referral; changing candidacy criteria.	Local calibration, monitored referral yield, specialist confirmation.
SNHL phenotyping	Research stratification	Heterogeneous evidence with low overall study quality [11].	Non-replicable clusters treated as disease entities.	Cross-cohort replication; biological/clinical validation; evidence of management benefit.
Temporal-bone/ neurotologic imaging	Pre-annotation, segmentation, measurement support	Active published literature [13,14,15].	Protocol and center shift; hidden correction workload.	Multi-center validation; workflow study; clinician correction and audit trail.

Note: The matrix represents a narrative, clinically oriented appraisal rather than a formal grading of evidence. AI, artificial intelligence; SNHL, sensorineural hearing loss.

igible studies, found no direct within-study comparisons of algorithms, and reported low overall study quality; consequently, it could not determine an optimal model or the actual number of subtypes [11].

Dimitrov and colleagues identified nine phenotypes in a large UK audiogram dataset, but these phenotypes only partly overlapped with earlier findings and varied across source datasets [12]. This result illustrates the instability that can arise when unsupervised methods are applied to heterogeneous audiological datasets. Before SNHL clusters are used in clinical decision-making, studies need to show that they recur across independent datasets, correspond to etiological or clinically meaningful characteristics, and have consequences for management or outcomes.

Temporal-bone and neurotologic imaging offers potential tasks in segmentation, landmarking, volumetry, classification, and longitudinal comparison [13,14,15]. Anatomical complexity and variation in scanners, protocols, contrast, and pathology make generalization central. A segmentation model may reduce manual work, but it does not itself show better treatment planning or patient outcomes. Quality control, measurement standardization, and clinician-supervised pre-annotation are more credible initial roles; prospective comparisons should assess time, correction burden, missed pathology, and cross-center agreement, rather than relying solely on retrospective overlap metrics. The clinical-readiness matrix in Table 1 (Ref. [3,4,5,6,8,9,10,11,13,14,15]) brings together differences in intended use, evidence, risk, and minimum deployment thresholds.

The need to specify an intended clinical use is reflected in reporting guidance for AI-guided decision-support interventions. DECIDE-AI asks investigators to describe the intended users, clinical setting, inputs, outputs, and pathway in which the system will operate; SPIRIT-AI and CONSORT-AI similarly require explicit reporting of the AI intervention and its evaluation [16,17,18]. These elements provide a more testable basis for assessment than a broad statement that a system “classifies hearing loss”. For example, an audiogram tool intended to flag adults for specialist review for possible cochlear-implant assessment can be evaluated against a defined referral process and its consequences for the receiving clinic.

The WHO guidance places these evaluation and reporting requirements within a broader governance context, including accountability, transparency, equity, and appropriate human oversight [19]. In this commentary, these sources are used to inform the implementation considerations below; the specific examples represent the authors’ interpretation of how these principles apply to hearing-care workflows.

External evaluation is particularly relevant when a model developed at one institution is considered for use elsewhere. Internal training and test sets may share equipment, annotation practices, clinical documentation, and case mix, so performance estimates may not transfer to sites with different populations or workflows. DECIDE-AI and related guidance therefore support prospective, context-specific evaluation of AI decision-support systems [16,17,18]. In hearing care, an initial silent deployment

could examine whether required inputs are available, how often the model produces unusable outputs, and whether particular error patterns would be unacceptable before the system influences care.

Human interaction with the output should also be evaluated as part of the intervention rather than treated as a separate technical issue. DECIDE-AI emphasizes the description of human–AI interaction and the management of system failures [16]. For hearing-care applications, this may include image-quality alerts, checks for incomplete audiograms, a clear explanation of the proposed escalation, and documentation of clinician acceptance or override. Whether these features improve safety, efficiency, or clinician confidence should be evaluated within the intended workflow rather than assumed from retrospective model performance.

The outcome measures should reflect the clinical purpose of each application. For otoscopic triage, relevant measures may include the time to appropriate assessment, diagnostic yield of referral, and missed disease. For audiogram support, concordance with expert review and delays in referral are more informative than classification accuracy alone. In imaging, reporting time, correction burden, agreement between readers, and downstream effects on decisions may be relevant. This outcome-focused approach is consistent with guidance that early evaluations of AI decision-support systems should report the clinical pathway and consequences of use, rather than performance metrics in isolation [16,17,18].

Taken together, the cited guidance and the hearing-care examples above indicate that implementation should proceed through a defined clinical task, an appropriate reference process, evaluation in the intended-use setting, and a clear route for clinician confirmation or escalation [16,17,18,19]. Ongoing monitoring is also needed to identify missing data, performance changes, and uneven errors across patient groups. These are not a substitute for showing benefit against the current care pathway; they are conditions for evaluating whether a proposed system can be used safely and usefully.

Equity and governance considerations are addressed explicitly in the WHO guidance, which calls for attention to inclusiveness, accountability, data protection, and human oversight in the use of AI for health [19]. Their application to hearing care requires empirical assessment rather than general assurances. Audiograms, speech measures, images, device metadata, and linked records may reflect differences in access, language, age, comorbidity, and referral pathways. Studies should therefore report the populations represented in development and validation, examine clinically relevant subgroups where data permit, and specify how data are governed and model versions are managed. Clinical organizations also need an explicit allocation of responsibility

when model output and clinician assessment diverge; the operational form of that responsibility will depend on the local service and intended use.

The evidence reviewed here supports different levels of clinical readiness across applications. Systematic reviews of otoscopic image analysis describe a substantial diagnostic literature, although deployment still depends on image quality, reference standards, and the consequences of false reassurance [3,4,5]. Audiogram interpretation and cochlear-implant referral have more clearly specified clinician-supervised tasks, but require prospective evaluation of their effects on referral and care pathways [6,7,8,9,10]. By contrast, the limited and incompletely replicated literature on SNHL phenotyping, together with the predominantly technical evidence base for temporal-bone imaging, does not yet support routine AI-guided clinical decisions [11,12,13,14,15]. The principal gaps are external validation in representative settings, assessment of workflow and patient outcomes, and evidence on how errors and model outputs affect clinicians' decisions.

Availability of Data and Materials

Not applicable.

Author Contributions

BH is the sole contributor to this manuscript. The author confirms sole responsibility for the conception and design of the study; the acquisition, analysis, and interpretation of data; the preparation of the manuscript; and for being accountable for all aspects of the work.

Ethics Approval and Consent to Participate

Not applicable.

Acknowledgment

Not applicable.

Funding

This research received no external funding.

Conflict of Interest

The author declares no conflict of interest.

References

- [1] You E, Lin V, Mijovic T, Eskander A, Crowson MG. Artificial Intelligence Applications in Otolaryngology: A State of the Art Review. *Otolaryngology–head and Neck Surgery: Official Journal of American Academy of Otolaryngology-Head and Neck Surgery*. 2020; 163: 1123–1133. <https://doi.org/10.1177/0194599820931804>

- [2] Rapoport N, Pavelchek C, Michelson AP, Shew MA. Artificial Intelligence in Otolaryngology and Neurology. *Otolaryngologic Clinics of North America*. 2024; 57: 791–802. <https://doi.org/10.1016/j.otc.2024.04.009>
- [3] Habib AR, Kajbafzadeh M, Hasan Z, Wong E, Gunasekera H, Perry C, et al. Artificial intelligence to classify ear disease from otoscopy: A systematic review and meta-analysis. *Clinical Otolaryngology: Official Journal of ENT-UK ; Official Journal of Netherlands Society for Oto-Rhino-Laryngology & Cervico-Facial Surgery*. 2022; 47: 401–413. <https://doi.org/10.1111/coa.13925>
- [4] Song D, Kim T, Lee Y, Kim J. Image-Based Artificial Intelligence Technology for Diagnosing Middle Ear Diseases: A Systematic Review. *Journal of Clinical Medicine*. 2023; 12: 5831. <https://doi.org/10.3390/jcm12185831>
- [5] Rony MAH, Fatema K, Raiaan MAK, Hassan MM, Azam S, Karim A, et al. Artificial Intelligence-Driven Advancements in Otitis Media Diagnosis: A Systematic Review. *IEEE Access*. 2024; 12: 99282–99307. <https://doi.org/10.1109/access.2024.3428700>
- [6] Crowson MG, Lee JW, Hamour A, Mahmood R, Babier A, Lin V, et al. AutoAudio: Deep Learning for Automatic Audiogram Interpretation. *Journal of Medical Systems*. 2020; 44: 163. <https://doi.org/10.1007/s10916-020-01627-1>
- [7] Kassjański M, Kulawiak M, Przewoźny T. Development of an AI-based audiogram classification method for patient referral. *Annals of Computer Science and Information Systems*. 2022; 30: 163–168. <https://doi.org/10.15439/2022f66>
- [8] Kassjański M, Kulawiak M, Przewoźny T, Tretiakow D, Kuryłowicz J, Molisz A, et al. Automated hearing loss type classification based on pure tone audiometry data. *Scientific Reports*. 2024; 14: 14203. <https://doi.org/10.1038/s41598-024-64310-2>
- [9] Tbin M, Ben Salah M. Deep learning for automatic interpretation of pure-tone audiometry using InceptionV3. *The Egyptian Journal of Otolaryngology*. 2025; 41: 144. <https://doi.org/10.1186/s43163-025-00907-2>
- [10] Carducci V, Sanchez-Trigo H, Elgabori Y, Carlson M, Deep N, Romero-Brufau S. Advancing clinical utility of artificial intelligence: lessons from developing a model to predict cochlear implant eligibility. *JAMIA Open*. 2026; 9: ooag024. <https://doi.org/10.1093/jamiaopen/ooag024>
- [11] Dimitrov L, Barrett L, Chaudhry A, Muzaffar J, Lilaonitkul W, Mehta N. Uncovering Phenotypes in Sensorineural Hearing Loss: A Systematic Review of Unsupervised Machine Learning Approaches. *Ear and Hearing*. 2025; 46: 1401–1411. <https://doi.org/10.1097/AUD.0000000000001696>
- [12] Dimitrov L, Lilaonitkul W, Mehta N. Identification of sensorineural hearing loss subtypes using unsupervised machine learning and assessment of their replicability. *Scientific Reports*. 2026; 16: 3774. <https://doi.org/10.1038/s41598-025-33815-9>
- [13] Chawdhary G, Shoman N. Emerging artificial intelligence applications in otological imaging. *Current Opinion in Otolaryngology & Head and Neck Surgery*. 2021; 29: 357–364. <https://doi.org/10.1097/MOO.0000000000000754>
- [14] Alsaleh H. The impact of artificial intelligence in the diagnosis and management of acoustic neuroma: A systematic review. *Technology and Health Care: Official Journal of the European Society for Engineering and Medicine*. 2024; 32: 3801–3813. <https://doi.org/10.3233/THC-232043>
- [15] Spinos D, Martinos A, Petsiou DP, Mistry N, Garas G. Artificial Intelligence in Temporal Bone Imaging: A Systematic Review. *The Laryngoscope*. 2025; 135: 973–981. <https://doi.org/10.1002/lary.31809>
- [16] Vasey B, Nagendran M, Campbell B, Clifton DA, Collins GS, Denaxas S, et al. Reporting guideline for the early-stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI. *Nature Medicine*. 2022; 28: 924–933. <https://doi.org/10.1038/s41591-022-01772-9>
- [17] Liu X, Cruz Rivera S, Moher D, Calvert MJ, Denniston AK, SPIRIT-AI and CONSORT-AI Working Group. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. *Nature Medicine*. 2020; 26: 1364–1374. <https://doi.org/10.1038/s41591-020-1034-x>
- [18] Cruz Rivera S, Liu X, Chan AW, Denniston AK, Calvert MJ, SPIRIT-AI and CONSORT-AI Working Group, et al. Guidelines for clinical trial protocols for interventions involving artificial intelligence: the SPIRIT-AI extension. *Nature Medicine*. 2020; 26: 1351–1363. <https://doi.org/10.1038/s41591-020-1037-7>
- [19] World Health Organization. Ethics and governance of artificial intelligence for health: WHO guidance. World Health Organization: Geneva. 2021.