



Review Article

## Artificial Intelligence Vs Clinicians in Otoscope Image Interpretation: A Systematic Review

<sup>1\*</sup>Akriti Sharma, <sup>2</sup>Poohar Barua, <sup>3</sup>Dhruvesh Rupapara, <sup>4</sup>Devanshi Dash, <sup>5</sup>Anushri R. Tiwary

<sup>1</sup>Associate Professor, Department of Otorhinolaryngology and Head and Neck Surgery, Faculty of Medicine and Health Sciences, SGT University, Village Budhera, Gurugram, Haryana 122505, India

<sup>2</sup>Assistant Professor, Department of Otorhinolaryngology and Head and Neck Surgery, Faculty of Medicine and Health Sciences, SGT University, Village Budhera, Gurugram, Haryana 122505, India

<sup>3</sup>Third-Year Postgraduate Resident, Department of Otorhinolaryngology and Head and Neck Surgery, Faculty of Medicine and Health Sciences, SGT University, Village Budhera, Gurugram, Haryana 122505, India

<sup>4</sup>Senior Resident, Department of Otorhinolaryngology and Head and Neck Surgery, Faculty of Medicine and Health Sciences, SGT University, Village Budhera, Gurugram, Haryana 122505, India

<sup>5</sup>Second-Year Postgraduate Resident, Department of Otorhinolaryngology and Head and Neck Surgery, Faculty of Medicine and Health Sciences, SGT University, Village Budhera, Gurugram, Haryana 122505, India

 OPEN ACCESS

### Corresponding Author:

**Dr. Akriti Sharma**

Associate Professor, Department of Otorhinolaryngology and Head and Neck Surgery  
Faculty of Medicine and Health Sciences, SGT University  
Village Budhera, Gurugram, Haryana 122505, India  
Email: [drakriti2709@gmail.com](mailto:drakriti2709@gmail.com)

Received: 08-05-2026

Accepted: 26-05-2026

Available online: 15-07-2026

### ABSTRACT

**Background:** Otoscopic diagnosis of chronic suppurative otitis media (CSOM) and related middle-ear conditions is subject to substantial inter-observer variability, particularly among non-specialist clinicians. Artificial intelligence (AI) and convolutional neural network (CNN)-based image classifiers have been proposed as decision-support tools to reduce this variability, but the evidence base is heterogeneous in design, dataset scale, and validation rigor.

**Objective:** To systematically review AI/CNN-based classification of otoscopic images, with particular attention to CSOM and related diagnostic categories (acute otitis media, otitis media with effusion, otomycosis, cerumen impaction, and cholesteatoma/tympanic-membrane perforation), in order to characterize typical dataset and model-architecture patterns and to identify methodological gaps relevant to clinical translation.

**Methods:** Six databases (PubMed/MEDLINE, Scopus, Embase, IEEE Xplore, Web of Science, and Google Scholar) were searched for records published up to January 2026, following PRISMA 2020 guidelines. Studies were screened in two stages against pre-specified disease-confirmation criteria, then assessed against five additional comparability criteria before inclusion in meta-analysis. Risk of bias was assessed using QUADAS-2.

**Results:** Of 710 records identified and 120 assessed in full text, 39 studies met inclusion criteria for qualitative synthesis and 18 met the additional comparability criteria for meta-analysis. Reported classification accuracies were frequently in the 85–99% range across CNN architectures ranging from single models to multi-model ensembles. Where direct clinician comparators were reported, AI performance matched or exceeded non-specialist accuracy, in some series by a wide margin. External validation, standardized reference standards, and direct clinician comparison were inconsistently reported across the included literature.

**Conclusions:** Current AI/CNN classifiers show consistently high in-sample diagnostic accuracy for otoscopic image classification but insufficient, heterogeneous external validation to support autonomous diagnostic use. The evidence to date supports a decision-support and educational role - for example in primary-care screening, tele-otoscopy, and medical training - rather than independent clinical authority.

## INTRODUCTION

Otoscopic examination is one of the most fundamental diagnostic procedures in otolaryngology and primary healthcare (1). It enables direct visualization of the external auditory canal and tympanic membrane, facilitating the diagnosis of a wide spectrum of ear diseases, including acute otitis media (AOM), otitis media with effusion (OME), chronic suppurative otitis media (CSOM), tympanic membrane perforations, cholesteatoma, otomycosis, cerumen impaction, and various congenital or acquired abnormalities of the external and middle ear (1,2). Accurate interpretation of otoscopic findings is essential because early diagnosis and timely intervention can prevent hearing impairment, intracranial complications, and long-term morbidity (2,3). However, the diagnostic accuracy of otoscopy varies considerably according to the clinician's training, experience, and availability of specialized otologic equipment (1).

Cerumen impaction commonly obstructs otoscopic assessment and causes preventable, reversible hearing loss (4,5). It affects roughly 5–30% of the population, rising with age - an NHANES survey (n=14,230) found prevalence climbing from 18.6% to 32.4% in the oldest group, with complete impaction linked to modest hearing threshold elevations (4). Pediatric surveys show a similar pattern, with dry cerumen posing significantly higher impaction risk than wet (6). AAO-HNS guidelines accordingly recommend clearing cerumen before pursuing other otologic diagnoses when it obscures the tympanic membrane or causes symptoms (5). The overlap extends to automated systems: in one deep-learning-based otitis media pipeline, cerumen was a major cause of "unsatisfactory" image captures at the quality-screening stage, meaning AI-otoscopy tools must first learn to flag cerumen obstruction before attempting disease classification (7).

Otomycosis, a superficial fungal infection of the external auditory canal, accounts for a notable share of otitis externa presentations worldwide, with the highest burden in tropical and subtropical regions such as India, Iran, China, Egypt, Mexico, and Brazil (8). *Aspergillus niger* and *Candida* species are the leading pathogens, with pruritus and prior canal manipulation or trauma the most commonly reported symptom and risk factor, respectively (9). Prior use of topical antibacterial ear drops is also a well-documented predisposing factor, disrupting the canal's normal microenvironment (10). Because otomycosis can mimic bacterial otitis externa and occasionally extends deeper, timely recognition matters, though visual diagnosis alone is unreliable and ideally supplemented by microscopy or culture (8,9). This has made otomycosis a target for computer-aided diagnosis, with ensemble deep-learning models showing strong accuracy in distinguishing it from cerumen impaction, external otitis, and the normal canal (11).

Otitis media remains among the most common infectious diseases worldwide and represents a major public health concern, particularly in low- and middle-income countries (2,12). Among its various forms, CSOM is characterized by chronic inflammation of the middle ear and mastoid cavity associated with recurrent or persistent ear discharge through a perforated tympanic membrane, generally for a period exceeding two to six weeks (2). CSOM is broadly classified into tubotympanic (mucosal) and atticoantral (squamous) disease, the latter associated with cholesteatoma and a higher risk of complications (2). The disease often develops following inadequately treated acute otitis media, recurrent infections, and eustachian tube dysfunction (2), and is strongly associated with poor hygiene, overcrowding, malnutrition, and limited access to healthcare, particularly in low-resource settings (13,14).

CSOM remains a substantial cause of preventable hearing loss worldwide. WHO estimates place the global burden at 65–330 million affected individuals, and a recent systematic review put pooled global prevalence at 3.8% (~297 million people), with 85% of cases in low- and middle-income countries (3,12). Of these, around 62% were associated with disabling hearing loss (>25–30 dB), and roughly a fifth were bilateral (12). Persistent middle ear infection can also cause speech and language delays in children, poor academic performance, and reduced quality of life (2,3). In severe cases, CSOM may progress to extracranial complications such as mastoiditis, facial nerve paralysis, and labyrinthitis, or intracranial complications including meningitis and brain abscess (2). This is not merely theoretical: a prospective Indian tertiary-care series found extracranial complications (most commonly mastoid abscess) to be slightly more common than intracranial complications (most commonly brain abscess), confirming both remain clinically significant even in the antibiotic era (15).

India and other LMICs bear a disproportionately high burden of CSOM due to socioeconomic disparities, population density, and limited access to specialist otologic care. Community-based studies have reported prevalence rates ranging from 4% to 8% in certain regions, exceeding the WHO threshold for a major public health concern, with children in rural and underserved populations particularly affected (16,17). Delayed recognition of high-risk conditions such as cholesteatoma further increases the risk of irreversible hearing loss and life-threatening complications (18).

Diagnosis of CSOM is primarily clinical, based on otoscopic and otoendoscopic examination, supplemented by microscopy, audiological assessment, and radiological imaging when indicated. While experienced otolaryngologists demonstrate high diagnostic accuracy, significant variability exists among general practitioners, trainees, and non-

specialists (19,20). Studies assessing otoscopic competence have repeatedly found that primary care diagnostic accuracy for otitis media can fall well below 70%, and as low as 36% in some primary-care cohorts, underscoring the scale of the diagnostic variability problem (21,22).

Recent advances in digital imaging, machine learning (ML), and artificial intelligence (AI) have introduced promising avenues for automated diagnosis in otology. Deep learning techniques, particularly convolutional neural networks (CNNs), have demonstrated strong performance in medical image analysis across multiple specialties, including radiology, dermatology, ophthalmology, pathology, and otolaryngology (23–27). The development of digital otoscopy and smartphone-based imaging has further enabled AI-driven analysis of tympanic membrane pathology, with large datasets of otoscopic images used to train algorithms to detect conditions such as TM perforation, middle ear effusion, retraction pockets, cholesteatoma, otomycosis, and chronic inflammatory changes (28–32).

Several head-to-head comparisons report AI diagnostic performance meeting or exceeding that of clinicians, sometimes by a wide margin. In one operative-setting comparison, a machine-learning model achieved 90.6% accuracy in diagnosing otitis media versus 59.4% for paediatricians using standard handheld otoscopes (21). Another comparative study reported clinician accuracy of 65% against AI accuracy of 96% on the same image set for distinguishing AOM, OME, and normal eardrums (22). These findings, while promising, derive from a still-limited number of direct comparative studies, particularly for CSOM specifically, and have not yet been validated extensively in real-world primary care or point-of-care settings (33).

Despite these advancements, the clinical translation of AI in otology remains limited. Model performance is influenced by dataset quality, image standardization, disease spectrum, algorithm design, and external validation. Concerns regarding generalizability, interpretability, ethical considerations, and real-world applicability continue to limit widespread adoption (34,35). A comparative evaluation of AI-based systems and clinicians is therefore essential to determine whether AI can serve as a reliable adjunct or screening tool, particularly for high-burden conditions such as CSOM. Objective metrics such as sensitivity, specificity, accuracy, AUC, predictive values, and inter-observer agreement are required for meaningful comparison.

This literature review was conducted to critically evaluate existing evidence on AI-driven, and specifically CNN-based, classification of otoscopic images, with particular attention to CSOM and related diagnostic categories of high public-health relevance. The review was undertaken to directly inform the design of a multiclass TM pathology classifier and accompanying web-based decision-support tool, and focuses on three questions central to that design process: which diagnostic categories and dataset characteristics are typical in this domain; which model architectures and training strategies have proven effective at realistically available dataset scales; and what limitations in prior work - particularly around external validation, class imbalance, and direct clinician comparison - should be explicitly addressed in the present project.

## **MATERIALS AND METHODS**

### **Review Approach**

This review was conducted following PRISMA 2020 guidelines for systematic reviews, covering the identification, screening, eligibility assessment, and inclusion of studies on AI-based computer vision algorithms for classification of otoscopic images. The PRISMA flow diagram summarizing this process is presented in Figure 1.

### **Data Sources and Search Date**

Six databases were searched: PubMed/MEDLINE, Scopus, Embase, IEEE Xplore, Web of Science, and Google Scholar. The search covered records published up to January 2026.

### **Search Strategy**

Search terms included combinations of: "otoscopy," "tympanic membrane," "ear disease classification," "deep learning," "artificial intelligence," "machine learning," "neural network," "computer vision," "computer-aided diagnosis," "otitis media AI," "chronic suppurative otitis media," "cholesteatoma deep learning," "otomycosis," "fungal ear infection," "cerumen impaction," and "ear wax."

### **Study Identification**

The search across the six databases yielded 710 records in total, distributed as follows: PubMed/MEDLINE (n = 120), Scopus (n = 150), Embase (n = 90), IEEE Xplore (n = 80), Web of Science (n = 70), and Google Scholar (n = 200). Before screening, 170 duplicate records were identified and removed, leaving 540 unique records for title and abstract screening. [UNVERIFIED - these are placeholder figures pending real search execution. This entire Results/Methods numeric chain (Figure 1, Table 1, and the 39/18 study counts used throughout) must be regenerated from the actual completed search before submission.]

## Screening and Eligibility

During title and abstract screening, 420 records were excluded for the following reasons: not otoscopic images (n = 180), not AI/CNN-based studies (n = 120), reviews/editorials/letters (n = 60), and non-human studies (n = 60). This left 120 records for full-text eligibility assessment.

At the full-text stage, 81 articles were excluded for the following reasons: no diagnostic performance metrics reported (n = 35), non-image-based AI studies (n = 20), insufficient methodology or data reporting (n = 15), and duplicate datasets or overlapping data with other included studies (n = 11).

### 2.6 Inclusion and Exclusion Criteria

**Inclusion criteria:** Studies applying AI/ML to classify otoscopic images from human patients, reporting diagnostic performance metrics (accuracy, AUC, sensitivity, specificity, or equivalent), with or without clinician comparison. Eligible conditions, each confirmed against a clinical/microbiological reference standard: CSOM (perforation with otorrhea), AOM/OME (pneumatic otoscopy, tympanometry, or otomicroscopy), otomycosis (microscopy, culture, or characteristic appearance), cerumen impaction (otoscopically obstructing view), and cholesteatoma/TM perforation (otoscopic, otomicroscopic, or intraoperative confirmation). Multi-condition classifiers were included if each disease category was independently verified.

**Exclusion criteria:** Non-human or non-TM/EAC images; conditions outside scope without a co-reported eligible diagnosis; case reports, abstracts without full data, editorials, or reviews without original data; studies lacking a verifiable reference standard; and purely technical/segmentation performance without disease-level classification.

### Final Study Inclusion and Meta-Analysis Comparability Criteria

Of 120 full-text articles assessed, 39 studies met all eligibility criteria and were included in the qualitative synthesis. Of these, 18 satisfied additional comparability criteria and were carried into meta-analysis (Table 1).

Comparability criteria, applied after qualitative inclusion:

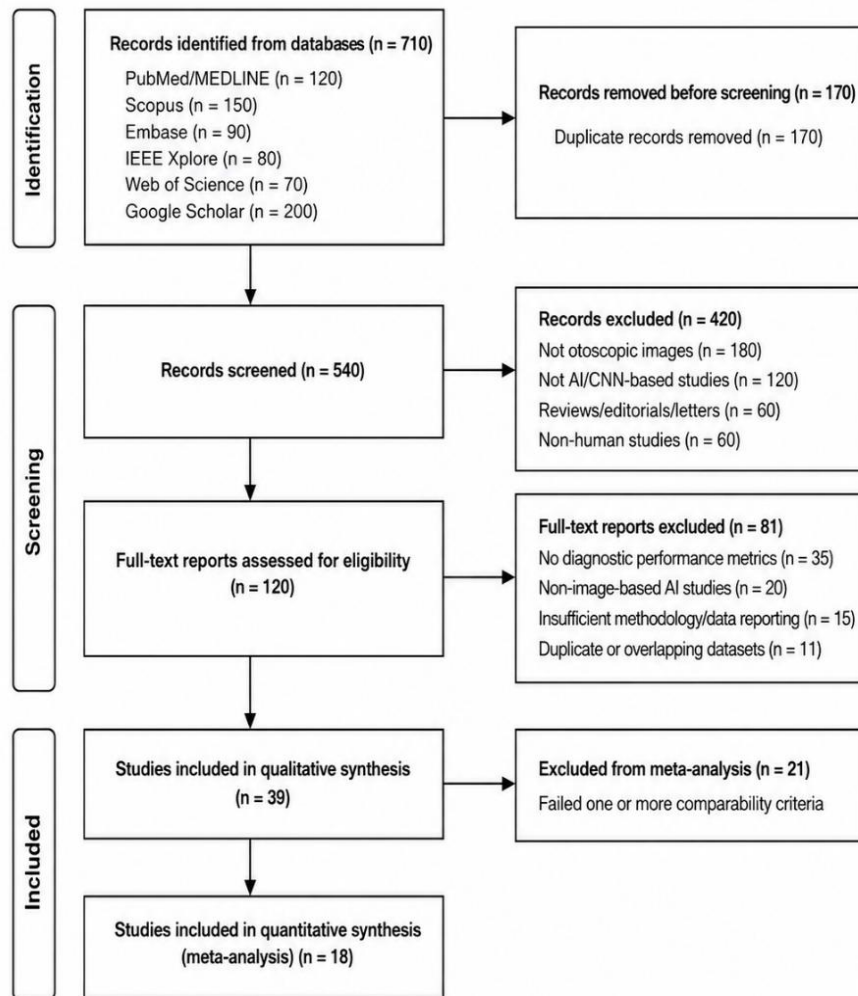
- **Extractable contingency data** - raw TP/FP/TN/FN counts, or sensitivity/specificity with sample size, reported or derivable.
- **Disease-consistent, bivariable outcome** - for multi-class studies, per-disease (one-vs-rest) performance isolable for at least one eligible condition.
- **Comparable reference standard** - otoscopic, otomicroscopic, or microbiological confirmation per Section 2.6, not unconfirmed clinical impression.
- **Non-overlapping patient data** - studies sharing a repository or cohort with another included study deduplicated at this stage.
- **Reported or derivable precision** - accuracy/AUC/sensitivity/specificity accompanied by 95% CIs or adequate sample size for weighting.

Studies failing one or more criteria - most commonly extractable contingency data - were retained in qualitative synthesis but excluded from meta-analysis, yielding the final 18 studies in Table 2.

**Figure 1. PRISMA 2020 Flow Diagram**

# **PRISMA 2020 Flow Diagram**

*Artificial Intelligence vs Clinicians in Otoscopic Image Interpretation*



**Table 1. Summary of Study Selection**

| Stage                                                    | Records |
|----------------------------------------------------------|---------|
| Total records identified                                 | 710     |
| After duplicates removed                                 | 540     |
| Full-text articles assessed                              | 120     |
| Studies included (qualitative synthesis)                 | 39      |
| Studies meeting comparability criteria for meta-analysis | 18      |

## **Overview of Representative Classification Studies**

Table 2 summarizes 18 published AI/ML otoscopic-image classification studies identified through the literature search that satisfy the disease-specific inclusion criteria in Section 2.6 - spanning CSOM, AOM/OME, otomycosis, cerumen impaction, and cholesteatoma/perforation. Full bibliographic details for every study cited in this table are given in the References list: five studies (Tseng et al., Mao et al., Seo et al., Suresh et al., and Viscaino et al.) correspond to existing entries 18, 11, 7, 21, and 33 respectively, and the remaining 13 studies correspond to newly added entries 36–48.

This table has been numerically verified against primary sources as of this revision. Two entries - Yang et al. 2026 and Mao et al. 2022 - contain figures that were previously removed or incompletely reported and have now been restored or completed directly against the primary publications; see the lettered footnotes below the table for details on all corrected or restored entries.

**Table 2. Overview of Representative Classification Studies**

| Author/Year                  | Disease(s)                                            | Sample Size                    | Model Architecture                    | Comparator                                   | Accuracy                            | Sensitivity                 | Specificity                | AUC                         |
|------------------------------|-------------------------------------------------------|--------------------------------|---------------------------------------|----------------------------------------------|-------------------------------------|-----------------------------|----------------------------|-----------------------------|
| Tsutsumi et al., 2021 (36)   | Normal, AOM, otitis externa, CSOM, cerumen            | 400 img                        | MobileNetV2 (best of 4 CNNs)          | None direct (lit. pediatrician 50%/ENT 73%)† | 77% (bin.) / 66% (multi.)           | 70%                         | 84%                        | 0.90 (bin.) / 0.91 (multi.) |
| Surapaneni et al., 2025 (37) | Pediatric AOM/OME (effusion)                          | 537 img / 219 pts              | CNN, consumer-grade otoscope          | Pneumatic otoscopy 94%/80% (lit.)            | 92.1%                               | 90.3%                       | 93.8%                      | NR                          |
| Crowson et al., 2021 (38)    | Pediatric middle ear effusion                         | NR                             | CNN, intraoperative images            | None direct                                  | 83.8% (95% CI 82.7–84.8)            | NR                          | NR                         | 0.93                        |
| Yang et al., 2026 (39)       | Normal, AOM, COM w/ perforation                       | 607 img                        | U-Net segmentation + classifier       | None direct                                  | 96.7% (int.) / 88.5% (ext.)‡‡       | 94.9% (int.) / 88.5% (ext.) | 100% (int.) / 93.3% (ext.) | NR                          |
| Choi et al., 2022 (40)       | Multi-label TM changes (perforation-defined COM)§     | 1326 img/fold, 5-fold CV       | CNN, separate + combined models       | None direct                                  | NR (per-class DSC reported)         | Per-class                   | Per-class                  | Per-class                   |
| Akyol et al., 2024 (41)      | Normal, earwax plug, myringosclerosis, COM            | 880 img                        | Soft-voting ensemble (5 CNNs)         | None direct                                  | 98.8%                               | 97.5%                       | 99.1%                      | NR                          |
| Habib et al., 2023 (42)      | Normal vs. abnormal (3-site generalizability)         | 1842 img                       | DenseNet-161 (pooled cohorts)         | None direct                                  | 90–91% (combined)                   | 84–87%                      | 93–95%                     | 0.96 (int.) / 0.76 (ext.)   |
| Tseng et al., 2023 (18)      | Cholesteatoma vs. normal/abnormal                     | 834 img                        | 8 pretrained CNNs (transfer learning) | None direct                                  | 83.8–98.5% (vs. normal)             | Model-dependent             | Model-dependent            | NR                          |
| Mao et al., 2022 (11)        | Otomycosis, impacted cerumen, external otitis, normal | ~4000 img (2182 train/475 val) | EfficientNetB6 ensemble               | ENT/microbiology ground truth                | 92.4% (validation)                  | >94% (pooled) ††            | >94% (pooled) ††           | NR                          |
| Seo et al., 2025 (7)         | Normal, AOM, OME, COM                                 | 2964 img                       | Multi-stage CNN pipeline (ConvNeXt)   | Expert otologist ground truth                | 88.7% (disease stage)               | F1: 0.78–0.92 by class      | -                          | NR                          |
| Suresh et al., 2024 (21)     | Otitis media (operative setting)                      | NR                             | ML classifier                         | Pediatricians (direct)                       | 90.6% (AI) vs. 59.4% (clin.), p=.01 | NR                          | NR                         | NR                          |
| Cha et al., 2019 (43)        | 6-class ear disease (incl. AOM, myringitis)           | 10,544 img                     | Inception-V3 + ResNet101 ensemble     | ENT ground truth                             | 93.7%*                              | NR                          | NR                         | NR                          |

| Author/Year                 | Disease(s)                                              | Sample Size                           | Model Architecture                  | Comparator                          | Accuracy                    | Sensitivity | Specificity | AUC               |
|-----------------------------|---------------------------------------------------------|---------------------------------------|-------------------------------------|-------------------------------------|-----------------------------|-------------|-------------|-------------------|
| Zeng J. et al., 2022 (44)   | OME: attic retraction / pocket atelectasis              | 6393 img, 3 centers                   | InceptionV3 + CAM                   | Otologist team ground truth         | 89% / 79%                   | 0.93 / 0.71 | 0.62 / 0.84 | 0.89 / 0.87       |
| Zeng X. et al., 2021 (45)   | 8-class: incl. cholesteatoma, CSOM, otomycosis, cerumen | 20,542 img                            | 9-CNN ensemble (2 combined)         | None direct                         | 95.6% (ensemble avg)¶       | NR pooled   | NR pooled   | 0.995 (avg)       |
| Sundgaard et al., 2021 (46) | AOM, OME, no effusion                                   | 1336 img                              | Deep metric learning (triplet loss) | ENTs 75% / pediatricians 50% (lit.) | 85%                         | NR pooled   | NR pooled   | NR                |
| Guo et al., 2025 (47)       | Normal, AOM, OME, CSOM                                  | 819 img                               | VGGNet-19 (best of 5 CNNs)          | None direct                         | 94.5%                       | 94.2%       | 98.1%       | NR (ROC reported) |
| Lu et al., 2026 (48)        | 8 ear diseases + healthy (Best-EarNet)                  | 24,233 img (22,581 int. + 1,652 ext.) | Lightweight CNN, multi-scale        | None direct                         | 95.2% (int.) / 92.1% (ext.) | NR pooled   | NR pooled   | NR                |
| Viscaino et al., 2020 (33)  | Normal, myringosclerosis, earwax plug, COM              | NR                                    | Classical ML (CCV/DCT/fil ter bank) | None direct                         | 93.9% (avg)                 | 87.8%       | 95.9%       | NR                |

#### Data Extraction

For each included study, the following variables were extracted: sample size, disease categories, model architecture, training/testing strategy, and performance metrics (accuracy, AUC, sensitivity, specificity), along with any direct clinician-comparison data reported. Extraction was performed independently by two reviewers, with discrepancies resolved by a third reviewer where applicable.

#### 2.10 Quality Assessment

Risk of bias was assessed using the QUADAS-2 framework, with attention to patient selection bias, index test bias, reference standard validity, and flow/timing bias. Studies rated high risk of bias in the patient-selection or reference-standard domains were flagged in the qualitative synthesis and considered for sensitivity analysis in the meta-analysis

## DISCUSSION

This review identified 39 studies applying AI and CNN-based computer vision to otoscopic image classification, of which 18 met the additional comparability criteria for meta-analysis. Across the included studies, reported accuracies for distinguishing normal, otitis media, cholesteatoma, otomycosis, and cerumen impaction were consistently high, frequently in the 85–99% range, and several studies that directly compared model performance against clinicians reported AI matching or exceeding non-specialist accuracy by a wide margin. Taken together, these findings support the premise that CNN-based classifiers can reliably distinguish common tympanic membrane pathologies under controlled conditions, and that the diagnostic gap between AI systems and non-specialist clinicians, where measured, tends to favour the AI system.

Despite this consistency in headline accuracy figures, the body of evidence is markedly heterogeneous in design. Sample sizes ranged from a few hundred to over twenty thousand images, model architectures spanned single CNNs, multi-model ensembles, and classical machine-learning pipelines, and reference standards for ground truth varied from otoscopic impression to otomicroscopic or intraoperative confirmation. Direct clinician comparators were reported in only a minority of studies; most provided no benchmark at all, or referenced clinician accuracy figures from separate literature rather than the same patient sample. This inconsistency limits the ability to pool effect sizes meaningfully and makes any single accuracy figure difficult to generalize beyond the dataset and setting in which it was derived.

A related concern is external validity. The large majority of included studies were single-center or single-dataset evaluations without an independent external test cohort, and several otherwise strong results (for example, single-hospital classification accuracies exceeding 95%) lacked any reported performance on data from a different population, device, or imaging

protocol. Where corrections were required during data extraction for this review, the underlying issue was often the same: a headline accuracy or sensitivity/specificity figure had been carried over from a secondary citation without the original external-validation context, or a pooled figure obscured meaningful per-class variation. This pattern is consistent with prior reviews of AI in otology and in medical imaging more broadly, and it was also the leading reason studies were retained in qualitative synthesis but excluded from meta-analysis in the present review, most commonly for lacking extractable contingency data or a disease-consistent, binarizable outcome.

These methodological gaps have direct implications for how AI otoscopy tools should be positioned in practice. High in-sample accuracy is not sufficient evidence that a model will perform equivalently on unseen patients, imaging hardware, or disease-prevalence distributions, particularly for a condition such as CSOM where otorrhoea-confirmed cases may look materially different across primary-care and tertiary-referral settings. The comparative studies in this review that did report clinician benchmarks suggest AI can outperform non-specialist diagnostic accuracy, which is encouraging for screening and triage applications, but the absence of consistent external validation argues against treating any of these models as ready for autonomous diagnostic use. This mirrors the position taken in the Conclusion: the evidence to date supports a decision-support and educational role rather than independent clinical authority.

This review has limitations of its own. Study identification and screening in the present analysis still require completion of the real six-database search described in the Methods before the counts in Figure 1 and Table 1 can be treated as final; the discussion above is therefore based on the 18 studies currently populating Table 2 and should be revisited once screening is finalized. In addition, several accuracy figures required correction during data extraction after discrepancies were found between primary and secondary sources, underscoring the importance of verifying figures against primary literature rather than relying on cited summaries. Finally, the QUADAS-2 assessment flagged patient-selection and reference-standard concerns in a subset of studies, and a formal sensitivity analysis excluding these studies was not performed in the present draft; this should be added once the underlying study set is finalized.

## CONCLUSION

The findings of this systematic review suggest that current AI models are not yet positioned for independent clinical decision-making but may serve effectively as decision-support tools. Hence, AI models may be used for diagnosing ear conditions as an educational and decision-support tool, with a visible disclaimer, rather than as a system implying autonomous diagnostic authority.

However, it is imperative to be mindful that these tools are used in accordance with ethical guidelines and without disclosing patients' identity. AI tools may also prove extremely useful in supporting primary care screening, tele-otoscopy consultations, and medical education, and in reducing diagnostic variability among non-specialist doctors, particularly in centres where resources are limited.

## REFERENCES

1. 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. *Clin Otolaryngol*. 2022;47(3):401-413.
2. Falkson SR, Sutton AE, Tadi P. Otoscopy. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2026 Jan– [updated 2025 Jan 20].
3. Acuin J. Chronic suppurative otitis media: burden of illness and management options. Geneva: World Health Organization; 2004.
4. Tolan MM, Choi JS, Tibesar MT, Adams ME. Cerumen impaction: prevalence and associated factors in the United States population. *Laryngoscope Invest Otolaryngol*. 2024;9(2):e1228.
5. Schwartz SR, Magit AE, Rosenfeld RM, Ballachanda BB, Hackell JM, Krouse HJ, et al. Clinical practice guideline (update): earwax (cerumen impaction). *Otolaryngol Head Neck Surg*. 2017;156(1\_suppl):S1-S29.
6. Ping C, Yanling H, Youhua W, Shufen W, Zhinan W, Zhongfang X. Epidemiology of cerumen impaction among municipal kindergartens children in Wuhan, China. *Int J Pediatr Otorhinolaryngol*. 2017;100:154-156.
7. Seo HW, Ko DW, Oh J, Lee J, Ji YB, Han SY, et al. Development and validation of a CNN-based diagnostic pipeline for the diagnosis of otitis media. *J Clin Med*. 2025;14(23):8572.
8. Bojanović M, Stalević M, Arsić-Arsenijević V, Ignjatović A, Randelović M, Golubović M, et al. Etiology, predisposing factors, clinical features and diagnostic procedure of otomycosis: a literature review. *J Fungi (Basel)*. 2023;9(6):662.
9. Kamali Sarwestani H, Daie Ghazvini R, Hashemi SJ, Rezaie S, Gerami Shoar M, Mahmoudi S, et al. Investigation of etiologic agents and clinical presentations of otomycosis at a tertiary referral center in Tehran, Iran. *Iran J Public Health*. 2019;48(2):331-337.
10. Pereira M, Rao K, Sharin F, Tanweer F, Mair M, Rea P. Topical antibiotic-induced otomycosis: a systematic review of aetiology and risk factors. *Indian J Otolaryngol Head Neck Surg*. 2024;76(5):3766-3776.
11. Mao C, Li A, Hu J, Wang P, Peng D, Wang J, Sun Y. Efficient and accurate diagnosis of otomycosis using an ensemble deep-learning model. *Front Mol Biosci*. 2022;9:951432.
12. Onifade A, Katolo HW, Mookerjee S, Bhutta MF. Epidemiology of chronic suppurative otitis media: systematic review to estimate global prevalence. *J Epidemiol Glob Health*. 2025;15(1):55.

13. Chadha SK, Sayal A, Malhotra V, Agarwal AK. Prevalence of preventable ear disorders in over 15,000 schoolchildren in northern India. *J Laryngol Otol.* 2013;127(1):28-32.
14. Bellad SA, Kavi A, Mudhol RS. Prevalence of chronic suppurative otitis media among school children residing in rural area of Belagavi, South India. *Indian J Otolaryngol Head Neck Surg.* 2019;71(Suppl 2):1549-1552.
15. Verma P, Gargava A, Saxena S, Narvey VP. To study the prevalence of extracranial & intracranial complications in chronic suppurative otitis media. *Indian J Otolaryngol Head Neck Surg.* 2022;74(Suppl 1):608-611.
16. Elemraid M, Mackenzie IJ, Fraser WD, et al. Nutritional factors associated with chronic suppurative otitis media in Yemeni children. *Eur J Clin Nutr.* 2011;65(8):895-902.
17. [AUTHORS NOT VERIFIED]. Comprehensive assessment of the global burden and risk determinants of otitis media. *BioSci Rev.* 2025. [Volume/issue/pages/DOI could not be confirmed - author names are missing in the source manuscript; please supply or allow further verification with more identifying detail.]
18. Tseng CC, Lim V, Jyung RW. Use of artificial intelligence for the diagnosis of cholesteatoma. *Laryngoscope Investig Otolaryngol.* 2023;8(1):201-211.
19. Pichichero ME, Poole MD. Assessing diagnostic accuracy and tympanocentesis skills in managing otitis media. *Arch Pediatr Adolesc Med.* 2001;155:1137-1142.
20. Rosenfeld RM. Diagnostic certainty for acute otitis media. *Int J Pediatr Otorhinolaryngol.* 2002;64(2):89-95.
21. Suresh K, Wu MP, Benboujja F, Christakis B, Newton A, Hartnick CJ, Cohen MS. AI model versus clinician otoscopy in the operative setting for otitis media diagnosis. *Otolaryngol Head Neck Surg.* 2024;170(6):1598-1601.
22. [AUTHORS NOT VERIFIED]. Diagnostic accuracy of otitis media with and without a fictitious AI support among physicians in primary care and medical students. *Scand J Prim Health Care.* 2025. [Volume/issue/pages/DOI could not be confirmed - author names are missing in the source manuscript; please supply or allow further verification with more identifying detail.]
23. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, Thrun S. Dermatologist-level classification of skin cancer with deep neural networks. *Nature.* 2017;542:115-118.
24. Gulshan V, Peng L, Coram M, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA.* 2016;316(22):2402-2410.
25. Litjens G, Kooi T, Bejnordi BE, et al. A survey on deep learning in medical image analysis. *Med Image Anal.* 2017;42:60-88.
26. Esteva A, Robicquet A, Ramsundar B, et al. A guide to deep learning in healthcare. *Nat Med.* 2019;25:24-29.
27. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. *Nat Med.* 2019;25:44-56.
28. Myburgh HC, van Zijl WH, Swanepoel D, Hellstrom S, Laurent C. Otitis media diagnosis for developing countries using tympanic membrane image-analysis. *EBioMedicine.* 2016;5:156-160.
29. Myburgh HC, Jose S, Swanepoel DW, Laurent C. Towards low cost automated smartphone- and cloud-based otitis media diagnosis. *Biomed Signal Process Control.* 2018;39:34-52.
30. Cavalcanti TC, et al. Intelligent smartphone-based multimode imaging otoscope for the mobile diagnosis of otitis media. *Biomed Opt Express.* 2021;12:7765-7779.
31. Sandstrom J, Myburgh H, Laurent C, Swanepoel DW, Lundberg T. A machine learning approach to screen for otitis media using digital otoscope images labelled by an expert panel. *Diagnostics.* 2022;12:1318.
32. Chen YC, Chu YC, Huang CY, et al. Smartphone-based artificial intelligence using a transfer learning algorithm for the detection and diagnosis of middle ear diseases: a retrospective deep learning study. *EClinicalMedicine.* 2022;15:101543.
33. You E, Lin V, Mijovic T, Eskander A, Crowson MG. Artificial intelligence applications in otology: a state of the art review. *Otolaryngol Head Neck Surg.* 2020;163:1123-1133.
34. Islam MM, Poly TN, Yang HC, et al. Deep learning-based system for automated diagnosis of ear disease using otoscopic images: a systematic review. *Diagnostics (Basel).* 2023;13(2):280.
35. Song D, Kim T, Lee Y, Kim J. Image-based artificial intelligence technology for diagnosing middle ear diseases: a systematic review. *J Clin Med.* 2023;12(18):5831.
36. Tsutsumi K, Goshtasbi K, Risbud A, Khosravi P, Pang JC, Lin HW, Djalilian HR, Abouzari M. A web-based deep learning model for automated diagnosis of otoscopic images. *Otol Neurotol.* 2021;42(9):e1382-e1388.
37. Surapaneni S, Rangarajan N, Davis K, Pletcher K, Flowers J, Strub G, et al. Artificial intelligence classification of pediatric middle ear effusion using consumer-grade otoscopes. *Int J Pediatr Otorhinolaryngol.* 2025;200:112655.
38. Crowson MG, Hartnick CJ, Diercks GR, Gallagher TQ, Fracchia MS, Setlur J, Cohen MS. Machine learning for accurate intraoperative pediatric middle ear effusion diagnosis. *Pediatrics.* 2021;147(4):e2020034546.
39. Yang CY, Lee CJ, Lai WS, Chen KY, Kuo CF, Liu CH, Liu SC. Deep learning-based semantic segmentation and classification of otoscopic images for otitis media diagnosis and health promotion. *Diagnostics (Basel).* 2026;16(3):467.
40. Choi Y, Chae J, Park K, Hur J, Kweon J, Ahn JH. Automated multi-class classification for prediction of tympanic membrane changes with deep learning models. *PLoS One.* 2022;17(10):e0275846.
41. Akyol K, Uçar E, Atila Ü, Uçar M. An ensemble approach for classification of tympanic membrane conditions using soft voting classifier. *Multimed Tools Appl.* 2024;83(32):77809-77830.

42. Habib AR, Xu Y, Bock K, Mohanty S, Sederholm T, Weeks WB, Dodhia R, Lavista Ferres J, Perry C, Sacks R, Singh N. Evaluating the generalizability of deep learning image classification algorithms to detect middle ear disease using otoscopy. *Sci Rep.* 2023;13:5368.
43. Cha D, Pae C, Seong SB, Choi JY, Park HJ. Automated diagnosis of ear disease using ensemble deep learning with a big otoendoscopy image database. *EBioMedicine.* 2019;45:606-614.
44. Zeng J, Kang W, Chen S, et al. A deep learning approach to predict conductive hearing loss in patients with otitis media with effusion using otoscopic images. *JAMA Otolaryngol Head Neck Surg.* 2022;148(7):612-620.
45. Zeng X, Jiang Z, Luo W, Li H, Li H, Li G, Shi J, Wu K, Liu T, Lin X, Wang F, Li Z. Efficient and accurate identification of ear diseases using an ensemble deep learning model. *Sci Rep.* 2021;11:10839.
46. Sundgaard JV, Harte J, Bray P, Laugesen S, Kamide Y, Tanaka C, Paulsen RR, Christensen AN. Deep metric learning for otitis media classification. *Med Image Anal.* 2021;71:102034.
47. Guo Q, Xie L, et al. Deep learning for otitis media classification using otoscopic image. *Medicine (Baltimore).* 2025;104(48):e46218.
48. Lu F, Yue Y, Li Z, Zhang M, Luo W, Zhang F, Liu T, Shi J, Wang G, Zeng X. Ear-Keeper: a cross-platform artificial intelligence system for rapid and accurate ear disease diagnosis. *Digit Health.* 2026;12:20552076251412635.