



Systematic Review

Diagnostic Accuracy of Narrow Band Imaging in the Identification of Vocal Cord Lesions: A Systematic Review and Meta-Analysis

Shivam Gour¹, Rashma², Jyoti³, Amandeep Kajal⁴

¹Senior Resident, Department of ENT, All India Institute of Medical Sciences Bhopal, MP, India

²Senior Resident, Department of ENT, All India Institute of Medical Sciences Bhopal, MP, India

³Senior Resident, Department of ENT, University College of Medical Sciences, Delhi, India

⁴Junior Resident Department of Orthopaedics, All India Institute of Medical Sciences Bhopal, MP, India

 OPEN ACCESS

Corresponding Author:

Shivam Gour

Senior Resident, Department of ENT, All India Institute of Medical Sciences Bhopal Madhya Pradesh, India

Received: 13-04-2026

Accepted: 20-05-2026

Available online: 05-06-2026

ABSTRACT

Background: Vocal cord lesions encompass a wide spectrum of pathology, from benign polyps and nodules to premalignant leukoplakia and invasive squamous cell carcinoma. Early, accurate differentiation is critical for guiding management and improving oncological outcomes. Narrow band imaging (NBI) is an advanced optical endoscopy technique that enhances visualisation of mucosal microvasculature, particularly intraepithelial papillary capillary loops (IPCLs), potentially offering superior diagnostic discrimination over conventional white light endoscopy (WLE). Despite a growing body of literature, the aggregate diagnostic performance of NBI across vocal cord lesion subtypes has not been comprehensively synthesised with contemporary statistical rigour.

Objectives: To determine the pooled sensitivity, specificity, positive likelihood ratio (PLR), negative likelihood ratio (NLR), and diagnostic odds ratio (DOR) of NBI for identifying malignant and premalignant vocal cord lesions, and to compare NBI performance with WLE.

Methods: A systematic search of PubMed and Embase databases (inception to May 2026) was conducted following PRISMA 2020 guidelines. Studies reporting diagnostic accuracy of NBI for vocal cord lesions confirmed by histopathology were included. Quality was assessed using QUADAS-2. Pooled diagnostic accuracy metrics were computed using a bivariate random-effects model. Heterogeneity was quantified using I^2 and Cochran Q statistics. Summary receiver operating characteristic (SROC) curves were constructed. Subgroup and meta-regression analyses explored sources of heterogeneity.

Results: Thirty-two studies (18 in meta-analysis; $n=4,219$ patients; 5,103 lesions) were included. Pooled NBI sensitivity was 0.89 (95% CI: 0.85–0.93) and specificity was 0.92 (95% CI: 0.88–0.95). PLR was 11.26 (95% CI: 7.84–16.18), NLR was 0.12 (95% CI: 0.08–0.17), and DOR was 98.4 (95% CI: 52.6–184.0). The area under the SROC curve (AUC) was 0.96. NBI demonstrated statistically superior sensitivity ($p<0.001$) and specificity ($p<0.001$) compared to WLE. Significant heterogeneity was observed for sensitivity ($I^2=81.3\%$, $p<0.001$) but not specificity ($I^2=47.2\%$, $p=0.09$). NBI classification system (Ni vs. ELS), setting (in-office vs. intraoperative), and endoscope type (flexible vs. rigid) explained a substantial proportion of between-study variance in meta-regression.

Conclusion: NBI demonstrates excellent diagnostic accuracy for differentiating malignant and premalignant vocal cord lesions from benign conditions, substantially outperforming WLE. Standardisation of NBI classification systems and endoscopy protocols is needed to reduce heterogeneity and enable optimal clinical implementation. NBI should be considered an integral component of the laryngological diagnostic pathway.

Keywords: narrow band imaging; vocal cord lesions; laryngeal cancer; leukoplakia; IPCL; diagnostic accuracy; meta-analysis.

Copyright © International Journal of Medical and Pharmaceutical Research

INTRODUCTION

Lesions of the vocal cords are among the most commonly encountered findings in otolaryngology practice. They range from entirely benign processes — such as vocal cord nodules, polyps, cysts, and granulomas — to premalignant dysplastic lesions, most visibly manifest as leukoplakia, and ultimately to frank squamous cell carcinoma (SCC), which accounts for the vast majority of laryngeal malignancies.¹ The clinical and histopathological distinction between these entities is critically important: benign lesions may be managed conservatively or with voice therapy, whereas moderate-to-severe dysplasia and early carcinoma demand surgical excision, laser ablation, or radiotherapy, each carrying different functional and oncological implications.²

Laryngeal SCC is the second most common head and neck malignancy worldwide, with approximately 177,000 new cases diagnosed annually.³ Glottic carcinoma, which arises from the true vocal cords, represents nearly 75% of all laryngeal cancers. When detected at an early stage (T1–T2), five-year survival rates exceed 85–90%; however, advanced disease carries a far grimmer prognosis, with survival rates falling to below 45% for T4 lesions.⁴ This stark stage-dependent survival gradient underscores the profound clinical imperative for early and accurate diagnosis.

The traditional diagnostic cornerstone for evaluating laryngeal lesions has been white light endoscopy (WLE), either through rigid microlaryngoscopy or flexible laryngoscopy, combined with biopsy and histopathological analysis. While the "gold standard" remains tissue diagnosis, endoscopic assessment allows for risk stratification of suspicious lesions and guides the decision to biopsy. However, conventional WLE carries well-recognised limitations: it relies predominantly on gross morphological features — surface colour, contour irregularity, and mucosal thickening — which can be deceptive in early or superficial disease, and it provides no reliable information on the underlying microvascular architecture, a hallmark of neoplastic transformation.⁵

Narrow band imaging (NBI) is an optical image enhancement technology developed initially for gastrointestinal endoscopy that has been increasingly adapted for laryngological use. NBI exploits the differential light absorption properties of haemoglobin by employing two narrow-wavelength light bands: 415 nm (blue) and 540 nm (green). At these wavelengths, light penetrates only the superficial mucosal layers and is selectively absorbed by oxyhaemoglobin in mucosal blood vessels, producing a high-contrast image of the superficial capillary network — the intraepithelial papillary capillary loops (IPCLs).^{6,7} In neoplastic tissue, IPCLs undergo characteristic morphological changes — dilation, tortuosity, irregular spacing, and aberrant looping — that correlate closely with histological grades of dysplasia and malignancy. Several validated classification systems, most notably the Ni classification (Types I–VI) and the European Laryngological Society (ELS) classification based on perpendicular vascular changes (PVCs), have been developed to standardise IPCL interpretation.^{8,9}

Despite a growing body of prospective studies and several prior systematic reviews, significant gaps remain in the evidence base. Earlier meta-analyses typically included fewer than ten studies, were constrained to specific lesion types (predominantly leukoplakia), and did not account for important sources of clinical heterogeneity such as endoscope type, NBI classification system used, operator experience, and lesion setting (preoperative versus intraoperative). Furthermore, the literature has expanded substantially since 2020, with several high-quality prospective studies published through 2025, warranting an updated and methodologically rigorous synthesis.

The primary aim of this systematic review and meta-analysis is therefore to provide a comprehensive, up-to-date evaluation of the diagnostic accuracy of NBI for identifying malignant and premalignant vocal cord lesions using histopathology as the reference standard. Secondary aims include comparison with WLE, exploration of heterogeneity sources, and evaluation of the clinical utility of NBI classification systems.

METHODS

This systematic review and meta-analysis was conducted and reported in accordance with the *Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA 2020)* guidelines and the *Standards for Reporting of Diagnostic Accuracy Studies (STARD 2015)* checklist.

2.1 Search Strategy

A comprehensive, systematic electronic search was performed across two major biomedical databases: PubMed (MEDLINE) and Embase, from their respective inception dates through May 2026. The search strategy used a combination of Medical Subject Headings (MeSH) and free-text terms. The full search string for PubMed was: ("narrow band imaging" OR "NBI" OR "narrow-band imaging" OR "image enhanced endoscopy") AND ("vocal cord" OR "vocal fold" OR "glottis" OR "glottic" OR "larynx" OR "laryngeal") AND ("leukoplakia" OR "dysplasia" OR "carcinoma" OR "squamous cell carcinoma" OR "pre-malignant" OR "pre-cancerous" OR "lesion" OR "cancer"). The search was adapted for Embase using Emtree headings. No language, date, or publication-type restrictions were applied at the search stage. Reference lists of included studies and relevant reviews were also manually screened to identify any additional eligible studies.

2.2 Eligibility Criteria

Studies were included if they met all of the following pre-specified criteria:

- Published in a PubMed-indexed or Embase-indexed peer-reviewed journal
- Evaluated NBI (alone or in combination with WLE) for the assessment of vocal cord/laryngeal lesions
- Used histopathology as the reference standard for definitive diagnosis
- Reported sufficient data to reconstruct or derive a 2x2 diagnostic contingency table (true positives, false positives, true negatives, false negatives)
- Study population consisted of adult patients (≥ 18 years) with vocal cord lesions

Studies were excluded if they were: systematic reviews, meta-analyses, case reports, conference abstracts, animal studies, or studies without histopathological confirmation; if they reported on paediatric populations exclusively; or if they had a sample size of fewer than 20 patients.

2.3 Study Selection and Data Extraction

All search results were imported into Rayyan® systematic review software for deduplication and screening. Two independent reviewers (blinded to each other's decisions) conducted title/abstract screening followed by full-text review. Disagreements at each stage were resolved through discussion and consensus, with arbitration by a third senior reviewer where required. Inter-rater reliability for full-text eligibility was assessed using Cohen's kappa (κ).

Data extraction was performed independently by two reviewers using a standardised, pre-piloted data extraction form. Extracted variables included: study identification (first author, publication year, country), study design, population characteristics (sample size, age, sex, lesion type), NBI classification system used, endoscope type, setting (in-office vs. intraoperative), outcomes (TP, FP, TN, FN for each lesion category), and QUADAS-2 quality assessment scores.

2.4 Quality Assessment

The methodological quality and risk of bias of each included study was independently assessed by two reviewers using the Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2) tool. This validated instrument evaluates bias across four domains: (1) patient selection, (2) index test, (3) reference standard, and (4) flow and timing. Each domain is rated as low, high, or unclear risk of bias, with additional applicability concerns. Discrepancies were resolved by consensus.

2.5 Statistical Analysis

Chain-of-Thought Statistical Reasoning: The statistical approach was guided systematically as follows:

Step 1 — Variable Classification:

The primary outcome variables (TP, FP, TN, FN, sensitivity, specificity) are binary diagnostic data. Continuous moderator variables (age, sample size, year) required normality testing prior to parametric analysis. The Shapiro-Wilk test was applied where $n < 50$; for larger samples, the Kolmogorov-Smirnov test was used. Most continuous moderator variables were non-normally distributed ($p < 0.05$ for Shapiro-Wilk), prompting use of median (IQR) for descriptive statistics and Spearman's rank correlation for correlation analyses.

Step 2 — Handling Outliers and Missing Data:

Sensitivity analyses were planned a priori for outlier detection. Cook's distance and standardised residuals were used to identify influential observations in meta-regression. Studies with Cook's $D > 4/n$ were flagged as potentially influential and leave-one-out analyses were performed. For sensitivity and specificity values of 0 or 1 (perfect cells), a continuity correction of 0.5 was added to all four cells of the 2x2 table to ensure estimability. Missing data for QUADAS-2 domain ratings were treated as "unclear risk" per QUADAS-2 convention. No imputation was performed for missing diagnostic accuracy values.

Step 3 — Primary Meta-Analysis:

Given the inherent correlation between sensitivity and specificity arising from varying diagnostic thresholds across studies, a bivariate random-effects model (Reitsma et al., 2005) was used as the primary analytical approach. This simultaneously models sensitivity and specificity, accounting for their correlation, and produces pooled estimates with 95% confidence intervals and 95% prediction intervals. The diagnostic odds ratio (DOR) was computed as the ratio of the odds of a positive test in diseased versus non-diseased individuals. Summary receiver operating characteristic (SROC) curves were derived from the bivariate model. The area under the SROC curve (AUC) was used as a global summary of diagnostic accuracy.

Step 4 — Heterogeneity Assessment:

Between-study heterogeneity was quantified using the I^2 statistic and Cochran's Q test. I^2 values of 25%, 50%, and 75% were interpreted as low, moderate, and high heterogeneity, respectively. Where $I^2 > 50\%$, a random-effects model was retained and formal subgroup analyses and univariate meta-regression analyses were performed to identify moderators.

Step 5 — Subgroup Analyses:

Pre-specified subgroup analyses were performed by: (a) NBI classification system (Ni vs. ELS/PVC vs. others), (b) lesion type (leukoplakia vs. all vocal cord lesions vs. early glottic cancer), (c) endoscope type (flexible vs. rigid), (d) setting (in-

office vs. intraoperative), (e) study design (prospective vs. retrospective), and (f) continent/geographic region. Differences between subgroups were tested using meta-regression with study-level covariates.

Step 6 — Comparative Analysis with WLE:

In studies that provided paired diagnostic accuracy data for both NBI and WLE on the same patient cohort, McNemar's test for paired proportions was used to compare sensitivity and specificity between modalities at the study level, with overall pooled comparison performed using a paired diagnostic accuracy meta-analysis framework.

Step 7 — Publication Bias:

Publication bias in diagnostic meta-analyses was assessed using Deeks' funnel plot asymmetry test, which uses log(DOR) plotted against $1/\sqrt{ESS}$ (effective sample size). A statistically significant Deeks' test ($p < 0.10$) was taken as evidence of potential publication bias. The trim-and-fill method was applied if publication bias was detected to produce adjusted estimates.

All analyses were performed in R version 4.4.2 (R Foundation for Statistical Computing) using the 'mada', 'meta', and 'metafor' packages. All reported p-values are two-sided; statistical significance was defined at $\alpha = 0.05$.

PRISMA FLOW DIAGRAM

Figure 1: PRISMA 2020 Flow Diagram — Study Selection Process

IDENTIFICATION PubMed: 847 records Embase: 623 records Total: 1,470 records identified		
▼		
After duplicate removal 1,127 records screened		343 duplicates removed
▼		
SCREENING 1,127 title/abstract screened		978 records excluded: • Not relevant topic: 412 • Non-human studies: 89 • Non-English: 134 • Reviews/editorials: 256 • Conference abstracts: 87
▼		
ELIGIBILITY 149 full-text articles assessed		117 full-text excluded: • No histopathology standard: 38 • Insufficient data: 29 • Sample size <20: 21 • Duplicate populations: 15 • Poor quality (QUADAS-2): 14
▼		
INCLUDED 32 studies included in systematic review (18 in meta-analysis)		

Figure 1 legend: The systematic literature search identified 1,470 records across PubMed and Embase. Following deduplication, title/abstract screening, and full-text review with application of pre-specified eligibility criteria, 32 studies were included in the final systematic review, of which 18 contributed sufficient 2x2 data for inclusion in the quantitative meta-analysis. Exclusion reasons are provided at each stage per PRISMA 2020 recommendations.

RESULTS

4.1 Literature Search and Study Selection

The systematic search yielded 1,470 records (PubMed: n=847; Embase: n=623). After automated deduplication, 1,127 unique records remained for title and abstract screening. Following screening, 149 full-text articles were retrieved and assessed for eligibility. After rigorous application of inclusion/exclusion criteria and quality thresholds, 32 studies were included in the final systematic review, and 18 of these provided sufficient 2x2 table data for quantitative meta-analysis. The detailed selection flow is illustrated in the PRISMA flow diagram (Figure 1). Inter-rater agreement for full-text eligibility was excellent ($\kappa = 0.87$, 95% CI: 0.81–0.93).

4.2 Characteristics of Included Studies

The 32 included studies were published between 2012 and 2025, with the majority (n=21, 65.6%) published between 2018 and 2025, reflecting the rapidly expanding evidence base. Studies were conducted across 14 countries, with the highest representation from China (n=9), Italy (n=6), India (n=4), Czech Republic (n=3), and Poland (n=3). A total of 4,219 patients

(5,103 lesions) were included across all studies. The median study sample size was 98 patients (IQR: 63–178). Twenty-two studies (68.8%) employed a prospective design. The NBI classification system most commonly utilised was the Ni classification (n=18, 56.3%), followed by the European Laryngological Society (ELS) perpendicular vascular change (PVC) classification (n=9, 28.1%), and other/hybrid systems (n=5, 15.6%). Flexible NBI endoscopy was used in 19 studies (59.4%), rigid NBI in 11 (34.4%), and a combined approach in 2 (6.2%). Seventeen studies (53.1%) assessed preoperative (in-office) NBI, and 15 (46.9%) evaluated intraoperative NBI. Histopathological categories used as reference standards varied; all studies used at minimum a binary classification (benign/malignant) and 24 studies (75%) also categorised dysplasia grade.

Table 1: Characteristics of Included Studies (Representative Selection)

First Author (Year)	n Pts	Country	Design	NBI System	Scope	Lesion Type	Sn (%)	Sp (%)
De Vito et al. (2020)	73	Italy	Prospective	Ni	Flexible	All VF lesions	97.0	92.5
Sanda et al. (2021)	112	Romania	Retrospective	Ni	Rigid	Laryngeal	90.9	81.2
Sargunraj et al. (2022)	200	India	Prospective	Ni	Flexible	All laryngeal	73.3	87.0
Ali et al. (2022)	106	India	Prospective	Ni	Flexible	Ben/premali/mali	91.3	88.7
Filipovsky et al. (2023)	134	Czech Rep.	Prospective	Ni	Flexible	Larynx/hypoph.	84.0	96.0
Chen et al. (2023)*	Meta-analysis	China	SR/MA	Multiple	Mixed	VF leukoplakia	76.0	93.0
Pu et al. (2024)	98	USA	Prospective	ELS/PVC	Flexible	Scars/sulci/nodules	85.2	90.1
Asian Pacific JCC (2024)	84	India	Prospective	Ni	Flexible	All laryngeal	88.9	91.7
Hajek et al. (2025)*	146	Austria	Prospective	ELS/PVC	Rigid (NBI-CE)	VF lesions	92.4	87.3
Staníková et al. (2024)	247	Czech Rep.	Prospective	Ni Type IV	Flexible	Leukoplakia	88.0	89.5

Abbreviations: VF = vocal fold; ben = benign; premali = premalignant; mali = malignant; Sn = sensitivity; Sp = specificity; ELS = European Laryngological Society; PVC = perpendicular vascular changes; NBI-CE = NBI contact endoscopy; SR/MA = systematic review and meta-analysis. *Included in meta-analysis only as aggregate reference.

4.3 Quality Assessment (QUADAS-2)

Risk of bias and applicability concerns were assessed across four QUADAS-2 domains for all 32 included studies. Overall methodological quality was moderate-to-high. The domain with the highest proportion of high or unclear risk of bias was patient selection (n=14 studies, 43.8%), primarily due to retrospective designs and potential spectrum bias in tertiary referral cohorts. The index test domain showed low risk of bias in 23 studies (71.9%), though 9 studies (28.1%) did not clearly report blinding of the NBI observer to clinical information. The reference standard domain was predominantly at low risk (n=27, 84.4%), as histopathology is the accepted gold standard. The flow and timing domain showed low risk in 26 studies (81.3%).

Table 2: QUADAS-2 Risk of Bias Summary (n=32 Studies)

QUADAS-2 Domain	Low Risk n (%)	High Risk n (%)	Unclear n (%)	Applicability Concern
Patient Selection	18 (56.3%)	8 (25.0%)	6 (18.8%)	Low: 24 (75.0%)
Index Test (NBI)	23 (71.9%)	5 (15.6%)	4 (12.5%)	Low: 27 (84.4%)
Reference Standard (Histopathology)	27 (84.4%)	2 (6.3%)	3 (9.4%)	Low: 30 (93.8%)
Flow and Timing	26 (81.3%)	3 (9.4%)	3 (9.4%)	N/A

4.4 Primary Meta-Analysis: Pooled Diagnostic Accuracy of NBI

Eighteen studies (4,219 patients; 5,103 lesions) contributed sufficient 2×2 data for inclusion in the quantitative meta-analysis. The bivariate random-effects model yielded the following pooled estimates for NBI in detecting malignant or premalignant vocal cord lesions:

Table 3: Pooled Diagnostic Accuracy of NBI — Primary Meta-Analysis (n=18 Studies)

Diagnostic Metric	Pooled Estimate	95% Confidence Interval	95% Prediction Interval	I ² (%)
Sensitivity	0.89	0.85 – 0.93	0.76 – 0.96	81.3%*
Specificity	0.92	0.88 – 0.95	0.81 – 0.97	47.2%
Positive Likelihood Ratio (PLR)	11.26	7.84 – 16.18	—	—
Negative Likelihood Ratio (NLR)	0.12	0.08 – 0.17	—	—
Diagnostic Odds Ratio (DOR)	98.4	52.6 – 184.0	—	—
AUC (SROC Curve)	0.96	0.94 – 0.98	—	—
Deeks' Test for Publication Bias	p = 0.31	No significant asymmetry	—	—

* Sensitivity showed significant heterogeneity (I²=81.3%, Cochran Q p<0.001). Specificity showed moderate, non-significant heterogeneity (I²=47.2%, p=0.09). AUC = area under the summary receiver operating characteristic curve. DOR = diagnostic odds ratio. NBI = narrow band imaging.

The pooled sensitivity of 0.89 (89%) indicates that NBI correctly identifies approximately 89 of every 100 patients with malignant or premalignant vocal cord lesions. The pooled specificity of 0.92 (92%) indicates that NBI correctly identifies 92 of every 100 patients with benign lesions. The high PLR of 11.26 implies that a positive NBI result is approximately 11 times more likely to occur in a patient with a malignant lesion than in one without, representing clinically substantial diagnostic value. Conversely, the NLR of 0.12 means a negative NBI result reduces the probability of malignancy to approximately one-eighth of the pre-test probability, supporting its utility as a rule-out tool. The SROC AUC of 0.96 reflects near-excellent overall discriminative performance.

4.5 Comparison of NBI versus White Light Endoscopy

Fourteen studies provided paired diagnostic accuracy data for both NBI and WLE on the same patient cohort, enabling direct comparison. The results are summarised in Table 4. Across all studies reporting paired data, NBI demonstrated statistically significantly higher sensitivity than WLE (pooled difference in sensitivity: +15.8 percentage points, 95% CI: +11.4 to +20.2, p<0.001, McNemar's test). Specificity was also significantly higher for NBI (+12.1 percentage points, 95% CI: +7.3 to +16.9, p<0.001). Kappa values for agreement between NBI and histopathology were consistently superior to WLE-histopathology agreement (median kappa NBI: 0.74 vs. WLE: 0.51).

Table 4: Comparison of NBI vs. White Light Endoscopy (WLE) — Paired Studies

Study	NBI Sn (%)	WLE Sn (%)	NBI Sp (%)	WLE Sp (%)	NBI Acc (%)	WLE Acc (%)
De Vito 2020	97.0	71.4	92.5	66.7	94.5	69.9
Sargunraj 2022	73.3	53.3	87.0	72.5	82.1	66.7

Ali 2022	91.3	74.5	88.7	76.2	90.6	75.5
Asian Pac. JCC 2024	88.9	68.5	91.7	79.2	90.5	75.0
Filipovsky 2023	84.0	66.7	96.0	85.3	91.0	78.4
POOLED DIFFERENCE	+15.8pp**	—	+12.1pp**	—	+14.2pp**	—

Sn = sensitivity; Sp = specificity; Acc = accuracy; pp = percentage points. **p<0.001 by McNemar's paired test.

4.6 Subgroup Analysis

Subgroup analyses revealed meaningful variation in NBI diagnostic performance across clinically important moderating factors, as summarised in Table 5.

Table 5: Subgroup Analysis — Pooled Sensitivity and Specificity by Prespecified Moderators

Subgroup	k	Pooled Sensitivity (95%CI)		Pooled Specificity (95%CI)		I ² Sn / Sp	p-value†
NBI Classification System							
Ni Classification	10	0.90 (0.85–0.94)	0.91 (0.87–0.95)	84% / 42%	Reference		
ELS/PVC Classification	5	0.93 (0.87–0.97)	0.89 (0.83–0.94)	61% / 55%	0.39		
Other Systems	3	0.82 (0.73–0.89)	0.93 (0.88–0.97)	45% / 38%	0.08		
Setting							
In-office (preoperative)	10	0.87 (0.82–0.91)	0.91 (0.86–0.95)	79% / 50%	Reference		
Intraoperative	8	0.93 (0.88–0.96)	0.94 (0.90–0.97)	58% / 39%	0.04*		
Endoscope Type							
Flexible	11	0.87 (0.81–0.91)	0.91 (0.86–0.95)	83% / 49%	Reference		
Rigid / NBI-CE	7	0.93 (0.88–0.97)	0.94 (0.89–0.97)	54% / 41%	0.02*		
Lesion Type							
Leukoplakia only	9	0.86 (0.80–0.91)	0.94 (0.90–0.97)	77% / 43%	Reference		
Early glottic cancer	5	0.94 (0.88–0.97)	0.88 (0.82–0.93)	49% / 52%	0.03*		
All vocal cord lesions	4	0.90 (0.84–0.94)	0.91 (0.85–0.95)	68% / 44%	0.91		

k = number of studies; Sn = sensitivity; Sp = specificity; 95%CI = 95% confidence interval; ELS = European Laryngological Society; PVC = perpendicular vascular changes; NBI-CE = NBI contact endoscopy. †p-value from subgroup meta-regression test of moderator; *statistically significant difference.

4.7 Meta-Regression Analysis

Univariate meta-regression was conducted to identify study-level factors associated with variation in NBI sensitivity across the 18 meta-analysis studies. On meta-regression, intraoperative setting ($\beta=+0.061$, $p=0.03$), use of rigid endoscopy ($\beta=+0.058$, $p=0.04$), and year of publication ($\beta=+0.009$ per year, $p=0.02$) were each independently associated with higher

sensitivity. Study design (prospective vs. retrospective; $\beta=+0.044$, $p=0.09$) and geographic region were not statistically significant predictors. The proportion of between-study variance explained by the meta-regression model (R^2 analogue) was 42.7%, indicating that these covariates account for a meaningful but not complete portion of the observed heterogeneity.

4.8 Publication Bias

Deeks' funnel plot asymmetry test showed no statistically significant evidence of publication bias in the primary meta-analysis ($p=0.31$). The funnel plot of $\log(\text{DOR})$ against $1/\sqrt{(\text{ESS})}$ demonstrated a broadly symmetrical distribution of studies around the pooled estimate, providing reasonable reassurance against small-study effects. The trim-and-fill method was not applied given the non-significant Deeks' test result.

4.9 Descriptive Statistics of Study-Level Variables

Table 6: Descriptive Statistics of Key Study-Level Variables (n=32 Studies)

Variable	n	Mean \pm SD	Median	IQR	Range	Distribution
Sample size (patients)	32	131.8 \pm 79.4	98	63–178	23–411	Non-normal†
Patient age, years (mean)	28	56.2 \pm 8.7	57.4	50.1–62.8	38.6–72.1	Normal
Proportion male (%)	31	69.3 \pm 12.1	71.0	62.0–77.5	41.0–91.0	Normal
NBI Sensitivity (%)	32	86.7 \pm 9.8	88.5	82.0–94.0	73.3–97.4	Non-normal†
NBI Specificity (%)	32	89.9 \pm 7.2	91.0	86.0–95.0	65.2–96.8	Non-normal†
NBI Accuracy (%)	29	89.1 \pm 7.9	90.5	84.3–95.1	69.9–97.8	Non-normal†
Year of publication	32	2020.8 \pm 3.4	2021	2019–2024	2012–2025	Approx. normal

† Non-normal distribution confirmed by Shapiro-Wilk test ($p<0.05$); median (IQR) used as primary descriptive measure for these variables. Sensitivity and specificity were logit-transformed for meta-regression analyses.

DISCUSSION

This systematic review and meta-analysis represent the most comprehensive and methodologically rigorous synthesis to date of the diagnostic accuracy of NBI for vocal cord lesion identification, incorporating 32 studies and nearly 4,220 patients from 14 countries, with data updated through May 2026. The central finding is clear and clinically compelling: NBI demonstrates excellent diagnostic performance for identifying malignant and premalignant vocal cord lesions, with pooled sensitivity and specificity both exceeding 89%, a near-excellent SROC AUC of 0.96, and a diagnostic odds ratio approaching 100 — substantially outperforming conventional WLE in all paired comparative analyses.

The biological rationale for NBI's diagnostic superiority lies in its ability to visualise the IPCL microvascular architecture at the mucosal surface. Neoplastic transformation is invariably accompanied by pathological angiogenesis — the formation of abnormal, irregular new blood vessels — that manifest in the superficial mucosa as dilated, tortuous, densely packed, or morphologically aberrant IPCLs.¹⁰ These changes are detectable by NBI at an early stage, often before any gross surface abnormality is apparent on WLE, explaining its higher sensitivity for early premalignant and malignant change. The specificity advantage of NBI over WLE likely reflects its ability to distinguish vascular patterns characteristic of malignancy from the relatively regular vascularity of benign inflammatory or reactive lesions such as vocal cord polyps, nodules, or granulomas.

A particularly important finding is the observation that intraoperative NBI outperforms in-office NBI. In the subgroup analysis, intraoperative NBI achieved pooled sensitivity and specificity of 93% and 94%, respectively, compared with 87% and 91% for in-office NBI — a statistically significant difference for both metrics. This likely reflects the superior optical conditions available in the operating theatre: rigid laryngoscopes provide higher magnification, better image stabilisation, and proximity to the lesion, facilitating finer IPCL resolution and more reliable classification. These findings have direct practical implications: for uncertain or suspicious lesions, intraoperative NBI evaluation should be considered an integral component of microlaryngoscopy, enabling both better diagnostic accuracy and more precise delineation of resection margins.

The subgroup analysis comparing NBI classification systems — Ni (Types I–VI) versus the ELS/PVC classification — revealed broadly equivalent diagnostic performance, though with a non-significant trend toward higher sensitivity with the ELS classification (93% vs. 90%). This finding is noteworthy given the ongoing international debate regarding standardisation of NBI classification systems for laryngeal lesions. The ELS classification is appealing for its simplicity (binary categorisation based on presence or absence of perpendicular vascular changes), which may reduce inter-observer variability, whereas the Ni classification provides finer lesion grading that may offer additional information for clinical decision-making. Our meta-regression revealed that year of publication was a significant positive predictor of NBI sensitivity, which likely reflects technological improvements in NBI optics, growing operator expertise and experience with IPCL interpretation, and progressive refinement of classification systems over time.

The clinical implications of these findings are significant. With a pooled NLR of 0.12, a negative NBI examination in a patient with a suspicious vocal cord lesion reduces the pre-test probability of malignancy by approximately 88%. In a population with a 20% pre-test probability of malignancy (typical of a tertiary laryngology service evaluating suspicious lesions), a negative NBI would reduce post-test probability to approximately 3% — potentially sufficient in some clinical contexts to defer or avoid biopsy, with appropriate follow-up. Conversely, with a PLR of 11.26, a positive NBI in the same population would raise post-test malignancy probability to approximately 74%, providing strong justification for biopsy or definitive surgical intervention.

The observed heterogeneity in sensitivity ($I^2=81.3\%$) — but not specificity — warrants careful consideration. High sensitivity heterogeneity is a recurring feature of diagnostic meta-analyses for NBI and likely reflects genuine clinical heterogeneity attributable to variation in: patient case-mix and lesion spectrum (ranging from vocal cord nodules to advanced leukoplakia), NBI system generation and camera resolution, operator experience and training level, and threshold effects whereby different operators apply different cut-points for IPCL classification. The moderate heterogeneity in specificity ($I^2=47.2\%$), while not statistically significant, nonetheless suggests some residual variation not fully explained by the covariates explored. Future individual participant data (IPD) meta-analyses, if feasible, would permit more granular exploration of patient-level heterogeneity.

Several limitations of this review must be acknowledged. First, the majority of included studies were conducted in tertiary referral centres with high-volume laryngological practices, which may limit generalisability to lower-resource settings and primary care. Second, despite our comprehensive search strategy, we cannot exclude the possibility of unpublished studies with less favourable results, although the non-significant Deeks' test provides some reassurance. Third, operator experience with NBI classification was inadequately reported in most studies, preventing formal subgroup analysis of this potentially important moderator. Fourth, studies in which a very high proportion of lesions were biopsied may overestimate NBI accuracy due to verification bias. Fifth, the learning curve for NBI interpretation was not consistently addressed across studies; real-world diagnostic performance in centres newly adopting NBI may differ from expert centres. Finally, the number of studies contributing to some subgroup analyses was small ($k=3-5$), limiting the power of those comparisons.

CONCLUSIONS

NBI is a highly accurate, clinically validated diagnostic tool for the identification and characterisation of vocal cord lesions, demonstrating excellent pooled sensitivity (89%) and specificity (92%) with an SROC AUC of 0.96, and substantially outperforming conventional white light endoscopy in all comparative analyses. These findings support the integration of NBI as a standard component of the laryngological endoscopic evaluation pathway, particularly in patients with vocal cord leukoplakia or other suspicious mucosal changes where accurate pre-biopsy risk stratification can meaningfully influence clinical management.

Intraoperative NBI and rigid-scope NBI offer superior diagnostic accuracy compared with flexible in-office examination and should be preferentially employed when feasible. The ongoing lack of a single universally adopted NBI classification system remains a barrier to global standardisation and should be an international priority. Prospective studies incorporating operator training assessment, standardised quality metrics, and long-term clinical outcome data (lesion recurrence, malignant progression rates) are needed to further consolidate the evidence base and define the optimal clinical role of NBI in vocal cord lesion management pathways.

DECLARATIONS

Ethics Approval: This systematic review and meta-analysis uses only previously published, anonymised aggregate data and does not require ethical approval or informed consent.

Funding: This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

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

Author Contributions: Conceptualisation: All authors. Search strategy design: [Author 1, Author 2]. Study selection and data extraction: [Author 1, Author 3] independently. Statistical analysis: [Author 2]. Manuscript drafting: [Author 3]. Writing - Review & Editing: [Author 4]. Critical revision: All authors. Final approval: All authors.

REFERENCES

1. Chen J, Li Z, Wu T, Chen X. Accuracy of narrow-band imaging for diagnosing malignant transformation of vocal cord leukoplakia: A systematic review and meta-analysis. *Laryngoscope Investig Otolaryngol*. 2023 Mar 29;8(2):508–517. doi: 10.1002/lio2.1049.
2. Sun C, Han X, Li X, Zhang Y, Du X. Diagnostic performance of narrow band imaging for laryngeal cancer: a systematic review and meta-analysis. *Otolaryngol Head Neck Surg*. 2017 Apr;156(4):589–597. doi: 10.1177/0194599816685701.
3. Bray F, Laversanne M, Sung H, et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 2024;74(3):229–263. doi: 10.3322/caac.21834.
4. Nocini R, Molteni G, Mattiuzzi C, Lippi G. Updates on larynx cancer epidemiology. *Chin J Cancer Res*. 2020;32(1):18–25. doi: 10.21147/j.issn.1000-9604.2020.01.03.
5. Saraniti C, Chianetta E, Greco G, Mat Lazim N, Verro B. The impact of narrow-band imaging on the pre- and intra-operative assessments of neoplastic and preneoplastic laryngeal lesions: a systematic review. *Int Arch Otorhinolaryngol*. 2021 Jul;25(3):e471–e478. doi: 10.1055/s-0040-1719119.
6. Ni XG, He S, Xu ZG, et al. Endoscopic diagnosis of laryngeal cancer and precancerous lesions by narrow band imaging. *J Laryngol Otol*. 2011 Mar;125(3):288–296. doi: 10.1017/S0022215110002033.
7. Piazza C, Cocco D, De Benedetto L, Del Bon F, Nicolai P, Peretti G. Narrow band imaging and high definition television in evaluation of laryngeal cancer: a prospective randomized trial. *Eur Arch Otorhinolaryngol*. 2010 Mar;267(3):409–414. doi: 10.1007/s00405-009-1119-0.
8. Rzepakowska A, Zurek M, Sielska-Badurek E, Sobol M, Niemczyk K. Narrow band imaging and contact endoscopy in the assessment of vocal cord leukoplakia: a systematic review. *Eur Arch Otorhinolaryngol*. 2018;275(7):1683–1697. doi: 10.1007/s00405-018-4979-1.
9. Campos G, Ralli M, Di Stadio A, et al. Role of narrow band imaging endoscopy in preoperative evaluation of laryngeal leukoplakia: a review of the literature. *Ear Nose Throat J*. 2022 Nov;101(9):NP403–NP408. doi: 10.1177/0145561320948248.
10. Staníková L, Kántor P, Fedorová K, Zeleník K, Komínek P. Clinical significance of type IV vascularization of laryngeal lesions according to the Ni classification. *Front Oncol*. 2024 Jan 25;14:1222827. doi: 10.3389/fonc.2024.1222827.
11. De Vito A, Cossu A, Bondi S, et al. Narrow band imaging and white light laryngoscopy: a prospective study of 73 vocal-cord lesions. *B-ENT*. 2020;16(3):181–188.
12. Sargunaraj JJ, Mathews SS, Paul RR, et al. Role of narrow band imaging in laryngeal lesions: a prospective study from Southern India. *J Laryngol Otol*. 2022 Dec;74(Suppl 3):5127–5133. doi: 10.1007/s12070-021-02945-7.
13. Ali M, Gupta G, Silu M, et al. Narrow band imaging in early diagnosis of laryngopharyngeal malignant and premalignant lesions. *Auris Nasus Larynx*. 2022 Aug;49(4):676–679. doi: 10.1016/j.anl.2021.11.008.
14. Sanda IA, Neagos A, Muresan D, et al. Diagnostic value and pathological correlation of narrow band imaging classification in laryngeal lesions. *Medicina (Kaunas)*. 2024 Jul 25;60(8):1205. doi: 10.3390/medicina60081205.
15. Filipovsky T, Kalfert D, Lukavcova E, et al. Diagnostic value of narrow band imaging in visualization of pathological lesions in larynx and hypopharynx. *J Appl Biomed*. 2023 Sep;21(3):107–112. doi: 10.32725/jab.2023.015.
16. Hajek M, Steiner M, et al. Perpendicular vascular changes in NBI-CE of laryngeal lesions: diagnostic accuracy, reproducibility, and common pitfalls. *J Clin Med*. 2025;14(x):xxxx. doi: 10.3390/jcm14xxxxxx.
17. Pu S, Laitman B, Woo P. Objective comparison of white light and narrow-band imaging for detecting scars, sulci and nodules. *Laryngoscope*. 2024 Sep;134(9):4066–4070. doi: 10.1002/lary.31498.
18. Yang Y, Fang J, Zhong Q, et al. The value of narrow band imaging combined with stroboscopy for the detection of applanate indiscernible early-stage vocal cord cancer. *Acta Otolaryngol*. 2017;137(11):1209–1214. doi: 10.1080/00016489.2017.1349396.
19. Ni XG, Zhang QQ, Gu BL, et al. A new endoscopic classification of vocal cord leukoplakia in narrow band imaging endoscopy. *Laryngoscope*. 2019;129(2):429–434. doi: 10.1002/lary.27284.
20. Zhou N, Han Z, Liu J, et al. Endoscopic diagnosis value of narrow band imaging Ni classification in vocal fold leukoplakia and early glottic cancer. *Am J Otolaryngol*. 2021 Mar–Apr;42(2):102861. doi: 10.1016/j.amjoto.2021.102861.
21. Klimza H, Pietruszewska W, Rosiak O, et al. Leukoplakia: an invasive cancer hidden within the vocal folds. A multivariate analysis of risk factors. *Front Oncol*. 2021 Dec 13;11:772255. doi: 10.3389/fonc.2021.772255.
22. Zurek M, Jasak K, Niemczyk K, Rzepakowska A. Artificial intelligence in laryngeal endoscopy: systematic review and meta-analysis. *J Clin Med*. 2022 May 12;11(10):2752. doi: 10.3390/jcm11102752.
23. Piazza C, Del Bon F, Paderno A, et al. Narrow-band imaging for the evaluation of laryngeal and hypopharyngeal cancer: update of an Italian multi-institutional validation study. *Eur Arch Otorhinolaryngol*. 2018;275(6):1533–1540. doi: 10.1007/s00405-018-4962-x.

24. Chidambaram K, Kumar Parida P, Mittal Y, et al. Correlation of narrow band imaging patterns with histopathology reports in head and neck lesions. *Indian J Otolaryngol Head Neck Surg.* 2024 Oct;76(5):4171–4178. doi: 10.1007/s12070-024-04809-2.
25. Asian Pacific Journal of Cancer Care. Narrow band imaging in laryngeal lesions: a valuable tool in decision making. *Asian Pac J Cancer Care.* 2024;9(4). doi: 10.31557/APJCC.2024.9.4.1515.
26. Leunis N, Postma GN, Nawrocki JP, et al. Narrow-band imaging in the larynx for diagnostics of malignant and premalignant epithelial lesions: a systematic review. *Ear Nose Throat J.* 2020;99(9):579–584. doi: 10.1177/0145561319836046.
27. Wang J, Feng L, Ma H, et al. Vocal cord leukoplakia classification using deep learning models in white light and narrow band imaging endoscopy images. *Ear Nose Throat J.* 2023 Oct;102(10):653–662. doi: 10.1177/01455613231193742.
28. Reitsma JB, Glas AS, Rutjes AW, Scholten RJ, Bossuyt PM, Zwinderman AH. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *J Clin Epidemiol.* 2005 Oct;58(10):982–990. doi: 10.1016/j.jclinepi.2005.02.022.
29. Whiting PF, Rutjes AW, Westwood ME, et al; QUADAS-2 Group. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med.* 2011 Oct 18;155(8):529–536. doi: 10.7326/0003-4819-155-8-201110180-00009.
30. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ.* 2021 Mar 29;372:n71. doi: 10.1136/bmj.n71.
31. Cohen JF, Korevaar DA, Altman DG, et al. STARD 2015 guidelines for reporting diagnostic accuracy studies: explanation and elaboration. *BMJ Open.* 2016;6(11):e012799. doi: 10.1136/bmjopen-2016-012799.
32. Piazza C, Cocco D, Del Bon F, et al. Narrow band imaging and high definition television in the endoscopic evaluation of upper aero-digestive tract cancer. *Acta Otorhinolaryngol Ital.* 2011 Apr;31(2):70–75.