EUROPEAN RESPIRATORY REVIEW
REVIEW
K.M. COLD ET AL.

Artificial intelligence in bronchoscopy: a systematic review

Kristoffer Mazanti Cold[1,2], Anishan Vamadevan[1], Christian B. Laursen[3,4], Flemming Bjerrum[1,2,5], Suveer Singh[6,7,8] and Lars Konge[1,2]

[1]Copenhagen Academy for Medical Education and Simulation (CAMES), Center for HR & Education, The Capital Region of Denmark, Copenhagen, Denmark. [2]Department of Clinical Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark. [3]Department of Respiratory Medicine, Odense University Hospital, Odense, Denmark. [4]Odense Respiratory Research Unit (ODIN), Department of Clinical Research, University of Southern Denmark, Odense, Denmark. [5]Gastrounit, Surgical Section, Copenhagen University Hospital – Amager and Hvidovre, Hvidovre, Denmark. [6]Chelsea and Westminster Hospital, London, UK. [7]Royal Brompton Hospital, London, UK. [8]Faculty of Medicine, Imperial College London, London, UK.

Corresponding author: Kristofer Mazanti Cold (kristoffer.mazanti.cold.01@regionh.dk)

Shareable abstract (@ERSpublications)
AI appears to be a promising tool for enhancing bronchoscopy performance; however, implementation studies evaluating its impact in clinical settings are urgently needed
https://bit.ly/41z9sAh

Cite this article as: Cold KM, Vamadevan A, Laursen CB, *et al*. Artificial intelligence in bronchoscopy: a systematic review. *Eur Respir Rev* 2025; 34: 240274 [DOI: 10.1183/16000617.0274-2024].

Received: 4 Dec 2024
Accepted: 23 Feb 2025

Abstract

Background Artificial intelligence (AI) systems have been implemented to improve the diagnostic yield and operators' skills within endoscopy. Similar AI systems are now emerging in bronchoscopy. Our objective was to identify and describe AI systems in bronchoscopy.

Methods A systematic review was performed using MEDLINE, Embase and Scopus databases, focusing on two terms: bronchoscopy and AI. All studies had to evaluate their AI against human ratings. The methodological quality of each study was assessed using the Medical Education Research Study Quality Instrument (MERSQI).

Results 1196 studies were identified, with 20 passing the eligibility criteria. The studies could be divided into three categories: nine studies in airway anatomy and navigation, seven studies in computer-aided detection and classification of nodules in endobronchial ultrasound, and four studies in rapid on-site evaluation. 16 were assessment studies, with 12 showing equal performance and four showing superior performance of AI compared with human ratings. Four studies within airway anatomy implemented their AI, all favouring AI guidance to no AI guidance. The methodological quality of the studies was moderate (mean MERSQI 12.9 points, out of a maximum 18 points).

Interpretation 20 studies developed AI systems, with only four examining the implementation of their AI. The four studies were all within airway navigation and favoured AI to no AI in a simulated setting. Future implementation studies are warranted to test for the clinical effect of AI systems within bronchoscopy.

Introduction

Artificial intelligence (AI) is increasingly replacing and supplementing human tasks, with new AI applications to be beneficial in increasing the effectiveness of bronchoscopy. AI applications in other endoscopic procedures such as colonoscopy and gastroscopy have already proven their effect by increasing the detection rate of tumours [1, 2] through computer-aided detection (CADe) and computer-aided diagnosis (CADx), where an AI identifies and classifies lesions [1], and computer-aided quality assurance, where the AI provides the endoscopist with feedback to increase their technical abilities [3]. As novices have lower yields of diagnostic biopsy material, higher complication rates and increased patient discomfort [4–6], several efforts have been made to improve bronchoscopy's safety and diagnostic yield. Emerging techniques in navigational bronchoscopy such as electromagnetic navigation bronchoscopy and robotic bronchoscopy increase the diagnostic yield [7–12]; however, the summary diagnostic yield is still only 77.5% [7]. Rapid on-site evaluation (ROSE) of biopsy material can also increase the diagnostic yield [13, 14] and decrease the need for additional bronchoscopy procedures [15]. However, it requires an on-site trained cytopathologist. The proposed efforts towards the best possible yield seem to come with the same limitation: the need for highly specialised and skilled healthcare personnel [4–6, 16–19]. AI can assist in developing the required skills by potentially replacing more supervision from more experienced bronchoscopists for correct navigation and identification of tumours as seen in endoscopy or replace the cytopathologists using ROSE to improve the diagnostic yield of bronchoscopy. The evidence is still scarce, and no combined overview exists.

This systematic review aims to identify and describe AI-based systems in bronchoscopy.

Methods

The systematic review adhered to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA-AI) guidelines (see supplementary material for the PRISMA-AI checklist) [20]. The systematic literature review utilised searches in MEDLINE (*via* PubMed), Embase and Scopus. The search strategy, developed by the primary investigator (K.M. Cold) in collaboration with an information specialist from the University of Copenhagen, employed Boolean operators and focused on the two key elements of the research question (bronchoscopy and AI). The initial search string was developed for PubMed and adapted for Embase and Scopus (see supplementary material for applied search strings).

All retrieved articles were imported into Covidence systematic review software (Veritas Health Innovation, Melbourne, Australia) to identify and remove duplicates. Subsequently, two authors (K.M. Cold and A. Vamadevan) independently screened the remaining articles by title and abstract, followed by full-text assessment for eligibility and then data extraction. Any discrepancies were resolved through consensus between K.M. Cold and A. Vamadevan; in cases where consensus could not be reached, a third author (L. Konge) provided the final decision. Just before submission, the search was updated to include all eligible articles published since the initial search to ensure a timely review. These articles underwent the same screening process and data extraction as the initial search.

Additionally, forward and backward citation searches were employed, meaning that articles citing the included study and the references of those included studies were assessed for eligibility through the Web of Science platform (Clarivate, Philadelphia, PA, USA). References of identified systematic reviews were also screened for additional relevant studies.

Eligibility criteria

All original studies investigating or developing adaptable AI doing the bronchoscopy procedure from insertion to extraction of the scope were considered for inclusion. To meet eligibility criteria, articles would have to satisfy the requirements of the inclusion and exclusion criteria.

Inclusion criteria

The two inclusion criteria adhered to the two elements of the search string (supplementary material).

Inclusion criteria 1 (bronchoscopy). The study must focus on any part of the bronchoscopy procedure from insertion to extraction of the scope. Pre-procedure route planning, risk analysis and virtual bronchoscopy were therefore not included. Both pre-clinical and clinical studies would be included along with procedures as part of the bronchoscopy, such as ROSE.

Inclusion criteria 2 (AI). Defining AI systems can be difficult. We used a broad definition through automatic image or video identification, relying on a training and test dataset. The AI had to be tested by medical staff (medical students, doctors, *etc.*) and compared with human raters to determine whether it could replace or assist human decision making.

Exclusion criteria

Conference abstracts and articles not available in English were excluded.

Data extraction

Data extracted from the included studies were categorised into four tables to summarise the findings. Table 1 summarises the general study characteristics, including year of publication, country, study design, *etc.* Table 2 summarises the descriptions of identified AI system categories: the AI systems were categorised and their standard features within their category were described. Table 3 summarises the development of the AI systems: data sources used for training, internal and external validation, testing, and comparison to human ratings. Table 4 summarises the performance of the AI systems: reported AI performance metrics for internal and external validation were chosen based on a model card for developing AI within endoscopy [21].

TABLE 1 General study characteristics

Study	Year	Category	Country	Setting: clinical or simulation	Patient population or phantom (amount)	Bronchoscopies collected (n)	Endoscopists performing the procedure	Mean bronchoscopies per endoscopist (n)	Study design
Ozcelik [32]	2020	CADx in EBUS	Turkey	Clinical	NA	NA	4 experienced bronchoscopists	NA	Observational study
Matava [23]	2020	Airway anatomy and navigation	Canada	Clinical	Age 1–23 years	NA	NA	NA	Observational study
Yoo [24]	2021	Airway anatomy and navigation	South Korea	Clinical	NA	3216	NA	NA	Observational study
Lin [39]	2021	ROSE	Taiwan (China)	Clinical	Age 23–92 years	104	Bronchoscopists with >10 years experience	104	Observational study
Li [33]	2021	CADx in EBUS	China	Clinical	NA	NA	NA	NA	Observational study
Ai [41]	2022	ROSE	China	Clinical	NA	NA	NA	NA	Observational study
Li [25]	2022	Airway anatomy and navigation	China	Clinical	NA	342	NA	NA	Observational study
Hotta [34]	2022	CADx in EBUS	Japan	Clinical	NA	NA	NA	NA	Observational study
Wang [40]	2022	ROSE	Taiwan (China)	Clinical	NA	NA	NA	NA	Observational study
Yong [35]	2022	CADx in EBUS	South Korea	Clinical	Patients suspected of lung cancer referred for EBUS	310	1 experienced bronchoscopist	310	Observational study
Zhou [38]	2023	CADx in EBUS (EB-OCT)	China	Clinical	Asthma or COPD	25	NA	NA	Observational study
Xu [36]	2023	CADx in EBUS elastography	China	Clinical	NA	NA	NA	NA	Observational study
Chen [26]	2023	Airway anatomy and navigation	China	Clinical	Patients with normal bronchoscopy anatomy >14 years old	200	NA	NA	Observational study
Cold [27]	2024	Airway anatomy and navigation	Denmark	Simulation	Bronchoscopy phantom[¶]	24	24 untrained medical students	1	RCT

Continued

https://doi.org/10.1183/16000617.0274-2024

TABLE 1 Continued

Study	Year	Category	Country	Setting: clinical or simulation	Patient population or phantom (amount)	Bronchoscopies collected (n)	Endoscopists performing the procedure	Mean bronchoscopies per endoscopist (n)	Study design
Yan [42]	2024	ROSE	China	Clinical	NA	575	NA	NA	Observational study
Zhang [31]	2024	Airway anatomy and navigation	China	Simulation#	Porcine model	NA	NA	NA	Prospective non-randomised trial
Cold [28]	2024	Airway anatomy and navigation	Denmark	Simulation	Bronchoscopy phantom¶ (52 bronchoscopies)	52	52 bronchoscopists of varying experience level	1	Observational study (prospective)
Cold [29]	2024	Airway anatomy and navigation	Denmark	Simulation	Bronchoscopy phantom¶ (202 bronchoscopies)	202	101 bronchoscopists of varying experience level	2	RCT
Ervik [37]	2024	CADx in EBUS	Norway	Clinical	Referred for EBUS	30	2 pulmonologists with experience of >500 EBUS procedures conducted all study acquisitions	15	Observational study
Cold [30]	2024	Airway anatomy and navigation	Denmark	Simulation	Bronchoscopy phantom¶ (24 bronchoscopies)	24	24 untrained medical students	1	RCT
Mean (range) or summary number of studies	2023 (2020–2024)	Airway anatomy and navigation: 9 CADx in EBUS: 7 ROSE: 4	Asia: 14 Europe: 5 North America: 1	Clinical: 15 Simulation: 5	Patient population: 6 Porcine model: 1 Bronchoscopy phantom¶: 4 NA: 9	425 (24–3216)	26 (1–101)	62 (1–310)+	Observational: 16 Prospective non-randomised: 1 RCT: 3

CADx: computer-aided diagnosis; EBUS: endobronchial ultrasound; NA: not applicable (not included by the authors in the article); ROSE: rapid on-site evaluation; EB-OCT: endobronchial optical coherence tomography; RCT: randomised controlled trial. #: the study was also conducted in a porcine model, but since no human subjects were used, it was classified as simulation; ¶: Ultrasonic Bronchoscopy Simulator LM-099 (Koken, Tokyo, Japan); +: non-weighted mean of bronchoscopies per bronchoscopist.

https://doi.org/10.1183/16000617.0274-2024

TABLE 2 Description and primary findings of identified artificial intelligence (AI) applications

AI applications	Description and subclassification	Primary findings
Airway anatomy and navigation	Real-time navigation [27, 29, 30]: three RCTs tested the use of a real-time navigational system that identifies bronchial segments in a simulated setting, and one study assessed its performance compared to expert ratings	Navigation was superior with than without AI for all experience levels; mastery learning was superior to directed self-regulated learning
	Bronchial identification of images [23–26, 28]: four studies developed AI to identify bronchial anatomy from the glottis to the bronchial segmental level [23–26]; one study gathered validity evidence for the AI used in the included RCTs [28]	AI showed equal or superior performance to experienced bronchoscopists in classifying bronchoscopy images to their correct anatomical location; AI could accurately assess performance in a simulated setting
	AI copilot [31]: one study developed an AI copilot for steering to the fifth bronchial level	One junior bronchoscopists using the AI copilot was able to navigate and steer as well as a senior bronchoscopists not using it in a porcine model
CADe or CADx in EBUS	CADe or CADx in EBUS: six studies developed CADe to identify or CADx to classify malignant or benign LNs in EBUS [32–37]	AI showed equal or superior performance to experienced bronchoscopists in predicting LNs as malignant or benign based on the EBUS image
	EBUS elastography: one study applied EBUS elastography to check illness progression in asthma and chronic obstructive lung disease [38]	AI showed equal performance to experienced bronchoscopists in predicting wall thickness based on the EBUS image
ROSE	Four studies [39–42] developed AI in ROSE to detect malignant from benign lesions	AI showed equal or inferior performance to experienced histopathologists detecting malignant from benign lesions

RCT: randomised controlled trial; CADe: computer-aided detection; CADx: computer-aided diagnosis; LN: lymph node; EBUS: endobronchial ultrasound; ROSE: rapid on-site evaluation.

Quality assessment

Each included study was evaluated using the Medical Education Research Study Quality Instrument (MERSQI) on an 18-point scale [22]. The included AIs function as assessment systems, best analysed through a validity framework. We chose this instrument to analyse their competence assessment capabilities as done in a similar review [3].

Results

Study selection

The searches were conducted on 19 February 2024 and 7 November 2024, identifying 1196 studies after removal of duplicates (see figure 1 for PRISMA flowchart). No additional studies were identified through forward and backward citation searches. 20 studies passed the eligibility criteria and could be categorised into the following three categories: nine studies within airway anatomy and navigation [23–31], seven studies within CADx in endobronchial ultrasound (EBUS) [32–38], and four studies within ROSE [39–42] (see table 2 for a description of the categories).

General study characteristics

General study characteristics are presented in table 1. The studies were recent publications from 2020 to 2024, spanning three continents (Asia: 14 studies, Europe: 5 studies and North America: 1 study). 18 studies were single-centre studies and two studies utilised data from two institutions [29, 39].

Categorisation, description and performance of the AIs

Details of the categorisation, description and performance comparisons of the AIs are presented in tables 2 and 3.

In the category of airway anatomy, three AIs showed equal performance [23, 24, 28] and two AIs showed superior performance [25, 26] to experienced bronchoscopists' ratings, indicating AI can be trained to recognise images of anatomical locations better than experienced raters.

In CADx in EBUS, five AIs showed equal performance [32, 35–38] and two AIs showed superior performance [33, 34] to experienced bronchoscopists' ratings in classifying lymph nodes (LNs) as benign or malignant based on EBUS images.

In ROSE, all four AIs showed equal or inferior performance [39–42] to experienced histopathologists' ratings.

https://doi.org/10.1183/16000617.0274-2024

TABLE 3 Development of the artificial intelligence (AI) systems

Study	AI capability	Procedures for training of AI	Procedures for validation of AI	Procedures for testing of AI (internal)	Procedures for testing of AI (external; video)	Human comparator	Assessment or implementation	Comparison to comparator
Ozcelik [32]	CADe: malignant or non-malignant LN	300 LN images	NA	45 LN images	NA	Performing bronchoscopist	Assessment	Equal
Matava [23]	Airway anatomy: identification of tracheal rings and vocal cords	395 bronchoscopy images	NA	62 bronchoscopy images	1 bronchoscopy video	Performing bronchoscopist	Assessment	Equal
Yoo [24]	Airway anatomy: identification of carina, right main bronchus and left main bronchus	6806 bronchoscopy images from 3216 patients	1191 images from 3216 patients	511 images from 3216 patients	NA	3 anaesthesiologists and 3 pulmonologists with varying experience levels	Assessment	Equal
Lin [39]	ROSE: malignant or non-malignant	499 ROSE images from 97 patients	NA	66 ROSE images from 14 patients	NA	Experienced cytopathologist	Assessment	Equal
Li [33]	CAD: malignant or non-malignant LN	Number of images not reported from 245 LNs	NA	Number of images not reported from 49 LNs	NA	3 experienced and 3 trainee bronchoscopists	Assessment	Superior
Ai [41]	ROSE: malignant or non-malignant	374 ROSE images	91 ROSE images	162 ROSE images	NA	1 experienced and 2 junior cytopathologists	Assessment	Inferior to expert cytopathologist, superior to junior cytopathologist
Li [25]	Airway anatomy: 31 anatomical locations	22 746 bronchoscopy images from 342 bronchoscopies	5695 bronchoscopy images	NA	372 images from 12 bronchoscopies	4 senior and 4 junior bronchoscopists	Assessment	Superior: improves bronchoscopists' identification of images regarding anatomical locations of the bronchial tree with feedback from AI
Hotta [34]	CAD in EBUS: malignant or non-malignant LN	2 421 360 EBUS images from 171 peripheral pulmonary lesions	NA	26 674 EBUS images from 42 peripheral pulmonary lesions	42 peripheral pulmonary lesions	2 experienced and 2 trainee bronchoscopists	Assessment	Superior
Wang [40]	ROSE: malignant or non-malignant	46 ROSE images	NA	76 ROSE images	NA	2 experienced cytopathologists	Assessment	Equal
Yong [35]	CAD in EBUS: malignant or non-malignant LN	2394 images from 310 patients	NA	NA	NA	2 experienced bronchoscopists	Assessment	Equal
Zhou [38]	CAD in EBUS: wall thickness using OCT	13 078 images from 25 patients	1620 images from 3 patients	2700 images from 5 patients	NA	2 experienced bronchoscopists' manual segmentation	Assessment	Equal
Xu [36]	CAD in EBUS: malignant or non-malignant	Number of images not reported, 727 EBUS elastography videos	NA	NA	NA	3 experienced bronchoscopists	Assessment	Equal

Continued

https://doi.org/10.1183/16000617.0274-2024

TABLE 3 Continued

Study	AI capability	Procedures for training of AI	Procedures for validation of AI	Procedures for testing of AI (internal)	Procedures for testing of AI (external; video)	Human comparator	Assessment or implementation	Comparison to comparator
Chen [26]	Airway anatomy: 9 anatomical locations	1527 images from 200 bronchoscopies	475 images from 72 bronchoscopies	50 images from the validation dataset	NA	21 bronchoscopists of varying experience level	Assessment	Superior
Cold [27][#]	Airway anatomy: 31 anatomical locations with navigational feedback scores	NA	NA	NA	24 bronchoscopies, inspecting the entire bronchial tree	AI to no navigational assistance	Implementation	AI superior to traditional training
Yan [42]	ROSE: malignant or non-malignant	4106 ROSE images from 575 patients	544 images from 347 patients	526 images from 337 patients	1181 ROSE images from 145 patients	1 expert cytopathologist	Assessment	Equal
Zhang [31]	Airway anatomy: path and co-steering assistant	NA	NA	NA	3 paths through the bronchial tree to the fifth bronchial level	Path and steering without AI assistance	Implementation	Making a novice perform as well as an expert in a porcine model
Cold [28][#]	Airway anatomy: 31 anatomical locations with navigational feedback scores	NA	NA	NA	52 bronchoscopies, inspecting the entire bronchial tree	2 blinded experienced bronchoscopists	Assessment	Equal; high level of validity evidence for assessing competence
Cold [29][#]	Airway anatomy: 31 anatomical locations with navigational feedback scores	NA	NA	NA	202 bronchoscopies, inspecting the entire bronchial tree	AI to no navigational assistance	Implementation	Superior to no navigation
Ervik [37]	EBUS: annotation of LNs and blood vessels	882 EBUS images from 30 bronchoscopies	145 EBUS images from 5 bronchoscopies	134 images from 5 bronchoscopies	NA	2 experienced bronchoscopists	Assessment	Equal
Cold [30][#]	Airway anatomy: 31 anatomical locations with navigational feedback scores	NA	NA	NA	24 bronchoscopies, inspecting the entire bronchial tree	Mastery learning to directed self-regulated learning	Implementation	Mastery learning superior to directed self-regulated learning
Mean (range) or summary number of studies	Airway anatomy: 9 CAD in EBUS: 7 ROSE: 4	190 347 (46–2 421 360)[¶]	1394 (91–5695)[¶]	2819 (45–26 674)[¶]	NA[+]	NA[+]	16 assessment 4 implementation	12 assessment studies found AI equal to and 4 studies found AI superior to human ratings 4 intervention studies were in favour of AI intervention

CADe: computer-aided detection; LN: lymph node; NA: not applicable (not included by the authors in the article); ROSE: rapid on-site evaluation; EBUS: endobronchial ultrasound; OCT: optical coherence tomography; [#]: the underlying AI is developed and patented by a private company, therefore their reporting metrics are not accessible and not reported, which is discussed in Cold *et al.* [29]; [¶]: only included studies that utilise images; [+]: mean and range not calculated due to heterogeneous data sources.

https://doi.org/10.1183/16000617.0274-2024

TABLE 4 Performance metrics of the artificial intelligence (AI) systems

Study	Internal validation						External validation					
	Accuracy (internal)	Sensitivity	Specificity	AUC	PPV	NPV	Accuracy (external)	Sensitivity	Specificity	AUC	PPV	NPV
Ozcelik [32]	0.82	0.89	0.72	0.83	0.81	0.78	NA	NA	NA	NA	NA	NA
Matava [23]	NA	0.85	0.98	NA	NA	NA	NA	NA	NA	NA	NA	NA
Yoo [24]	0.89	NA	NA	NA	NA	0.98	0.84	NA	NA	NA	NA	0.98
Lin [39]	0.99	0.99	0.99	0.99	0.98	NA	NA	NA	NA	NA	NA	NA
Li [33]	0.86	0.90	0.80	0.87	0.86	0.95	0.89	0.90	0.80	0.87	0.86	0.95
Ai [41]	0.85	0.96	0.97	NA	NA	0.98	NA	NA	NA	NA	NA	NA
Li [25]	0.97	NA	NA	NA	NA	0.96	0.54	NA	NA	NA	NA	NA
Hotta [34]	0.83	0.95	0.53	0.84	0.82	NA	0.83	1.00	0.42	0.81	1.00	NA
Wang [40]	NA	0.90	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Yong [35]	0.80	0.79	0.76	NA	NA	NA	NA	NA	NA	NA	NA	NA
Zhou [38]	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Xu [36]	0.81	0.85	0.75	0.88	0.76	0.87	NA	NA	NA	NA	NA	NA
Chen [26]	0.91	NA	NA	NA	NA	0.98	NA	NA	NA	NA	NA	NA
Cold [27][#]	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Yan [42]	0.93	0.94	0.90	0.95	0.90	NA	0.90	0.91	0.89	0.96	0.78	NA
Zhang [31]	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Cold [28][#]	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Cold [29][#]	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Ervik [37]	NA	0.71	0.99	NA	NA	NA	NA	NA	NA	NA	NA	NA
Cold [30][#]	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Non-weighted mean (range)	0.88 (0.80–0.99)	0.89 (0.71–0.99)	0.85 (0.78–0.99)	0.91 (0.83–0.99)	0.87 (0.76–0.98)	0.93 (0.78–0.98)	0.80 (0.54–0.9)	0.94 (0.90–1.00)	0.70 (0.42–0.89)	0.88 (0.81–0.96)	0.88 (0.78–1.00)	0.97 (0.95–0.98)

AUC: area under the curve; PPV: positive predictive value; NPV: negative predictive value; NA: not applicable (not included by the authors in the article), indicating the heterogenous and often missing reporting conducted by the studies. [#]: the underlying AI is developed and patented by a private company, therefore their reporting metrics are not accessible and not reported, which is discussed in Cold *et al.* [29].

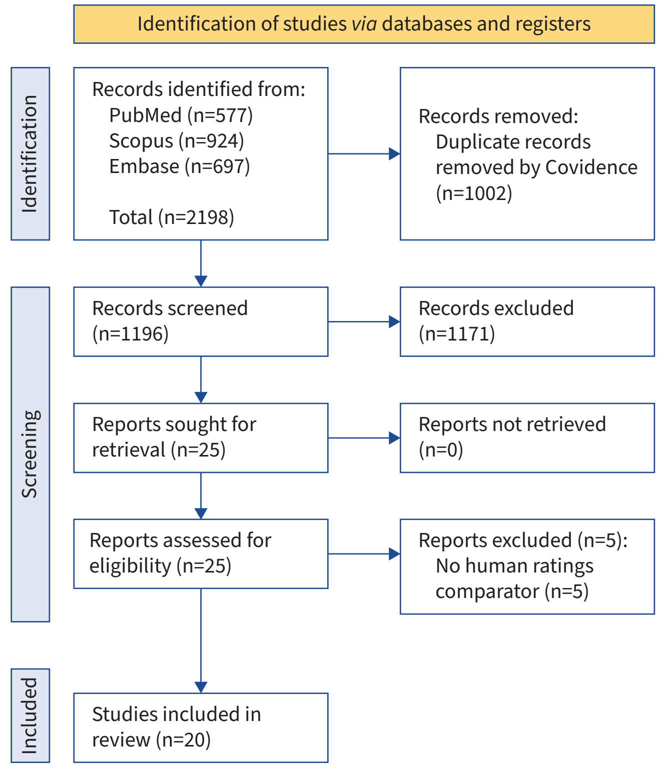

FIGURE 1 Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flowchart for the joint literature searches conducted on 19 February 2024 and 7 November 2024.

Four studies tested the implementation of AIs, all within the airway anatomy category [27, 29–31]. Three randomised controlled trials (RCTs) tested the implementation of the same AI airway navigational system, where the AI identified bronchial segments and kept track of inspected segments [27, 29, 30]. They showed that with just a limited amount of training with the AI (median training time 217 min), complete novices could systematically inspect the bronchial tree with the AI disabled [27]. Their performance was comparable to experts' performances [27]. The second RCT, with 101 participants spanning six continents, showed that bronchoscopists with experience of more than 500 bronchoscopies performed better than those without AI navigation [29]. Experienced bronchoscopists could inspect more segments in a more structured order, with less time spent navigating between the segments [29]. The third RCT showed mastery learning is superior to traditional directed self-regulated learning using the AI system [30]. The prospective non-randomised trial found that novices using an AI copilot could navigate the bronchial tree as well as experienced bronchoscopists in a porcine model [31].

The studies had a high range of images used for training (13 studies; mean 190 347, range 42–2 241 360), internal validation (seven studies; mean 1394, range 91–5695) and external validation (11 studies, mean 2819, range 45–26 674).

Performance metrics of the AIs
The studies used very heterogeneous reporting metrics of their AIs (table 4). Accuracy was the most frequently reported for internal validation (11 studies; mean 0.88, range 0.80–0.99) and for external validation (five studies; mean 0.80, range 0.54–0.90). Due to the low volume of intervention studies and heterogeneous study designs with varying outcome measures, a meta-analysis was not feasible.

Quality assessment
The MERSQI scores differed moderately (mean 12.9, range 11.5–16.5) (table 5). The three highest-scoring studies were the RCTs, scoring 15.5 [27], 15.5 [30] and 16.5 points [29]. All studies scored maximum points for "Response rate" (no studies reported on system failure and therefore all studies had an assumed response rate >75%), "Type of data" (inherent when using automated/AI rating) and "Validity (relationships to other variables)" (it was an inclusion criterion to compare to human ratings).

https://doi.org/10.1183/16000617.0274-2024

TABLE 5 Medical Education Research Study Quality Instrument (MERSQI) score

Study	Study design |3|	Sampling (institutions) |1.5|	Sampling (response rate) |1.5|	Type of data |3|	Validity (internal structure) |1|	Validity (content) |1|	Validity (relationships to other variables) |1|	Data analysis (appropriateness) |1|	Data analysis (complexity) |2|	Outcomes |3|	Total |18|
Ozcelik [32]	1	0.5	1.5	3	0	1	1	1	2	1.5	12.5
Matava [23]	1	0.5	1.5	3	0	0	1	1	2	1.5	11.5
Yoo [24]	1	0.5	1.5	3	0	0	1	1	2	1.5	11.5
Lin [39]	1	1	1.5	3	0	0	1	1	2	1.5	12
Li [33]	1	0.5	1.5	3	1	0	1	1	2	1.5	12.5
Ai [41]	1	0.5	1.5	3	1	1	1	1	2	1.5	13.5
Li [25]	1	0.5	1.5	3	0	0	1	1	2	1.5	11.5
Hotta [34]	1	0.5	1.5	3	0	1	1	1	2	1.5	12.5
Wang [40]	1	0.5	1.5	3	0	1	1	1	2	1.5	12.5
Yong [35]	1	0.5	1.5	3	0	1	1	1	2	1.5	12.5
Zhou [38]	1	0.5	1.5	3	0	1	1	1	2	1.5	12.5
Xu [36]	1	0.5	1.5	3	1	1	1	1	2	1.5	13.5
Chen [26]	1	0.5	1.5	3	0	0	1	1	2	1.5	11.5
Cold [27]	3	0.5	1.5	3	1	1	1	1	2	1.5	15.5
Yan [42]	1	0.5	1.5	3	1	1	1	1	2	1.5	13.5
Zhang [31]	2	0.5	1.5	3	0	1	1	0	1	1.5	11.5
Cold [28]	1	1.5	1.5	3	1	1	1	1	2	1.5	14.5
Cold [29]	3	1.5	1.5	3	1	1	1	1	2	1.5	16.5
Ervik [37]	1	0.5	1.5	3	0	0	1	1	2	1.5	11.5
Cold [30]	3	0.5	1.5	3	1	1	1	1	2	1.5	15.5
Mean (range)	1.4 (1–3)	0.6 (0.5–1.5)	1.5	3	0.4 (0–1)	0.7 (0–1)	1	1.0 (0–1)	2.0 (1–2)	1.5	12.9 (11.5–16.5)

Maximum points for each category indicated by |n|.

All studies except one scored maximum points for "Data appropriateness" and "Data complexity", as Zhang *et al.* [31] only included one participant in each group and therefore could not support their statements with statistical tests.

Discussion

This is the first systematic review of AI applications in bronchoscopy. We identified 20 studies within three elements of bronchoscopy: airway anatomy [23–31], CADx in EBUS [32–38] and ROSE [39–42].

Airway anatomy

Three RCTs in a simulated setting

In 2001, Colt *et al.* [43] highlighted a suboptimal performance by experienced physicians, as novices could surpass their performance in a simulated setting by training on a virtual reality (VR) bronchoscopy simulator. Ost *et al.* [44] showed that training on the same simulator translated into superior clinical performance than traditional apprenticeship clinical training. However, VR simulators often do not provide haptic feedback, and their automated metrics cannot always be used to assess competency [45]. With the first systematic review and meta-analysis from 2013 in simulation-based training of bronchoscopy, Kennedy *et al.* [46] concluded it to be effective, and that using plastic phantoms might be superior to VR simulators. AI has been proposed to fuse the best of VR simulators and phantoms by providing automated feedback metrics using real bronchoscopes providing haptic feedback [47]. An updated systematic review from 2023 confirmed simulation-based training in bronchoscopy to be effective [48]. However, no studies have followed Kennedy *et al.*'s [46] recommendations from 2013 on the use of mastery learning. Mastery learning is a recommended training principle where trainees practice until performance targets are met [49]. The third RCT in this systematic review showed mastery learning was superior to traditional directed self-regulated learning using the AI system [30]. The three RCTs conclude that AI provides novices with a large benefit from training [27], improves experienced bronchoscopists' performance [29] and is suitable with mastery learning as an effective training form [30]. In general, transfer studies to a clinical setting are sparse [48], and how training with AI translates into clinical performance needs to be tested in future studies [50, 51]. AI can induce dependence, causing users to rely on technology rather than honing their skills, a concept known as guidance theory [52], which can be detrimental to long-term learning [53]. For example, excessive use of Google Maps might hinder one's ability to navigate independently. Notably, the studies tested participants without AI assistance, highlighting its potential as a beneficial training tool without causing dependence.

Developing bronchial identification systems

Four studies developed bronchial identification systems based on clinical bronchoscopy images [23–26] and one study gathered validity evidence for the AI used in the RCTs [28]. Three AIs showed equal performance [23, 24, 28] and two AIs showed superior performance [25, 26] in classifying anatomical locations compared to experts. Such AIs might be developed into real-time clinical bronchial identification systems. However, testing the classification of still images of bronchial segments against expert ratings effectively evaluates the experts' assessments within the AI's environment. Flexible bronchoscopy is a dynamic procedure where bronchoscopists classify bronchial segments while manoeuvring through the bronchial tree, with the route behind an essential part of knowing your location. To add actual clinical benefit, AIs must be equal or superior in the real-life clinical environment to experts. If such AIs could be developed for real-life bronchial navigational systems, they could improve the performance of bronchoscopists, as illustrated in a simulated setting by the three RCTs [27, 29, 30]. To claim AI superiority should be done with the utmost caution, as the focus should be on improving patient outcomes and not just outperforming expert ratings in the AI training environment. This is highlighted by Li *et al.* [25], who trained an AI to classify 31 anatomical locations of the bronchial tree (glottis, trachea, carina and the third bronchial division). The AI's low test accuracies (0.54 overall, 0.34 for segmental bronchi) highlight how similar bronchial segments challenge even AI recognition [25]. However, the four expert raters had a lower accuracy (mean 0.39, range 0.35–0.42) but were able to improve their accuracy to the AI's level when receiving a recommendation from the AI (mean 0.54, range 0.52–0.58) [25]. The three other studies only trained their AIs to recognise bronchial anatomy to the carina [23], main bronchial [24] or lobar bronchial level [26]. None of the AIs were trained to identify bronchial lumens beyond the segmental level (RB1a, RB1b, *etc.*), where it might be more useful in reaching peripheral tumours. Li *et al.* [25] provided the most elaborate AI to one day function real-time in a clinical setting. However, with an accuracy of only 0.34 at the segmental bronchial level, much improvement is needed before the system will benefit bronchoscopists [25].

AI copilot

Zhang *et al.* [31] developed an AI copilot controlled through a remote console. They compared the performance of a junior doctor (less than 500 bronchoscopies) in steering to the fifth bronchial level on a porcine model, using the AI copilot, to a chief doctor (more than 4000 bronchoscopies), not using it. They could both follow the designated routes, but the authors concluded the junior doctor had safer steering [31]. The study only used descriptive statistics, and to make conclusions based on just two study participants seems unfeasible. The development appears interesting in allowing safer peripheral steering. Still, it needs to be tested on patients and in the hands of more doctors, like the similar Monarch (Johnson & Johnson, New Brunswick, NJ, USA) and Ion (Intuitive Surgical, Sunnyvale, CA, USA) applications for robot-assisted bronchoscopy [54].

CADx in EBUS

Six studies developed CADe/CADx to detect and/or discriminate malignant from benign LNs in EBUS [32–37], and one study applied EBUS elastography to check illness progression in asthma and chronic obstructive lung disease [38]. CADe was used as early as 1992 to help detect malignancy in breast ultrasound [55], with several AI systems developed that decreased the rate of unnecessary biopsies [56, 57]. In endoscopic ultrasound, a meta-analysis found CADe-assisted procedures superior to non-CADe-assisted procedures in detecting subepithelial lesions [58]. CADe has, therefore, proven its potential in other ultrasound disciplines to improve the detection of malignancy and lower the rate of biopsies from benign lesions. Recommendations have been made as to when to biopsy LNs in EBUS [59], and AI could help in this assessment by identifying malignant LNs and deciding when not to biopsy due to an overwhelming chance of benign diagnosis. Five studies found their AI equal [32, 35–38] and two studies found their AI superior [33, 34] to human ratings. The two studies showing superiority had the same limitations as discussed in the airway anatomy section, where bronchoscopists were tested in the AI's training environment by only assessing images of LNs [33, 34]. Therefore, while CADx could prove helpful in EBUS, it is beneficial to consider experiences from other endoscopic procedures before conducting implementation studies. A meta-analysis in colonoscopy on the CADx-based "resect and discard" strategy found no benefit or harm compared to sending polyps to histopathology, questioning its clinical value and emphasising the need for improved CADx accuracy and explainability [60]. Additionally, will we rely on an AI prediction, when we cannot remove the potential malignancy but only biopsy it as in EBUS? And if wrong, who is held liable? The first US Food and Drug Administration-approved AI in healthcare, IDx-DR (Digital Diagnostics, Coralville, IA, USA), detects diabetic retinopathy, and as an autonomous AI the company assumes liability [61, 62]. AI regulations have swiftly changed in the last years. They will probably do so in the future, leaving much uncertainty for the potential implementation of CADx in EBUS studies. Strong evidence from implementation studies is needed if we are to depend on an AI suggestion not to biopsy an LN in EBUS.

ROSE

Four studies assessed the use of AI in ROSE, with three studies showing equal performance to experienced histopathologists' ratings [39, 40, 42], and one study showing inferior performance of the AI to experienced histopathologist rating [41]. The studies indicate AI can potentially replace histopathologists in widening the use of ROSE. ROSE does not improve the diagnostic yield but is associated with fewer needle passes during EBUS-guided transbronchial needle aspiration and a lower requirement for additional bronchoscopy procedures to make a final diagnosis [15]. However, ROSE depends on an on-site cytopathologist, providing a good example of AI potentially widening the distribution of ROSE. Unfortunately, none of the studies tested how ROSE would impact their bronchoscopy suite or performance in a prospective way, but they were all assessment studies. Three studies showed equal performance [39, 40, 42] to expert cytopathologists and one study showed inferior performance [41]. The only commonly reported metric for all studies was sensitivity for internal validation (non-weighted mean 0.95, range 0.9–0.99) and only one study performed external validation (sensitivity drop of 9.5%) [42]. A systematic review of AI in radiology found a pooled performance drop of 6% at external validation [63]. As in the current study, it was also found that the reporting metrics were highly heterogeneous, with accuracy only reported in 15% of studies [63]. Other endoscopic societies have, therefore, issued guidelines when developing and reporting on the performance of AI [21]. We recommend following these guidelines until thoracic societies issue similar recommendations.

ROSE generally has a pooled sensitivity of 0.92 and a pooled specificity of 0.95 for diagnosing lung cancer during bronchoscopy [64]. The room for improvement is therefore limited, but as AI potentially can replace a human cytopathologist, this is an interesting area for AI to expand a technology that is not widely available.

Limitations and future perspectives

Several limitations and future perspectives have been raised throughout the discussion of the individual categories; however, the included studies had more general limitations. The studies had a moderate range in MERSQI scores (mean 12.9, range 11.5–16.5), with the RCTs scoring highest, indicating a general lack of implementation studies. Only one study used a validity framework [28]. In particular, Zhang *et al*. [31]

could have benefitted from using one, as general conclusions cannot be made from comparing one junior to one senior doctor, which is reflected in the "Data analysis" aspect of MERSQI, where it was the only study not to score maximum points. AI can amplify the bias it is trained on [65–67]. It is, therefore, important to include several institutions and raters, whereas most studies only included data from one institution and one rater for training. Less than half the studies scored points for "Validity (internal structure)", which could be made by assessing inter-rater reliability. Validity frameworks help to address all these issues to design the best study for gathering validity evidence for assessment tools. This review used MERSQI for methodological quality assessment, as it can be broadly applied across AI applications [3]. Future reviews focusing on diagnostic accuracy for CADx in EBUS could apply more specific tools like Quality Assessment of Diagnostic Accuracy Studies (QUADAS) [68]. This review only included AIs to be applied during the bronchoscopy. Pre-procedure route planning, risk analysis and virtual bronchoscopy were not included, even though they are important parts of the bronchoscopy procedure.

Only one of the studies on diagnostic accuracy [35] used a relevant reporting guideline (Standards for Reporting of Diagnostic Accuracy Studies (STARD)) [69]. This highlights the heterogeneous study design and reporting metrics of the included studies, which made a meta-analysis unfeasible for the categories identified. Bronchoscopist experience, patient population and AI performance metrics, among other important aspects, were not consequently reported by the studies, highlighted by the many "NA" entries in the summary tables. We therefore highly recommend future studies to follow the most recent guidelines and report both within the assessment and AI literature.

Conclusions

AI was equal or superior to expert ratings within airway anatomy in identifying the correct segments, and in CADx in EBUS and ROSE in classifying LNs as benign or malignant. AI-supported navigation through the bronchial tree is the only application that has been implemented. However, no human or transfer trials have been conducted. Future studies should adhere to recommended guidelines within developing assessment tools and AI.

Points for clinical practice

- AI has been implemented in other endoscopic disciplines to increase the performance of the endoscopist.
- This systematic review identified AIs within airway anatomy and navigation, CADe/CADx of nodules in EBUS, and ROSE, which showed promising results in improving bronchoscopy.

Questions for future research

- Implementation of the identified AIs in clinical studies is highly warranted.
- Future studies should follow current recommendations for developing and reporting AI.

Provenance: Submitted article, peer reviewed.

Conflict of interest: The authors have no conflicts of interest to disclose.

References

1 Hassan C, Spadaccini M, Iannone A, *et al.* Performance of artificial intelligence in colonoscopy for adenoma and polyp detection: a systematic review and meta-analysis. *Gastrointest Endosc* 2021; 93: 77–85.
2 El Hajjar A, Rey JF. Artificial intelligence in gastrointestinal endoscopy: general overview. *Chin Med J (Engl)* 2020; 133: 326–334.
3 Cold KM, Vamadevan A, Vilmann AS, *et al.* Computer-aided quality assessment of endoscopist competence during colonoscopy: a systematic review. *Gastrointest Endosc* 2024; 100: 167–176.
4 Hsu LH, Liu CC, Ko JS. Education and experience improve the performance of transbronchial needle aspiration: a learning curve at a cancer center. *Chest* 2004; 125: 532–540.
5 Ouellette DR. The safety of bronchoscopy in a pulmonary fellowship program. *Chest* 2006; 130: 1185–1190.
6 Stather DR, MacEachern P, Chee A, *et al.* Trainee impact on procedural complications: an analysis of 967 consecutive flexible bronchoscopy procedures in an interventional pulmonology practice. *Respiration* 2013; 85: 422–428.
7 Kops SEP, Heus P, Korevaar DA, *et al.* Diagnostic yield and safety of navigation bronchoscopy: a systematic review and meta-analysis. *Lung Cancer* 2023; 180: 107196.

8 Wang Memoli JS, Nietert PJ, Silvestri GA. Meta-analysis of guided bronchoscopy for the evaluation of the pulmonary nodule. *Chest* 2012; 142: 385–393.

9 Ali MS, Trick W, Mba BI, *et al.* Radial endobronchial ultrasound for the diagnosis of peripheral pulmonary lesions: a systematic review and meta-analysis. *Respirology* 2017; 22: 443–453.

10 Kuijvenhoven JC, Leoncini F, Crombag LC, *et al.* Endobronchial ultrasound for the diagnosis of centrally located lung tumors: a systematic review and meta-analysis. *Respiration* 2020; 99: 441–450.

11 Zhang W, Chen S, Dong X, *et al.* Meta-analysis of the diagnostic yield and safety of electromagnetic navigation bronchoscopy for lung nodules. *J Thorac Dis* 2015; 7: 799–809.

12 Ali MS, Ghori UK, Wayne MT, *et al.* Diagnostic performance and safety profile of robotic-assisted bronchoscopy: a systematic review and meta-analysis. *Ann Am Thorac Soc* 2023; 20: 1801–1812.

13 Chen CH, Cheng WC, Wu BR, *et al.* Improved diagnostic yield of bronchoscopy in peripheral pulmonary lesions: combination of radial probe endobronchial ultrasound and rapid on-site evaluation. *J Thorac Dis* 2015; 7: Suppl. 4, S418–S425.

14 Gex G, Pralong JA, Combescure C, *et al.* Diagnostic yield and safety of electromagnetic navigation bronchoscopy for lung nodules: a systematic review and meta-analysis. *Respiration* 2014; 87: 165–176.

15 Sehgal IS, Dhooria S, Aggarwal AN, *et al.* Impact of rapid on-site cytological evaluation (ROSE) on the diagnostic yield of transbronchial needle aspiration during mediastinal lymph node sampling: systematic review and meta-analysis. *Chest* 2018; 153: 929–938.

16 Rodrigues I, Estêvão Gomes R, Coutinho LM, *et al.* Diagnostic yield and safety of transbronchial lung cryobiopsy and surgical lung biopsy in interstitial lung diseases: a systematic review and meta-analysis. *Eur Respir Rev* 2022; 31: 210280.

17 Verhoeven RLJ, van der Sterren W, Kong W, *et al.* Cone-beam CT and augmented fluoroscopy-guided navigation bronchoscopy: radiation exposure and diagnostic accuracy learning curves. *J Bronchology Interv Pulmonol* 2021; 28: 262–271.

18 Toennesen LL, Vindum HH, Risom E, *et al.* Learning curves for electromagnetic navigation bronchoscopy using CUSUM analysis. *J Bronchology Interv Pulmonol* 2022; 29: 164–170.

19 Toennesen LL, Vindum HH, Risom E, *et al.* When pulmonologists are novice to navigational bronchoscopy, what predicts diagnostic yield? *Diagnostics* 2022; 12: 3127.

20 Page MJ, McKenzie JE, Bossuyt PM, *et al.* The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ* 2021; 372: n71.

21 Parasa S, Repici A, Berzin T, *et al.* Framework and metrics for the clinical use and implementation of artificial intelligence algorithms into endoscopy practice: recommendations from the American Society for Gastrointestinal Endoscopy Artificial Intelligence Task Force. *Gastrointest Endosc* 2023; 97: 815–824.

22 Reed DA, Cook DA, Beckman TJ, *et al.* Association between funding and quality of published medical education research. *JAMA* 2007; 298: 1002–1009.

23 Matava C, Pankiv E, Raisbeck S, *et al.* A convolutional neural network for real time classification, identification, and labelling of vocal cord and tracheal using laryngoscopy and bronchoscopy video. *J Med Syst* 2020; 44: 44.

24 Yoo JY, Kang SY, Park JS, *et al.* Deep learning for anatomical interpretation of video bronchoscopy images. *Sci Rep* 2021; 11: 23765.

25 Li Y, Zheng X, Xie F, *et al.* Development and validation of the artificial intelligence (AI)-based diagnostic model for bronchial lumen identification. *Transl Lung Cancer Res* 2022; 11: 2261–2274.

26 Chen C, Herth FJ, Zuo Y, *et al.* Distinguishing bronchoscopically observed anatomical positions of airway under by convolutional neural network. *Ther Adv Chronic Dis* 2023; 14: 20406223231181495.

27 Cold KM, Xie S, Nielsen AO, *et al.* Artificial intelligence improves novices' bronchoscopy performance: a randomized controlled trial in a simulated setting. *Chest* 2024; 165: 405–413.

28 Cold KM, Agbontaen K, Nielsen AO, *et al.* Artificial intelligence for automatic and objective assessment of competencies in flexible bronchoscopy. *J Thorac Dis* 2024; 16: 5718–5726.

29 Cold KM, Agbontaen K, Nielsen AO, *et al.* Artificial intelligence improves bronchoscopy performance: a randomized crossover trial. *ERJ Open Res* 2025; 11: 00395-2024.

30 Cold KM, Wei W, Agbontaen K, *et al.* Mastery learning guided by AI is superior to directed self-regulated learning in flexible bronchoscopy training – an RCT. *Respiration* 2024; 104: 206–215.

31 Zhang J, Liu L, Xiang P, *et al.* AI co-pilot bronchoscope robot. *Nat Commun* 2024; 15: 241.

32 Ozcelik N, Ozcelik AE, Bulbul Y, *et al.* Can artificial intelligence distinguish between malignant and benign mediastinal lymph nodes using sonographic features on EBUS images? *Curr Med Res Opin* 2020; 36: 2019–2024.

33 Li J, Zhi X, Chen J, *et al.* Deep learning with convex probe endobronchial ultrasound multimodal imaging: a validated tool for automated intrathoracic lymph nodes diagnosis. *Endosc Ultrasound* 2021; 10: 361–371.

34 Hotta T, Kurimoto N, Shiratsuki Y, *et al.* Deep learning-based diagnosis from endobronchial ultrasonography images of pulmonary lesions. *Sci Rep* 2022; 12: 13710.

35 Yong SH, Lee SH, Oh SI, *et al.* Malignant thoracic lymph node classification with deep convolutional neural networks on real-time endobronchial ultrasound (EBUS) images. *Transl Lung Cancer Res* 2022; 11: 14–23.

36 Xu M, Chen J, Li J, *et al.* Automatic representative frame selection and intrathoracic lymph node diagnosis with endobronchial ultrasound elastography videos. *IEEE J Biomed Health Inform* 2023; 27: 29–40.
37 Ervik Ø, Tveten I, Hofstad EF, *et al.* Automatic segmentation of mediastinal lymph nodes and blood vessels in endobronchial ultrasound (EBUS) images using deep learning. *J Imaging* 2024; 10: 190.
38 Zhou ZQ, Guo ZY, Zhong CH, *et al.* Deep learning-based segmentation of airway morphology from endobronchial optical coherence tomography. *Respiration* 2023; 102: 227–236.
39 Lin CK, Chang J, Huang CC, *et al.* Effectiveness of convolutional neural networks in the interpretation of pulmonary cytologic images in endobronchial ultrasound procedures. *Cancer Med* 2021; 10: 9047–9057.
40 Wang CW, Khalil MA, Lin YJ, *et al.* Deep learning using endobronchial-ultrasound-guided transbronchial needle aspiration image to improve the overall diagnostic yield of sampling mediastinal lymphadenopathy. *Diagnostics* 2022; 12: 2234.
41 Ai D, Hu Q, Chao Y-C, *et al.* Artificial intelligence-based rapid on-site cytopathological evaluation for bronchoscopy examinations. *Intell Based Med* 2022; 6: 100069.
42 Yan S, Li Y, Pan L, *et al.* The application of artificial intelligence for rapid on-site evaluation during flexible bronchoscopy. *Front Oncol* 2024; 14: 1360831.
43 Colt HG, Crawford SW, Galbraith O 3rd. Virtual reality bronchoscopy simulation: a revolution in procedural training. *Chest* 2001; 120: 1333–1339.
44 Ost D, DeRosiers A, Britt EJ, *et al.* Assessment of a bronchoscopy simulator. *Am J Respir Crit Care Med* 2001; 164: 2248–2255.
45 Colella S, Sondergaard Svendsen MB, Konge L, *et al.* Assessment of competence in simulated flexible bronchoscopy using motion analysis. *Respiration* 2015; 89: 155–161.
46 Kennedy CC, Maldonado F, Cook DA. Simulation-based bronchoscopy training: systematic review and meta-analysis. *Chest* 2013; 144: 183–192.
47 Cold KM, Konge L. Simulation-based training in flexible bronchoscopy: best practices and future directions. *Chest* 2023; 134: 820–821.
48 Gerretsen ECF, Chen A, Annema JT, *et al.* The effectiveness of flexible bronchoscopy simulation-based training: a systematic review. *Chest* 2023; 164: 952–962.
49 Cook DA, Brydges R, Zendejas B, *et al.* Mastery learning for health professionals using technology-enhanced simulation: a systematic review and meta-analysis. *Acad Med* 2013; 88: 1178–1186.
50 Cold KM, Konge L. Response. *Chest* 2024; 165: e61.
51 Huang J, Lin J, Lin Z, *et al.* Artificial intelligence feedback for bronchoscopy training: old wine in a new bottle or true innovation? *Chest* 2024; 165: e60–e61.
52 Lee TD, Carnahan H. When to provide knowledge of results during motor learning: scheduling effects. *Hum Perform* 1990; 3: 87–105.
53 Schmidt RA, Wulf G. Continuous concurrent feedback degrades skill learning: implications for training and simulation. *Hum Factors* 1997; 39: 509–525.
54 Ho E, Hedstrom G, Murgu S. Robotic bronchoscopy in diagnosing lung cancer – the evidence, tips and tricks: a clinical practice review. *Ann Transl Med* 2023; 11: 359.
55 Goldberg V, Manduca A, Ewert DL, *et al.* Improvement in specificity of ultrasonography for diagnosis of breast tumors by means of artificial intelligence. *Med Phys* 1992; 19: 1475–1481.
56 Kim SY, Choi Y, Kim EK, *et al.* Deep learning-based computer-aided diagnosis in screening breast ultrasound to reduce false-positive diagnoses. *Sci Rep* 2021; 11: 395.
57 Shen Y, Shamout FE, Oliver JR, *et al.* Artificial intelligence system reduces false-positive findings in the interpretation of breast ultrasound exams. *Nat Commun* 2021; 12: 5645.
58 Liu XY, Song W, Mao T, *et al.* Application of artificial intelligence in the diagnosis of subepithelial lesions using endoscopic ultrasonography: a systematic review and meta-analysis. *Front Oncol* 2022; 12: 915481.
59 Vilmann P, Clementsen PF, Colella S, *et al.* Combined endobronchial and esophageal endosonography for the diagnosis and staging of lung cancer: European Society of Gastrointestinal Endoscopy (ESGE) Guideline, in cooperation with the European Respiratory Society (ERS) and the European Society of Thoracic Surgeons (ESTS). *Endoscopy* 2015; 47: 545–559.
60 Hassan C, Rizkala T, Mori Y, *et al.* Computer-aided diagnosis for the resect-and-discard strategy for colorectal polyps: a systematic review and meta-analysis. *Lancet Gastroenterol Hepatol* 2024; 9: 1010–1019.
61 Abràmoff MD, Lavin PT, Birch M, *et al.* Pivotal trial of an autonomous AI-based diagnostic system for detection of diabetic retinopathy in primary care offices. *NPJ Digit Med* 2018; 1: 39.
62 American Medical Association. Augmented intelligence in health care – payment and regulation. 2019. www.ama-assn.org/system/files/2019-08/ai-2019-board-report.pdf Date last accessed: 20 January 2025.
63 Kelly BS, Judge C, Bollard SM, *et al.* Radiology artificial intelligence: a systematic review and evaluation of methods (RAISE). *Eur Radiol* 2022; 32: 7998–8007.
64 Chen CC, Lu SC, Chang YK, *et al.* Diagnostic performance of rapid on-site evaluation during bronchoscopy for lung cancer: a comprehensive meta-analysis. *Cancer Cytopathol* 2025; 133: e22908.

65 Tolsgaard MG, Pusic MV, Sebok-Syer SS, *et al.* The fundamentals of artificial intelligence in medical education research: AMEE Guide No. 156. *Med Teach* 2023; 45: 565–573.

66 Dhaliwal J, Walsh CM. Artificial intelligence in pediatric endoscopy: current status and future applications. *Gastrointest Endosc Clin N Am* 2023; 33: 291–308.

67 Miotto R, Wang F, Wang S, *et al.* Deep learning for healthcare: review, opportunities and challenges. *Brief Bioinform* 2017; 19: 1236–1246.

68 Whiting PF, Rutjes AW, Westwood ME, *et al.* QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* 2011; 155: 529–536.

69 Cohen JF, Korevaar DA, Altman DG, *et al.* STARD 2015 guidelines for reporting diagnostic accuracy studies: explanation and elaboration. *BMJ Open* 2016; 6: e012799.