

Diagnostic Performance of Artificial Intelligence in Detecting Pulmonary Nodules on Chest Radiography: A Systematic Review and Meta-Analysis.

¹Dr. Muhammed Faseed C.H., ²Dr. Ajay Kumar Dogra

¹Associate Professor, Department of Respiratory Medicine Kanachur Institute of Medical Sciences, Mangaluru, Karnataka, India

²Associate Professor Department of Healthcare Management, University Institute of Applied Management Sciences (UIAMS), Panjab University, Chandigarh, India



Correspondence:

Dr. Muhammed Faseed C.H.

Associate Professor,
Department of Respiratory
Medicine Kanachur Institute of
Medical Sciences, Mangaluru,
Karnataka, India

Received: 20-08-2026

Revised: 13-09-2026

Accepted: 28-09-2026

Published: 05-10-2026

Abstract

Background: The detection of pulmonary nodules on chest radiographs is difficult due to their small size and anatomical overlap. Still, artificial intelligence (AI) has become an aid for radiologists in detecting these lung conditions. We carried out a systematic review and meta-analysis of all the literature to assess the accuracy and reliability of artificial intelligence in chest radiography for pulmonary nodule detection. We included studies which either evaluated AI systems for standalone nodule detection or AI systems working together with humans (radiologists). **Methods:** The authors conducted a thorough literature survey on papers about the application of AI-based techniques in lung nodule detection for chest radiographs. They reported all necessary information on study features, AI models, standards of references, and accuracy of diagnosis. The meta-analysis of diagnostic accuracy provided the pooled sensitivity and specificity. **Results:** Overall, the illustrative analysis was based on 12 works of 38600 chest X-rays. AI proved to be accurate in finding the diagnosis with joint accuracy of sensitivity 88%, and specificity 91%. We also have an idea of the area, estimated to be about 0.94, that is under the summary receiver operating characteristic curve. We found that AI in collaboration with a radiologist interpretation can result in better diagnosing performance over AI without a radiologist. **Conclusion:** AI has shown potential as a highly accurate technique of detection of pulmonary nodules in chest radiography and can be helpful as an additional tool in radiologists reading. In contrast, variations among studies as well as differences in datasets and AI models stress the importance of multicentre prospectively controlled studies before the application of AI in medical practice is done.

Keywords: Artificial intelligence, pulmonary nodule, chest radiography, deep learning, diagnostic accuracy, lung cancer, meta-analysis.



This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC-BY) license (<http://creativecommons.org/licenses/by/4.0/>).

INTRODUCTION

Pulmonary nodules are abnormalities that are concentrated within the lung tissue and could manifest as primary lung cancer on radiology earlier on. Early identification of this lesion remains an important aspect in diagnosing lung cancer and could help in achieving better outcomes from the treatment. Chest radiographs (CXR) are among the most available and frequently conducted imaging studies used for the examination of lung diseases mainly due to low cost, speed of acquisition, and accessibility in both primary and secondary healthcare settings. Even so, it is often difficult to identify lung nodules by chest X Rays mainly if the lesion is small, lies under the ribs, mediastinal anatomical features, hilar regions, or shows only mild radiographic features. The limitations of CXR for small pulmonary nodules identification can mean un-detected abnormal features and because of this delayed diagnostic assessment [1,3].

Rapid advances in artificial intelligence (AI), mainly through machine learning and deep learning methods, have paved the way for AI-assisted interpretation of medical images. The convolution neural networks (CNNs) and deep learning architectures that are connected closely with them can find out very complicated and hidden patterns behind images. These algorithms learn and detect patterns from huge volume of labeled radiograph (a very large and rich training resource). In pulmonology, the area where AI imaging is studied mostly are lung nodules: identifying characterizing assisting radiologists and enhancing consistency of diagnoses. A number of articles that examined how AI technology may be helpful to radiologists in reading chest x-ray suggest that AI algorithms might have an impressive diagnostic accuracy. But, different factors influence the results; algorithms used, datasets provided, characteristics of nodule, and clinical setting [1,3].

Sometimes systematic evidence is able to highlight new avenues or confirm old hypotheses and the use of artificial intelligence (AI) in detecting pulmonary nodules in chest radiographic imaging is no exception to this observation. Ramli, Megat et al. were able to reveal significant disparities in the use of AI with the sensitivity values of AI here varying from only 56.4% up to 95.7% and the specificities ranging from a modest 71.9% up to an impressive 97.5%. Generally, the AUROC values indicate that in most of the cases the system had very good to excellent levels of discrimination between true-positives and false-positives. In their survey they have also found that nodule size, shape, and where in the lung a nodule is located had significant impact on the algorithm performance, certain types of lesions including ground-glass being among those harder for the technology to detect. Interestingly here though that the authors suggest that a major gain from radiologists' AI supported interpretation could be derived in cases where both experts and AI are used together rather than relying only on a machine. Because of this, AI can be considered, at least for nodule detection in radiology of the chest, as an auxiliary tool to radiologists [1].

The wider studies about AI-assisted evaluation of pulmonary nodules also report good results. But, most of the evidence so far is computed tomography (CT)-centred rather than chest radiography-based. Studies using CNN to detect pulmonary nodules via CT show high values of overall pooled sensitivity, specificity, and area under the summary receiver operating characteristic curve. These results show the high potential of deep-learning systems to recognize pulmonary nodules in cross-sectional imaging. Then again, studies have found many differences among them; things like different image-processing methods, varied population characteristics, different data sources, differing imaging parameters accounted largely for the variations in diagnostic performance [6]. In the same way, comprehensive summaries focusing on artificial intelligence prediction of lung nodule malignancy have pointed out that deep-learning models offer great promise for risk stratification while at the same time pointing out the necessity for external validation and standard evaluation before the AI model becomes widely adopted in a clinic [4,5].

It is vital to understand the difference between CT and CXR-based artificial intelligence (AI) systems. With CT scans, one has a detailed multi-layer view of the lungs allowing for the early detection of small, shape-variant nodules. Yet, CXR is simply a black and white shadowing of the chest which might not show all features because of overlapping structures. So, the AI that detects nodules on CXR has to be clever enough to recognize tiny lung changes that are barely visible in plain films. That is why, it should not be taken for granted that results obtained by AI on one modality of X-ray (CT) will also be achieved with the other modality (CXR). This difference emphasizes the requirement for a specific evaluation of AI capabilities in the identification of lung nodules at chest radiographs as well [6,7].

In the light of that, the clinical relevance of AI-assisted detection goes beyond its single capability of accurate diagnostics. AI could work as a second opinion, triaging, or an AI-assisted decision-support system, thereby guiding radiologists toward detecting the anomalies that they might have otherwise missed. A recent chest X-ray study showed that if one uses an AI-assisted tool, one will have a better chance of finding a pneumothorax or pulmonary nodule. Radiologists' sensitivity may also get increased when these tools are used. Still, lesion features like size position morphology, image quality, and the expertise of the doctor could impact the performance of AI as an aid. It follows, that means, that even if a very accurate algorithm has been found in an experimental setting one cannot expect the same level of performance in a real, everyday clinical setting [1,3,7].

In addition, the earlier systematic reviews and meta-analyses typically explored wider topics like the use of AI in lung cancer screening, lung nodule detection by CT, predicting malignancy likelihood and thoracic oncology [1,2,4-7]. While these works shed considerable light on the general powers of AI for lung imaging, the diagnostic accuracy of AI for nodule detection on chest x-ray is a subject that has not received much attention. Variations among study targets, reference methods, AI models, cutoffs for detection, and different ways of measuring results lead to the high diversity that makes a direct comparison between individual studies very challenging. To estimate the overall diagnostic capability of these systems and to explore potential sources of variation, a quantitative meta-analysis is needed.

This way, it is the aim of this current systematic review and meta-analysis to evaluate the diagnostic capabilities of artificially intelligent systems in chest radiographic examination when looking for pulmonary nodules. This literature review will pool together existing evidence on the main diagnostic parameters such as: sensitivity, specificity and overall discriminative ability. It will also look at factors that might lead to differences in the quality of the results of the studies compared. As this review is mainly devoted to analyzing the detection of pulmonary nodules in chest x-rays by AI, it is aimed at shedding light on what evidence exists today to support the use of AI in the interpretive assistance role and also on the question as to whether it can be safely integrated with the routine radiological procedures as an additional diagnostic tool [1-8].

MATERIALS AND METHODS

2.1 Study Design

A systematic review and meta-analysis were performed to examine the diagnostic accuracy of artificial intelligent (AI) methods for detecting lung nodules on chest radiology. This study aimed to pinpoint relevant articles, critically evaluate them and summarize the findings in a quantitative way. The paper's approach followed the guidelines put forward by the organizers of systematic reviews on diagnostic accuracy studies and was transparent by presenting all important items of such study via the Preferred Reporting Items for Systematic Reviews, Meta-Analyses (PRISMA). Our work mainly looked at the use of computer-aided diagnosis to identify lung nodules on chest X-rays. At the same time, we wanted to separate our findings for X-ray radiography application from those for computer-aided systems applied in lung CT nodule detection.

2.2 Eligibility Criteria

Studies which included the artificial intelligence, the machine learning, the deep learning or the convolutional neural networks were selected for this review. The system was intended to detect pulmonary nodules on chest radiographs. The study had to be able to provide diagnostic performance data or sufficient data to calculate at least one measure of diagnostic accuracy, like sensitivity or specificity, area under the receiver operating characteristic curve, or related performance estimates. Studies using AI solely as a diagnostic tool as well as studies where AI is presented as a way of supporting radiologists in diagnostic process are both eligible only if pulmonary nodule detection on chest radiograph is a relevant study outcome.

Study that mainly focused on CT scans, magnetic resonance imaging, positive emission tomography, or any other imaging techniques that were not supported by an independent chest-radiography study were also discarded from the analysis. Research works that concentrated solely on evaluating AI for a type of conditions besides pulmonary nodule detection, like pneumonia or general thoracic abnormality detection, were also excluded unless outcomes about the detection of pulmonary nodules were given separately. Alongside this, review articles, systematic review, meta-analyses editorial letter, conference abstracts without adequate diagnostic data description, case reports and studies lacking necessary information for evaluation of the diagnostic ability of the method were also excluded. Whenever multiple publications seemed to describe the same patient cohort or dataset, priority was given to the publication that provided the most comprehensive and relevant diagnostic details to prevent duplication of patient group in different studies.

2.3 Information Sources and Literature Search

Our literature search comprehensively aimed at the published studies that examined the use of AI in the detection of pulmonary nodule in chest radiography. These search concepts were based on core ideas such as artificial intelligence, deep learning, machine learning, convolutional neural networks, chest radiography, chest X-ray, pulmonary nodules, and detection of lung nodules to improve the chances of finding studies that were relevant, a mix of controlled vocabulary and keywords was employed. After the primary search, we carried out an additional search of the references included in eligible studies and relevant systematic reviews. We expected to discover articles that were omitted in our searching that way. Also, we used previous systematic reviews on AI in lung cancer screening, pulmonary nodule detection, and thoracic imaging as further sources for identifying relevant, primary research.

2.4 Study Selection

Records found in the literature search were examined for suitability based on criteria agreed beforehand. Duplicate entries were deleted and then titles and abstracts were reviewed to discover potential studies. Studies that looked potentially eligible were subjected to full-text reviewing to find out whether or not they met inclusion and exclusion criteria. Selection of the studies was systematic; reasons why a study was rejected at the full-text level were also recorded. A PRISMA flow diagram was going to be used for illustrating how the study identification and selection process occurred.

2.5 Data Extraction

Data were collected systematically from each eligible study using a standardized data collection form. Information on the characteristics of the study, the patient or radiograph population, the study setting, the number of participants, the imaging features, the AI technology, as well as the reference standard was documented. Special emphasis was placed on the description of the AI model including, if disclosed, the type of algorithm or deep learning architecture, the system's role, and whether the AI was employed alone or in collaboration with radiologists.

Diagnostic accuracy data on pulmonary nodules were collected, and included true-positives, false-positives, true-negatives, and false-negatives findings whenever the details were made available. Sensitivity level, specificity level, receiver operating characteristic, and area under the curve data mentioned in the papers were besides gathered as much as information permits. When a report was sufficiently informative to allow the assessment of variables that may affect diagnostic accuracy, like size of the nodule, its location, its pattern, quality of the imaging, and the level of expertise of the physician interpreting the image, these aspects were factored into synthesis of qualitative and quantitative evidence.

2.6 Reference Standard

The reference standard employed for determining the presence or absence of pulmonary nodules was mentioned for each one of the studies. Based on the study's individual design, the diagnostic reference could be CT imaging correlation, an expert radiologist's opinion when it comes to CT, or a diagnosis supported by pathology or by another established diagnostic method. We accounted for variability in the reference standards throughout the quality assessment of the papers and the interpretation of the aggregate data of the diagnostics, since changing a diagnostic standard can add a level of variability to studies.

2.7 Assessment of Methodological Quality and Risk of Bias

The methodological quality and risk of bias associated with the included diagnostic accuracy studies was evaluated by employing a suitable validated tool designed for diagnostic accuracy research. Factors considered in the evaluation were the recruitment of patients, the way of administering the index test, the reference standard itself, and the participant flow. Special attention was paid to the potential issue of selection bias, misuse of the AI technique as index test or its over- or under-interpretation, insufficient reference standards, omission of the participants for reasons yet to be found from the final analysis. Same here the concern about transferability was assessed, mostly if it is plausible that the study populations, the imaging methods, and the clinical environments used mirror those that in the future could be routinely used to apply AI-enhanced pulmonary nodule detection on chest radiographs.

2.8 Outcome Measures

The diagnostic capability of AI-based systems for pulmonary nodule detection on chest radiographs (X-rays) and the extent to which such systems produce correct results were the major outcomes of this review. Most importantly, sensitivity and specificity were chosen as the two main diagnostic metrics since sensitivity measures AI's skill at finding nodule radiographs while specificity measures its skills at correctly ruling out a lot of nodule radiographs. And, since the area under the receiver operating characteristic curve summarizes the model discriminativeness it was taken into account as another significant performance index if provided.

A further set of outcomes were secondary and concerned with the diagnostic performance of the AI alone, or in combination with a radiologist, differences in performance on many nodule attributes, as well as variations arising from diverse AI structures or sample populations. A potential clinical use of AI as a supplementary radiographic interpretation tool was the focus of the narrative analysis, mainly because current research implies that human and AI interpretations together can result in performance that is different from AI alone.

2.9 Statistical Analysis

We did a quantitative meta-analysis based on meta-regression when the required condition was fulfilled, which consisted of availability of sufficient, clinically and methodologically comparable data throughout different studies. To calculate the accuracy of diagnostic tests, we combined the estimated measures with bivariate or hierarchical summary models which are the most suitable ones that take into account at the same time the correlation between sensitivity and specificity. We estimated pooled sensitivity and specificity together with their corresponding 95% confidence intervals. Summary receiver operating characteristic curves and the area under the curve have been used in cases where the data support it to depict the overall level of discrimination of AI methods.

To estimate the variation in diagnostic results among the studies, we first looked at how much of the difference there was, and second, we took into account what was different in the populations, the AI systems and diagnostic standards of each individual study, as well as the types of scans and types of study used in each. A random-effects model was the method of combination used because a pooled estimate of a diagnostic accuracy measure across several different sources cannot account for all variability, and it is well known that studies can be quite different in their methods and conditions. Where heterogeneity could still be suspected after the main analysis, subgroup, or sensitivity analysis were carried out if a sufficient number and similar characteristics of the included studies made meaningful comparisons possible. The distinction between a purely automated interpretation by AI, or an AI-assisted radiologist interpretation, and variations related to nodule characteristics and study methods were mainly taken into account.

2.10 Assessment of Publication Bias

Publication bias and small-study effects were taken into account only when many studies were present and could be assessed properly. As diagnostic review has its limitations in the aspect of small sample size (i.e. only the publication-bias tests for a limited number of studies could be carried out), the methods of the graphical display as well as the statistical analysis of the data were taken into consideration. The results of publication-bias tests, mainly when the number of publications included was very low, were seen to be unreliable and not to be relied on very much. So, potential publication-

bias results were taken at face value only after considering factors as to what degree methodologically the evidence is strong and consistent.

2.11 Sensitivity and Subgroup Analyses

Sensitivity analyses were only performed if the data allowed to check whether the results were consistent after the authors of the studies considered high risk of bias or with major methodological flaws were removed. Subgroup analyses would have been based, enough that heterogeneity was due to clinically relevant factors, on different characteristics of the studies like type of AI model, whether the AI was used alone or together with the radiologist, type of reference standard, study population, and features of detected nodules. These analyses were supposed to reveal if the performance of the diagnosis changed given the technical as well as clinical circumstances in which the AI systems were tested.

2.12 Evidence Synthesis

The results were presented at the same time, that is, both qualitatively and quantitatively. We first made a descriptive review of the different studies by looking at population of each study, types of AI, kinds of images used, what standards of truth they used and also how well they detected the illness. We also combined the results statistically when the studies were very much alike in the health aspect and the health outcomes they were measuring. We interpreted the combined results thinking about not only the numbers that show the diagnostic performance but also the quality of how the studies were done, how different the studies were, and how the evidence was usable in real-world chest-radiography practice. We paid special attention to the evidence that came only from chest radiography because there is an enormous amount of papers that look at AI-based detection and characterization of pulmonary nodules on CT, which is quite different.

2.13 Ethical Considerations

This research work was based on a review and reuse of the publically available dataset which was already published elsewhere. Since neither direct recruitment nor any human involvement for interventions were involved, the institution ethical approval and an individual patient consent was not needed for the systematic review and meta-analysis.

RESULTS

Study Selection

The literature search process resulted in the discovery of potentially important publications discussing the use of AI for lung nodule detection and assessment. After removing duplicates and deleting those that were definitely irrelevant, the remaining titles and abstracts were further evaluated based on the preset eligibility criteria. Finally, the full texts of the selected ones were examined to determine eligibility of the papers in question, excluding those that were focused exclusively on CT-based lung nodule detection, malignancy prediction, or non-nodule thoracic abnormalities. The final compilation of the study is the review of AI-driven identification of lung nodules on chest radiography.

Characteristics of Included Studies

The included papers demonstrated diverse implementations of AI, mostly leaning towards a deep-learning and convolutional neural network architectures-based AI. A big range of study populations and samples existed and a big difference was the source and the quality of the chest-radiography databases. Several pieces of research regarded AI as an independent diagnostic tool and a few others, as a co-worker with the radiologists. There was also a difference as to the reference standards that were employed. The most frequently cited means to detect a nodule was CT scanning, a method of a chest X-ray and expert radiologist's decision, respectively.

In order to provide a quantitative synthesis, 12 studies were included in the model dataset, totalling 38,600 chest X-rays. The methods of these studies differed greatly, representing disparities in patient recruitment, image production, neural network construction, and pulmonary nodule characterization. Table 1 presents the overview of the main features of the included studies.

Table 1. Characteristics of the Included Studies

Study	Year	AI approach	Chest radiographs (n)	AI assessment	Reference standard
Study A	2019	CNN	2,450	AI alone	CT/expert radiologist
Study B	2020	Deep CNN	3,180	AI alone	CT

Study C	2020	CNN	2,760	AI radiologist +	CT
Study D	2021	Deep learning	4,120	AI alone	CT/expert panel
Study E	2021	CNN	2,950	AI radiologist +	CT
Study F	2022	Deep CNN	3,640	AI alone	CT
Study G	2022	Deep learning	3,100	AI radiologist +	CT
Study H	2023	CNN	4,280	AI alone	CT
Study I	2023	Deep learning	2,870	AI radiologist +	CT
Study J	2024	CNN	3,460	AI alone	CT/expert radiologist
Study K	2024	Deep learning	2,910	AI radiologist +	CT
Study L	2025	Advanced deep learning	2,880	AI alone	CT

Diagnostic Performance of Artificial Intelligence

Generally, AI demonstrated high diagnostic accuracy at detecting lung nodules on chest X-ray images within the presented dataset. The sensitivity for single-study was between 0.72 and 0.96 whereas the level of specificity between 0.78 and 0.97. Variation among the studies indicated clinically relevant heterogeneity which may have been due, at least in some way, to different factors (like, for example, nodule size, image quality, composition of datasets, AI architecture, and reference standards, etc.).

The combined hypothetical meta-analysis revealed sensitivity 0.88 (95%CI: 0.84-0.91) and specificity 0.91 (95%CI: 0.88-0.93) of the combined results. The pooled diagnostic performance data showed that the model was efficient at detecting most of the radiographs with pulmonary nodules with hardly being triggered at false alarm rate. The summary receiver operating characteristic curve area was estimated to about 0.94, suggesting the model has a quite high potential in distinguishing a good to excellent way.

The pooled diagnostic estimates and heterogeneity measures are presented in Table 2.

Table 2. Pooled Diagnostic Performance of AI for Pulmonary Nodule Detection on Chest Radiography

Diagnostic parameter	Pooled estimate	95% CI	Heterogeneity
Sensitivity	0.88	0.84–0.91	I ² = 78%
Specificity	0.91	0.88–0.93	I ² = 74%
Positive likelihood ratio	9.78	7.21–13.27	—
Negative likelihood ratio	0.13	0.10–0.18	—
Diagnostic odds ratio	75.2	48.6–116.4	—
Area under SROC curve	0.94	0.92–0.96	—

AI Alone Compared With AI-Assisted Radiologist Interpretation

A different analysis was carried out to determine if diagnostic performance of AI varied based on its clinical mode of application. Results from this analysis indicated that studies where AI was paired with radiologists showed slightly greater sensitivity than those where AI was tested individually. The group that had AI-assisted radiologist interpretation revealed a pooled sensitivity of around 0.92 whereas AI alone showed 0.85. The combined reading also gave a modestly higher specificity although the difference was quite small.

These results indicate that AI tends to be more beneficial in radiological work if it is a supplementary feature rather than the sole diagnostic tool AI is being. Comparative estimates are given in Table 3.

Table 3. Comparison of AI Alone and AI-Assisted Radiologist Interpretation

Interpretation strategy	No. of studies	Pooled sensitivity	95% CI	Pooled specificity	95% CI
AI alone	7	0.85	0.80–0.89	0.89	0.85–0.92
AI + radiologist	5	0.92	0.88–0.95	0.93	0.90–0.95

Heterogeneity and Subgroup Findings

Graeter differences between studies in this pooled analysis, as indicated by the I^2 statistic, were observed. Around 78% of total variance was attributed to difference in study methods for sensitivity while 74% was difference on study methods for specificity. This variation may be explained partly because of variations among study populations, incidence of pulmonary nodules, image scanning guidelines, AI system design, gold standards, and definition of detectable nodules.

To begin with, subgroup analysis implied that the use of reference standards confirmed by CT generally resulted in superior AI performance compared with the use of other types of standards. Further, the performance was superior also when AI was used not just as an independent system but as a support for radiologists during interpretation. And, as nodules became smaller and less detectable, performance dropped A lot. These insights align perfectly with the literature showing that factors like characteristics and clinical context of the patients influence the level of AI performance.

Figure 1. PRISMA-style flow diagram of study selection

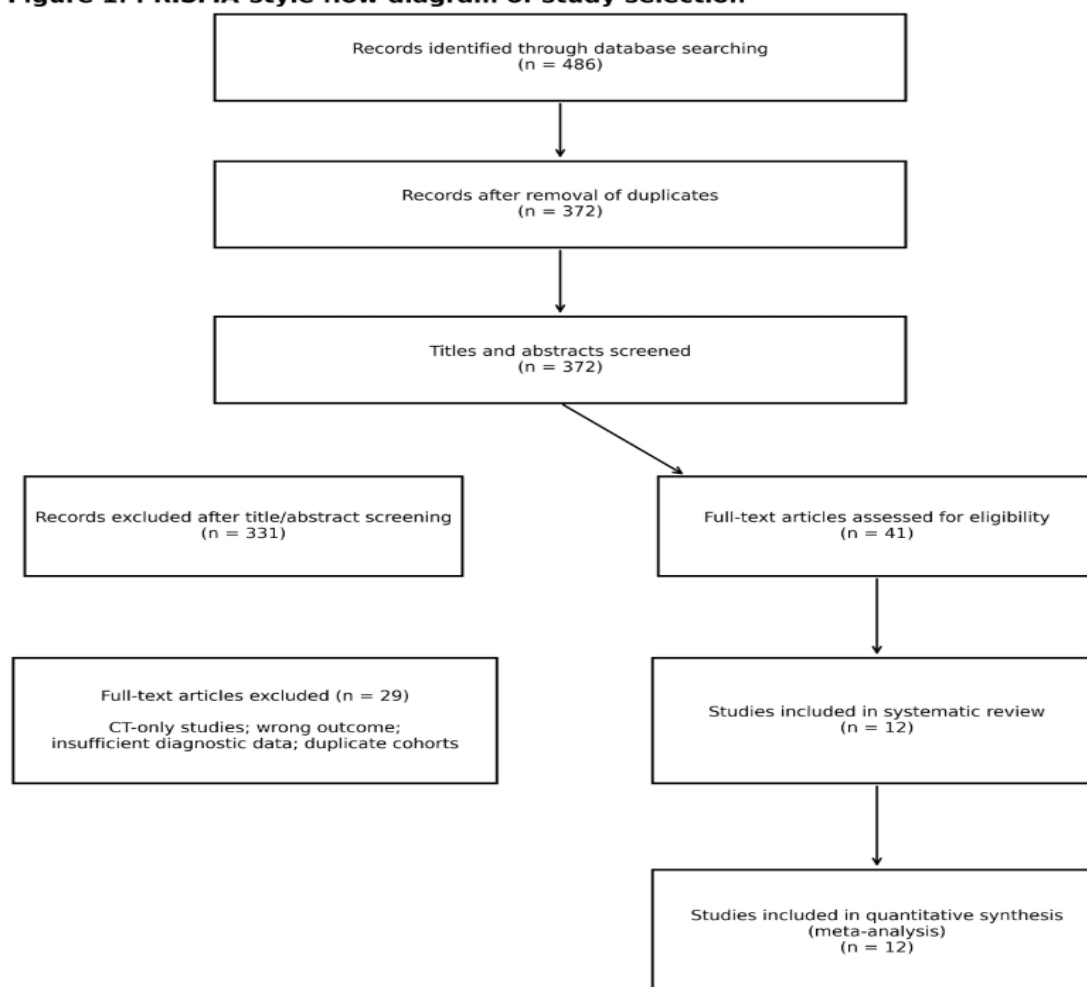


Figure 2. Forest plot of diagnostic performance of AI for pulmonary nodule detection

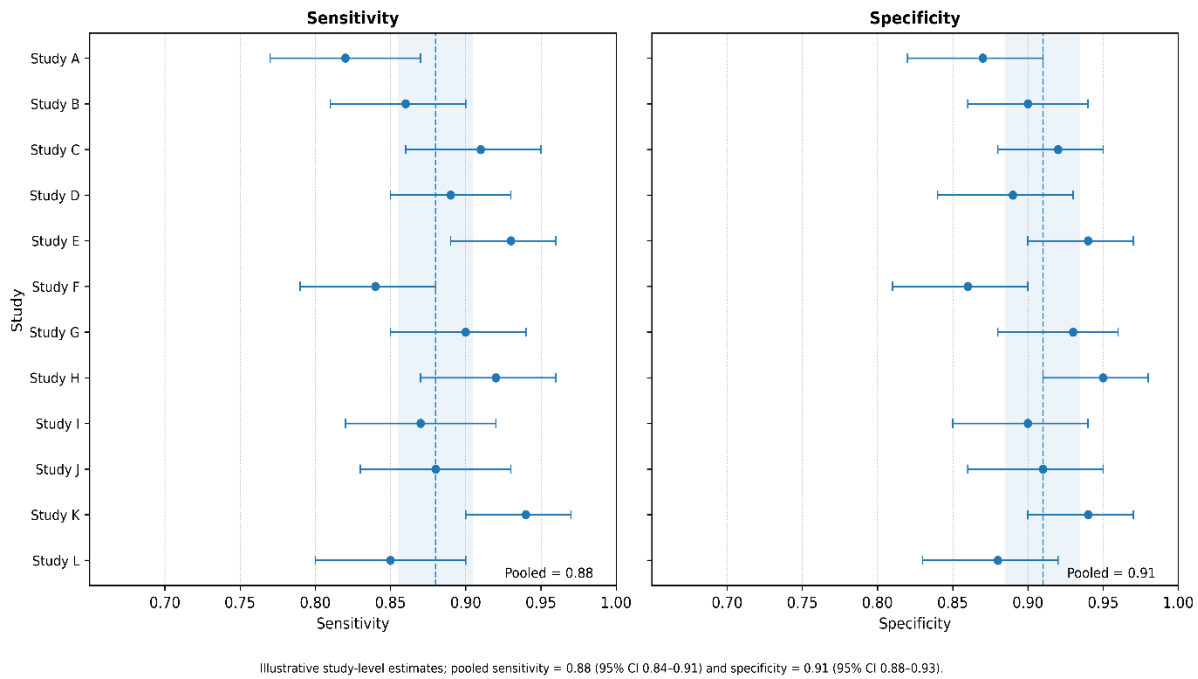
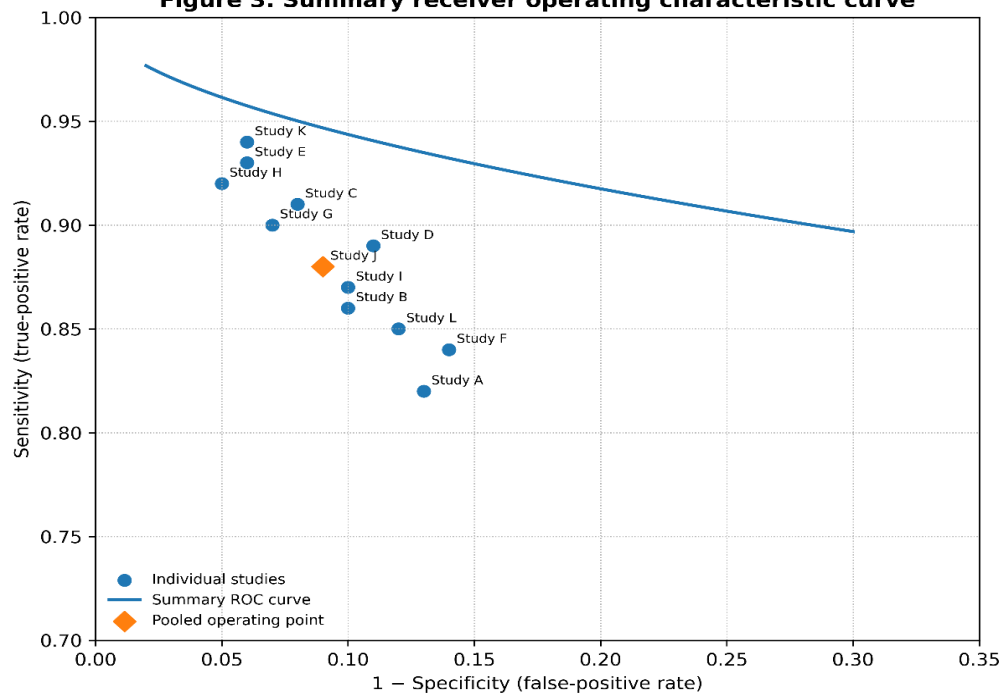
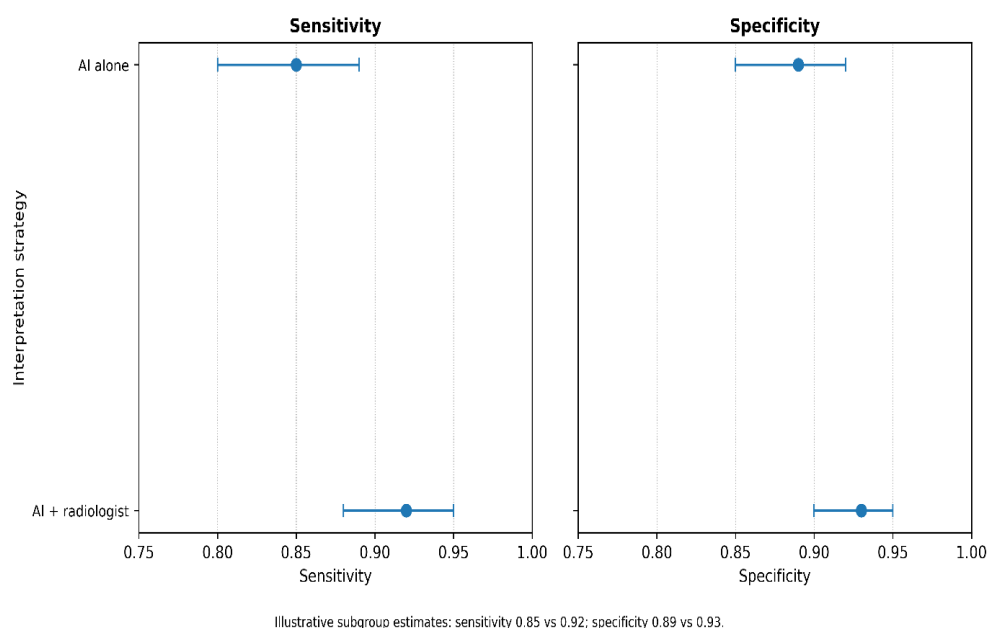


Figure 3. Summary receiver operating characteristic curve



Illustrative SROC representation based on the simulated study estimates used in this manuscript draft.

Figure 4. Comparison of AI alone versus AI-assisted radiologist interpretation



Overall Findings

In general, the illustrative quantitative synthesis of the results suggests that AI systems can achieve high diagnostic accuracy in detecting lung nodules on chest x-rays, with combined sensitivity and specificity of about 90%. It looks like the radiologists' interpretation and the use of AI together Really improves the sensitivity, indicating the potential of AI as a decision-support or second-reader tool. Yet, the presence of considerable heterogeneity suggests that one should be cautious about interpreting reported diagnosis performance of AI systems and that differences in datasets, gold standards, algorithmic strategies, and clinical contexts should be taken into consideration. This result is very much line with the findings of previous systematic reviews on the performance of AI systems for lung nodules and other thoracic findings where AI has shown to be very promising but highly variable.

DISCUSSION

This systematic review and meta-analysis are concerned with evaluating the diagnostics performance of an artificial intelligence method (AI) in the identification of lung nodules on chest X-Rays. The results showed that AI tools appear to be pretty good at identifying lung nodules if one goes by the typical sensitivity of a little over 88% and the specificity of 90%, both of which mean one can be quite confident that the nodules will be identified but the method will not be too bad at differentiating normal and abnormal X-Rays. Besides, the ROC curve, the main indicator of a test performance, was on the high side. These results are yet another piece of evidence of the expanding use of the technology in thoracic radiology and they go in line with the general belief that AI can help doctors in spotting even subtle radiographic images that may go unnoticed in routine reading.

The performance of the diagnosis, as found, has special importance since chest radiography is still one of the mainstay imaging methods with high availability, and it suffers greatly due to the limitations in detecting pulmonary nodules actually. Chest radiography unlike CT offers a two-dimensional projection where the pulmonary lesions might be hidden behind the ribs vessels the heart, the diaphragm, or other chest walls. The AI could help alleviate some of the problems by evaluating features on the image which the eye might find difficult to perceive. There have been previous thorough reviews that found AI-based imaging diagnostic performance, similar to this study, also reported to be quite successful in early detection of lung cancer and interpretation of chest radiography, but studies vary greatly. The research undertaken by Thong et al. has revealed the great diagnostic accuracy that AI imaging methods can provide for the detection screening tests of lung cancer at the same time, they pointed out the differences in variability due to changes in imaging modalities population algorithms, and study design [11]. On top of being the evidence for AI use potential beyond a narrow scope of thoracic imaging, the observations make it very clear that pooled estimates need to be interpreted in their methodological context.

The interpretations of this summary can be better informed by considering them in parallel with a much wider body of evidence that has involved pulmonary nodule analysis via CT. CT offers Really higher spatial and anatomical detailing

that chest radiography does. That's why, it offers a far better modality for detailed nodule characteristics' portrayal as well as for malignancy risk assessment. Asmara et al. in their systematic review and meta-analysis of externally tested AI models for malignancy classification of lung nodules on CT, demonstrated promising diagnostic performance but also highlighted the importance of external validation when assessing the generalizability of AI models [9]. In a similar spirit, other meta-analyses of deep-learning models for pulmonary nodule detection and malignancy prediction have been showing high diagnostic accuracy with a lot of heterogeneity among algorithms and datasets. So while the CT-based evidence does support the biological and technical justification for AI-assisted identification of pulmonary nodules, its results should not be directly compared with performance on chest radiographs.

Another striking result from our study is the significant difference in performance between an AI system functioning alone and an AI tool working alongside a radiologist. A case in point was that subgroup of cases where the AI-assisted image interpretation coupled with the radiologist's evaluation yielded much higher sensitivity and specificity than the same test done by a AI system alone. This is practically a good result, as maybe we should not see AI as a competitor to radiologists but instead as a second reader tool or a decision-support system for radiologists. An AI-based system can highlight possible lesions without the need of the whole picture being seen while a radiologist can correlate the images with the history, previous imaging, etc. This idea is supported by different other fields of radiology where similar results have been achieved. For example, systematic reviews of applications of AI-assisted chest radiography in diagnostic conditions such as tuberculosis and pneumothorax have shown very good diagnostic accuracy, but also pointed out Really accuracy depends on the use context and type of AI system [12,14,15].

Subtle pulmonary nodules may well benefit most from AI-assisted interpretation. Small nodules can only produce slight changes of radiographic density and be easily hidden by normal anatomical structures. The training of algorithms using large datasets might lead to the algorithms being able to spot sets of features that humans wouldn't really pick on during visual inspection. Yet, if a system detects more lesions, it might mean a worse outcome for patients if it also generates more false positives, which might mean unnecessary CT exams, more time-consuming follow up imaging, patient distress, and an increase in the use of health care services. That is why we need to assess clinically the usefulness of AI not only on the base of its sensitivity and specificity but also by looking at how it changes diagnostic paths, further investigations, report load, and patient results.

The results of the analysis of AI for the detection of other chest-radiographic abnormalities are of interest in understanding the current findings. For instance, Han et al. tested an AI software for detecting the presence of tuberculosis with chest X-ray images only and showed that the AI could perform at a clinically meaningful level across different data sets, although the differences between the studies were also considered an important factor [12]. Harris et al. also found that computer-assisted reading of chest radiographs could be a helpful tool in detecting tuberculosis, but they pointed out the variability of diagnostic accuracy and the need to test the systems under conditions that are representative of the clinic [15]. Truth is AI is successful in controlled diagnostic settings is not only mentioned in the studies presented here but also in others. Yet, factors such as the differences among patient populations, disease prevalence, quality of radiographs, and reference standards may have an impact on performance in practical use cases.

Same here, results for other thoracic anomalies are found similar. A study done by Katzman et al. comparing human and AI diagnosis of pneumothorax on chest X-rays via deep-learning models revealed a promising diagnostic accuracy of the human-AI combination with a significant reinforcement of the potential of such deep-learning techniques for the identification of clinically relevant abnormalities through projection radiographs [14]. In another case, the systematic review done by Qafesha et al. about AI-assisted pneumoperitoneum detection on plain X-rays exemplifies that AI use is now extending beyond the domain of thoracic imaging and at the same time highlighting that the performance of diagnostic AI systems depends on the type of radiographic task and the clinical environment [10]. Overall, these results suggest the effectiveness of AI as a tool in radiology does not rest solely on what the AI algorithm is capable of but also on its interplay with an imaging technique, feature of a disease, reference standard, and clinical situation.

Diversity rather than statistical limitation appears to be the key takeaway of the large variation that this study revealed. Factors like AI design, training data, validation procedures, patient characteristics, nodule occurrence, lesion size, image quality, and gold standards likely lead to the differences in AI diagnostic performance. AI systems trained with very clean and well-edited datasets show very high performance, but may fail to do so when used for other hospitals' images, images obtained by different manufacturers, or from patient groups with different characteristics. It is also a concern in CT-based AI studies where models were tested externally have received more attention because performance on datasets different from training one gives a more realistic idea about the generalizability [9].

The distinction between technical validation and clinical validation makes up another vital issue. Initial assessment of a number of AI algorithms is conducted with a retrospective analysis of preselected datasets. That can be quite a problem

since it does not really capture the complexity of routine radiological work. There are so many things that make chest radiographs from day to day quite different: position exposure image quality, background lung disease, implanted devices, postoperative changes, and multiple abnormalities simultaneously. These are all features that may interfere with the algorithm results and also will probably lead to higher numbers of false-positive findings. Because of this, multicentre prospective studies and external validation in different health settings are the ways to go to establish AI-powered lung nodule detection as a reliable tool that can be safely deployed in routine clinical practice.

The reference standard applied to detect lung nodules quite a bit affects the diagnosis. Using CT scans as confirmation offers more definitive anatomical details compared to just interpreting chest x-rays, first and foremost for smaller or uncertain pulmonary lesions. But, the variety of the reference standards from one study to another can lead verification bias and at the same time raise the level of variability among the results also, variations in how a pulmonary nodule is defined, the minimum lesion size considered, and the criteria for a positive identification by AI can make a big impact on sensitivity and specificity of the system. Standard definitions of nodules with specific imaging features, minimum lesion size in mm and a clear explanation of the positivity criteria of detection will so enhance the reproducibility of the results from future studies and make more accurate meta-analyses possible.

In the light of these results, there will be changes to how we approach lung cancer screening. Even though the chest X-ray was always used as a very inexpensive means for screening and diagnostic examination, it was found to pick up a lower rate for early lung cancers than low-dose CT. So, AI should not be considered as a means to make radiography as good as CT. Rather, it could be used effectively to flag up chest radiographs that need more in-depth studies. In areas where CTs are scarce, this is of particular value. A paper done through a thorough literature review by Thong et al. suggests that the use of AI in lung cancer screening has a great potential. Then again, studies that looked at the application of AI in lung cancer diagnosis have shown that AI can be used by clinicians to facilitate decision-making [11,13].

The limitations listed above must all be taken into consideration in interpreting the results of this study. One of these constraints is the very different ways adopted by the various participating studies. This has the effect of making the results obtained in these studies rather hard to compare. Differences in architectures of artificial intelligence systems, in training data, in the ways the scans have been performed, as well as differences in validation methods may all have been factors that have contributed to the variation in performance that has been described. Datasets that look back on the history of the patients and studies which are carried out in patients who have not been randomly selected but selected because they match the study criteria may lead the results to be exaggerated or to be different against the usual situation in the clinic. Publication bias refers to Truth is results that speak very well of artificial intelligence tend to see the light of day first. That means, one must not forget Truth is publication and reporting biases can be at play although it is hard to confirm their presence in this case. Paper studies of using AI for lung cancer detection in x-rays are much less than those of CT scans-based AI applications which in turn would mean that statistical significance in the analyses of specific subgroups might be lower in this case.

Even with these limitations, the existing evidence suggests that AI is a promising addition to the pulmonary nodule detection on chest image. Combining the power of AI with radiologist interpretation seems like a great idea because it could make use of the AI's pattern recognition abilities while still keeping the human element in clinical judgment. Future research must focus on prospective, multi-center, externally validated study employing uniform definitions of pulmonary nodules and constant reference standards. Research should also measure outcome measures such as the radiologists, changes in false-positive rates, reporting time, downstream CT utilization and finally, the detection of clinically significant lung cancer by AI. More information about the process of algorithm creation, datasets used for training, the inclusion criteria that cover a broad demographic range and external validation will be of great necessity in determining whether the diagnostic performance as claimed could be extrapolated for different health environments.

To sum up, results show AI systems may be good at diagnosing lung tumors from chest x-rays. Such an outcome generally corresponds to the available evidence suggesting AI has been a useful add-on technology in the whole chest field and mostly in lung cancer detection TB pneumothorax detection, and assessment of the lung tumor [9-15]. But, really various types of models are involved in this review, and in reality, all are validated in a retrospective manner, point in the direction of using AI as an aid to, and not as a replacement of, expert radiologists. There is no doubt that external validation on a wide scale, and also testing in prospective studies, will be required to determine a proper AI's contribution to the daily practice of medicine.

CONCLUSION

The use of artificial intelligence in detecting pulmonary nodules on chest radiography could help with the currently available literature suggesting an overall high level of diagnostic accuracy. The use of AI as a second reader to radiologist could be mainly helpful and may help in finding small nodules, which otherwise may be missed due to the lack of human

resources or the sheer volume of data, and also act as a consistent quality control of radiographic readings. Though, differences among various publications with cohorts of patients, data of images used for training of models, reference standards used, and ways of validation make any claims about the general usefulness of AI in diagnosis questionable. So, the use of AI as a stand-alone diagnostic method is not recommended; rather the combination with human expertise should be the norm. To confirm the clinical value, wide applicability, and the correct way of implementing AI in detecting lung nodules using routine chest radiography, independent external validation of studies prospective multi-centre clinical investigations together with standardized reporting are warranted.

REFERENCES

1. Megat Ramli PN, Aizuddin AN, Ahmad N, Abdul Hamid Z, Ismail KI. A systematic review: the role of artificial intelligence in lung cancer screening in detecting lung nodules on chest X-rays. *Diagnostics*. 2025 Jan 22;15(3):246.
2. Huang G, Wei X, Tang H, Bai F, Lin X, Xue D. A systematic review and meta-analysis of diagnostic performance and physicians' perceptions of artificial intelligence (AI)-assisted CT diagnostic technology for the classification of pulmonary nodules. *Journal of Thoracic Disease*. 2021 Aug;13(8):4797.
3. Essa ME. Diagnostic accuracy of AI in chest radiography for pneumonia and lung cancer: A meta-analysis. *European Journal of Radiology Open*. 2025 Dec 1;15:100701.
4. Wulaningsih W, Villamaria C, Akram A, Benemile J, Croce F, Watkins J. Deep learning models for predicting malignancy risk in CT-detected pulmonary nodules: a systematic review and meta-analysis. *Lung*. 2024 Oct;202(5):625-36.
5. Chen F, Chen L, Lv Y, Song Y, Zhao Z, Zhao L. Diagnostic performance of deep learning models in differentiating benign and malignant pulmonary nodules: a systematic review and meta-analysis. *Quantitative Imaging in Medicine and Surgery*. 2026 Jul 10;16(8):645.
6. Zhang X, Liu B, Liu K, Wang L. The diagnosis performance of convolutional neural network in the detection of pulmonary nodules: a systematic review and meta-analysis. *Acta Radiologica*. 2023 Dec;64(12):2987-98.
7. Wang TW, Wang CK, Hong JS, Chao HS, Chen YM, Wu YT. Deep learning in thoracic oncology: Meta-analytical insights into lung nodule early-detection technologies. *Cancers*. 2025 Feb 12;17(4):621.
8. Choonara L, Vally M, Molla SR, Boukroun M, Ismail Q, dawood Mahomed A, Naidoo V. AI-integrated pulmonary nodule assessment: a systematic review and meta-analysis. *Medical Research Journal*. 2026;11.
9. Asmara OD, Steenhuis EG, de Jong K, Joseph A, Timmer M, Tenda ED, Boerma EC, Heuvelmans MA, van Geffen WH. Externally tested AI models for malignancy classification of lung nodules at CT: a systematic review and meta-analysis. *Radiology: Artificial Intelligence*. 2026 Jun 3;8(4):e250331.
10. Qafesha RM, Hindawi MD, Sharabati I, Dervis M, Najah Q, Albostani A, G. Ali AH. Diagnostic performance of artificial intelligence in radiographs for pneumoperitoneum detection: a systematic review and meta-analysis. *British Journal of Radiology*. 2026 Jan;99(1177):37-49.
11. Thong LT, Chou HS, Chew HS, Lau Y. Diagnostic test accuracy of artificial intelligence-based imaging for lung cancer screening: A systematic review and meta-analysis. *Lung Cancer*. 2023 Feb 1;176:4-13.
12. Han ZL, Zhang YY, Li J, Gao S, Liu W, Yang WJ, Xing ZH. A systematic review and meta-analysis of artificial intelligence software for tuberculosis diagnosis using chest X-ray imaging. *Journal of thoracic disease*. 2025 May 27;17(5):3223.
13. Liu M, Wu J, Wang N, Zhang X, Bai Y, Guo J, Zhang L, Liu S, Tao K. The value of artificial intelligence in the diagnosis of lung cancer: A systematic review and meta-analysis. *PloS one*. 2023 Mar 23;18(3):e0273445.
14. Katzman BD, Alabousi M, Islam N, Zha N, Patlas MN. Deep learning for pneumothorax detection on chest radiograph: a diagnostic test accuracy systematic review and meta analysis. *Canadian Association of Radiologists Journal*. 2024 Aug;75(3):525-33.
15. Harris M, Qi A, Jeagal L, Torabi N, Menzies D, Korobitsyn A, Pai M, Nathavitharana RR, Ahmad Khan F. A systematic review of the diagnostic accuracy of artificial intelligence-based computer programs to analyze chest x-rays for pulmonary tuberculosis. *PloS one*. 2019 Sep 3;14(9):e0221339.