cancers

MDPI

Systematic Review

AI-Driven Models for Diagnosing and Predicting Outcomes in Lung Cancer: A Systematic Review and Meta-Analysis

Mohammed Kanan [1,*], Hajar Alharbi [2], Nawaf Alotaibi [3], Lubna Almasuood [4], Shahad Aljoaid [5], Tuqa Alharbi [6], Leen Albraik [7], Wojod Alothman [8], Hadeel Aljohani [9], Aghnar Alzahrani [10], Sadeem Alqahtani [11], Razan Kalantan [9], Raghad Althomali [12], Maram Alameen [12] and Ahdab Mufti [13]

1 Department of Clinical Pharmacy, King Fahad Medical City, Riyadh 12211, Saudi Arabia
2 Department of Medicine, Gdansk Medical University, 80210 Gdansk, Poland
3 Department of Clinical Pharmacy, Northern Border University, Rafha 73213, Saudi Arabia
4 Department of Pharmacy, Qassim University, Buraydah 52571, Saudi Arabia
5 Department of Medicine, University of Tabuk, Tabuk 47911, Saudi Arabia
6 Department of Medicine, Qassim University, Buraydah 52571, Saudi Arabia
7 Department of Medicine, Al-Faisal University, Riyadh 12385, Saudi Arabia; lsalbraik@alfaisal.edu
8 Department of Medicine, Imam Abdulrahman Bin Faisal University, Dammam 31411, Saudi Arabia
9 Department of Medicine and Surgery, King Abdulaziz University, Jeddah 22230, Saudi Arabia
10 Department of Medicine, Al-Baha University, Al Bahah 65964, Saudi Arabia
11 Department of Pharmacy, King Khalid University, Abha 62217, Saudi Arabia
12 Department of Medicine, Taif University, Taif 26311, Saudi Arabia
13 Department of Medicine, Ibn Sina National College, Jeddah 22230, Saudi Arabia
* Correspondence: mkaalshammari@kfmc.med.sa

Simple Summary: This research explores the transformative potential of artificial intelligence (AI) in the early detection of lung cancer. Through a comprehensive systematic review and meta-analysis, this study evaluates the effectiveness of AI models, emphasizing a promising avenue for improving diagnostic accuracy. Among 1024 identified records, 39 studies were meticulously selected and analyzed following the PRISMA guidelines. The findings highlight significant strides in AI's role, emphasizing the need for standardized protocols. Despite the observed heterogeneity, this study underscores AI's promising impact on lung cancer screening, laying the groundwork for future advancements in clinical practice. This research contributes crucial insights for healthcare professionals and researchers alike, aiming to enhance the early diagnosis and management of lung cancer.

Abstract: (1) Background: Lung cancer's high mortality due to late diagnosis highlights a need for early detection strategies. Artificial intelligence (AI) in healthcare, particularly for lung cancer, offers promise by analyzing medical data for early identification and personalized treatment. This systematic review evaluates AI's performance in early lung cancer detection, analyzing its techniques, strengths, limitations, and comparative edge over traditional methods. (2) Methods: This systematic review and meta-analysis followed the PRISMA guidelines rigorously, outlining a comprehensive protocol and employing tailored search strategies across diverse databases. Two reviewers independently screened studies based on predefined criteria, ensuring the selection of high-quality data relevant to AI's role in lung cancer detection. The extraction of key study details and performance metrics, followed by quality assessment, facilitated a robust analysis using R software (Version 4.3.0). The process, depicted via a PRISMA flow diagram, allowed for the meticulous evaluation and synthesis of the findings in this review. (3) Results: From 1024 records, 39 studies met the inclusion criteria, showcasing diverse AI model applications for lung cancer detection, emphasizing varying strengths among the studies. These findings underscore AI's potential for early lung cancer diagnosis but highlight the need for standardization amidst study variations. The results demonstrate promising pooled sensitivity and specificity of 0.87, signifying AI's accuracy in identifying true positives and negatives, despite the observed heterogeneity attributed to diverse study parameters. (4) Conclusions: AI demonstrates promise in early lung cancer detection, showing high accuracy levels in this systematic review. However, study variations underline the need for standardized protocols to fully leverage AI's potential in revolutionizing early diagnosis, ultimately benefiting

check for updates

Citation: Kanan, M.; Alharbi, H.; Alotaibi, N.; Almasuood, L.; Aljoaid, S.; Alharbi, T.; Albraik, L.; Alothman, W.; Aljohani, H.; Alzahrani, A.; et al. AI-Driven Models for Diagnosing and Predicting Outcomes in Lung Cancer: A Systematic Review and Meta-Analysis. *Cancers* **2024**, *16*, 674. https://doi.org/10.3390/cancers16030674

Academic Editors: Athanasia Pataka, Serafeim Chrysovalantis Kotoulas and Dionisios G. Spyratos

Received: 19 November 2023
Revised: 20 January 2024
Accepted: 25 January 2024
Published: 5 February 2024

Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

patients and healthcare professionals. As the field progresses, validated AI models from large-scale perspective studies will greatly benefit clinical practice and patient care in the future.

Keywords: AI-driven models; diagnosing; predicting; outcomes; lung cancer; systematic review; meta-analysis

1. Introduction

Lung cancer remains a formidable global health challenge, claiming the lives of millions of individuals each year [1]. The high mortality rate associated with lung cancer is primarily attributed to the advanced stage at which it is often diagnosed [2]. Lung cancer is notorious for its asymptomatic early stages, making it extremely difficult to diagnose until it has reached an advanced, often incurable, stage. The later the diagnosis, the more limited the treatment options, and the grimmer the prognosis for patients. In contrast, when lung cancer is detected at an early stage, the chances of successful treatment and long-term survival increase significantly. Consequently, there is a pressing need for innovative strategies to enable early detection, as this could significantly improve the prognosis and overall survival rates of lung cancer patients [3]. In recent years, the field of artificial intelligence (AI) has emerged as a promising avenue for achieving this goal. In the field of healthcare, AI has shown promise in improving diagnostic accuracy, predicting disease outcomes, and personalizing treatment plans [4]. In the context of lung cancer, AI systems can analyze vast datasets of medical images, patient records, and genetic information to identify patterns and abnormalities that may elude human perception. These systems can not only detect lung cancer at earlier stages, but also assist in risk assessment and treatment planning [5].

Current methods for early lung cancer detection include screening programs such as low-dose computed tomography (LDCT) and the analysis of biomarkers. While these approaches have demonstrated some success, they are not without limitations [6]. LDCT, for instance, may lead to overdiagnosis and increased healthcare costs. AI systems can potentially enhance the effectiveness of these methods by providing more precise and efficient analysis, reducing false positives and false negatives, and offering a complementary approach to existing techniques. Despite the potential benefits of AI in early lung cancer detection, several challenges and considerations must be addressed [7,8]. The performance of AI models can vary depending on the quality and diversity of the data used for training [9,10]. Therefore, the selection and curation of data are fundamental to the success of AI-based systems in this context.

This systematic review and metanalysis endeavors to provide a comprehensive evaluation of the performance of AI systems for the early detection of lung cancer. This paper analyzed the current state of AI applications in lung cancer detection, including the various techniques and approaches being utilized. Furthermore, the study critically assessed the advantages and limitations of AI-based methods compared to traditional approaches.

2. Materials and Methods

In conducting this systematic review and meta-analysis evaluating the performance of AI systems for the early detection of lung cancer, we meticulously adhered to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 guidelines. The following detailed Method Section outlines the steps taken in this comprehensive review.

2.1. Protocol Development

The research question was formulated to assess the performance of AI systems in early lung cancer detection. A detailed protocol was developed, outlining the inclusion and exclusion criteria, search strategy, and methods for data extraction and analysis. In accordance with the journal's guidelines, this systematic review was not registered in

any specific database prior to its initiation. While the journals encourage registration for systematic reviews, it is not a mandatory requirement universally practiced in the field. This decision was aligned with established practices within this domain, considering the extensive body of previously published systematic reviews without prior registration in reputable peer-reviewed journals.

2.2. Literature Search

Comprehensive searches were conducted in electronic databases, including PubMed, Google Scholar, Science direct, and Embase, to identify relevant studies published up to October 2023. The inclusion of Google Scholar helped us to identify the grey literature, conference papers, and other non-traditional sources of information.

2.3. Search Strategy

The search strategy included a combination of keywords and Medical Subject Heading (MeSH) terms related to lung cancer, artificial intelligence, and early detection. The search strategy was tailored to each database to account for variations in syntax and indexing. This ensured that no relevant studies were missed.

2.4. Study Selection

Two independent reviewers screened titles and abstracts for eligibility based on predefined criteria. The initial screening of titles and abstracts helped in rapidly identifying studies that met the inclusion criteria and eliminating those that did not. Full-text articles of potentially relevant studies were retrieved for further assessment. A full-text review was conducted on potentially relevant studies to ensure that the selection process was rigorous, and that the final dataset was of high quality.

2.5. Eligibility Criteria

2.5.1. Inclusion Criteria

Studies evaluating the performance of AI systems for the early detection of lung cancer.

Original research articles published in English.

Studies reporting sensitivity, specificity, and other relevant diagnostic performance metrics.

2.5.2. Exclusion Criteria

Studies lacking sufficient data on AI system performance.

Reviews, commentaries, and conference abstracts without primary data.

2.6. Data Extraction

Relevant data, including study characteristics, AI system details, validation methods, and diagnostic performance metrics, were extracted using a standardized data extraction form. Two reviewers independently extracted data from selected studies. Discrepancies were resolved through discussion or consultation with a third reviewer.

2.7. Quality Assessment

The quality of the included studies was assessed using appropriate tools, considering factors such as study design, patient selection, and AI system evaluation. A risk-of-bias graph and summary were generated to visually represent the methodological quality of the included studies.

2.8. Data Synthesis and Analysis

The meta-analysis was carried out using R software (Version 4.3.0, Vienna, Austria) along with the RStudio interface (Version 2023.03.0, Boston, MA, USA). Packages and libraries such as meta and metafor were utilized to calculate key performance metrics,

including pooled sensitivity and specificity, along with their associated confidence intervals, i.e., 95%. Moreover, the presence of heterogeneity among the included studies was evaluated using a chi-square test and I^2 index statistics.

2.9. Reporting

A PRISMA flow diagram was used to illustrate the study selection process, including the number of studies identified, screened, assessed for eligibility, and included in the final analysis.

3. Results

The flow diagram in Figure 1 shows that the researchers identified 1024 records from the databases, but only 116 records were assessed for eligibility. At the identification stage, 326 records were excluded due to duplication. During the screening stage, 28 records were excluded because they were not in English. Some records lacked the essential data required for the systematic review. Records that did not have full-text versions available for review were excluded. Review articles, which summarize and analyze existing research, were excluded during the screening stage. After completing the identification and screening stages, the research team identified 39 studies that met the inclusion criteria and were relevant to the systematic review. These studies formed the basis for the subsequent data extraction and analysis, contributing to the comprehensive evaluation of AI systems for early lung cancer detection in the systematic review.

In Table 1, we present an overview of the characteristics of the included studies in our systematic review, each focusing on the application of AI models for the early detection of lung cancer. The table encompasses a diverse range of studies conducted across different countries and utilizing various AI models and data sources. When comparing and contrasting the results of these studies, several key insights emerge. While studies such as Wu et al. (2022) and Alexander et al. (2020) achieved notably high specificity levels, suggesting the potential for reducing false positives in clinical settings, Baldwin et al. (2020) achieved exceptionally high sensitivity, minimizing the risk of missing cancer cases [4,7,10]. On the other hand, the study by Chen (2022) showcases the effectiveness of AI models, specifically CNN and RNN, in improving the overall accuracy of lung cancer prediction [8]. Notably, Huang et al. (2018) integrated sensor array technology with machine learning, demonstrating its promise in the precise identification of lung cancer, especially when compared to traditional models [11]. Li et al. (2019) conducted a retrospective study in China using 3D deep learning technology on CT scans [9]. Their AI system achieved a sensitivity of 75% and specificity of 82%, resulting in an overall accuracy of 88.8%. This research emphasized the potential of AI as a diagnostic tool capable of providing more precise and unbiased outcomes in the diagnosis of pulmonary nodules, ultimately reducing the interpretation time for radiologists. Choi et al. (2018), from the USA, conducted a retrospective study using Support Vector Machine (SVM) and LASSO on LIDC-IDRI data [5]. Their AI model achieved an accuracy of 84.6%, which was notably 12.4% higher than the accuracy for Lung-RADS. This result demonstrated the potential of AI in substantially improving the accuracy of lung cancer detection. In another study from China employed a 3D CMixNet model on LUNA-16 and LIDC-IDRI datasets. Their system achieved a sensitivity of 94.0% and specificity of 91.0%, showcasing better results compared to existing methods for lung cancer detection. These variations in results highlight the trade-offs between sensitivity and specificity, as well as the distinct strengths of different AI models and approaches. While some studies emphasize the potential of AI in overcoming specific challenges, such as PD-L1 assessment or eligibility assessment, others underscore the efficiency and reliability of AI in lung cancer screening. Collectively, these findings underscore the transformative potential of AI in enhancing the accuracy and efficiency of lung cancer diagnosis, promising significant benefits to both patients and healthcare professionals. Collectively, these results underscore the transformative role AI can play in improving the accuracy, efficiency, and reliability of lung cancer diagnosis, ultimately

benefiting patients and healthcare professionals. Our systematic review incorporates these findings to offer a holistic understanding of the state of AI in lung cancer detection, shedding light on the remarkable potential of these technologies in the field of oncology.

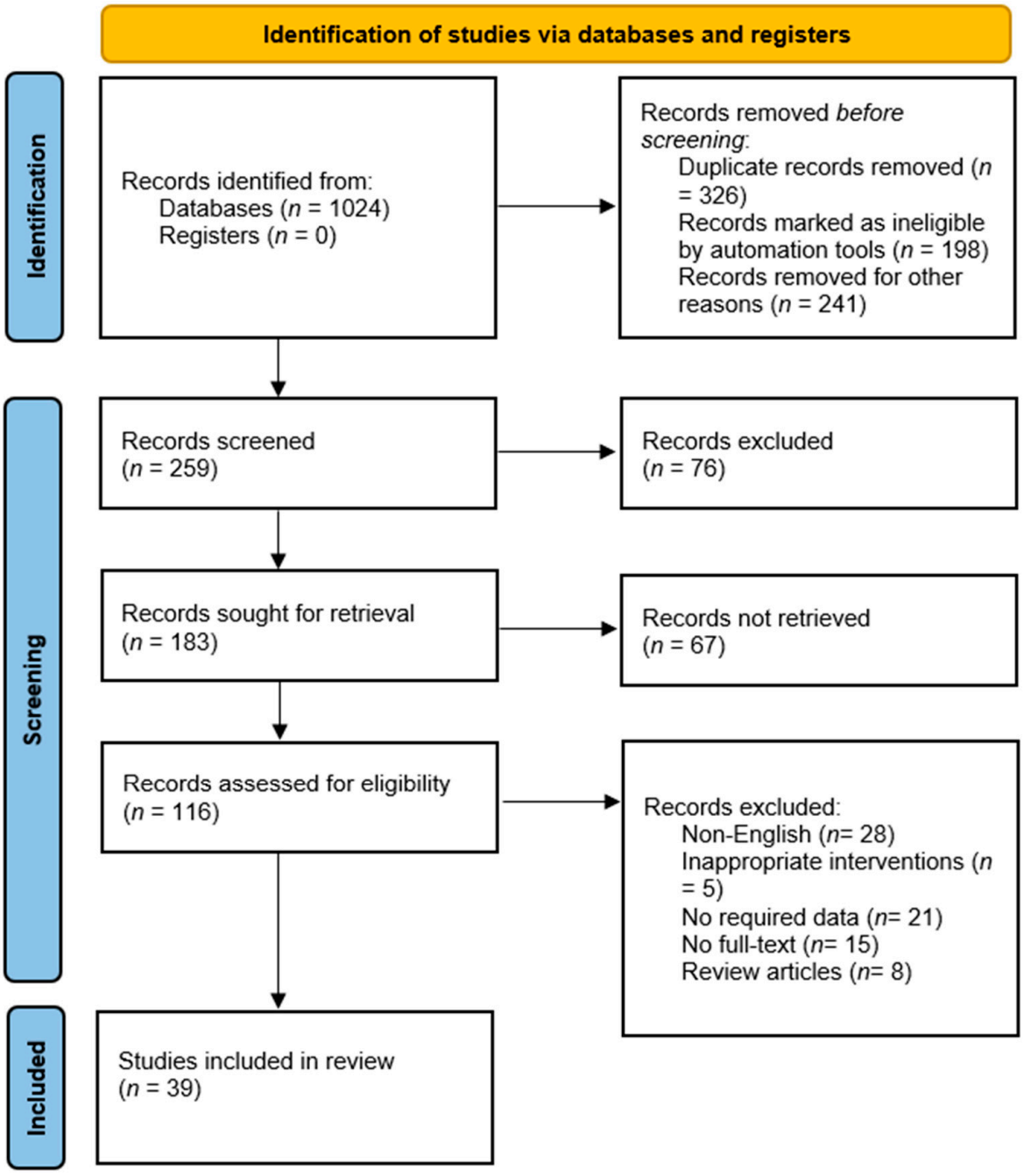

Figure 1. Flowchart of included studies.

Table 1. Characteristics of included studies.

Author and Year	Country	Study Design	AI Models Used	Source of Dataset	Key Performance Metrics of AI Models			Validation Method	Outcomes
					Sensitivity	Specificity	Accuracy		
Wu et al., 2022 [7]	China	Retrospective study	DL	Whole slide images	86%	96.4%	93.2%	NA	AI-assisted system can be an effective and valuable tool to overcome the challenges of PD-L1 assessment.
Chen 2022 [8]	China	Retrospective study	CNN and RNN	H-E stained pathological slices	95.1%	34.4%	92%	NA	AI-based approach using CNN and RCN could be very effective for improving the accuracy of predicting lung cancer.
Li et al., 2019 [9]	China	Retrospective study	3D deep learning technology	CT scans	75%	82%	88.8%	Internal validation	AI may represent a valuable diagnostic tool that shows more precise and unbiased outcomes in the diagnosis of pulmonary nodules, thus minimizing the time required for interpretation of results by radiologists.
Alexander et al., 2020 [10]	Australia	Retrospective study	ML and NLP	TMC	83.3%	93.8%	91.6%	Internal validation	The AI-assisted system enables efficient and reliable screening of cancer patients, with an accuracy of 91.6% for overall eligibility assessment.
Baldwin et al., 2020 [4]	UK	Retrospective study	LCP-CNN	EDC	99.5%	28%	-	External validation	The LCP-CNN approach minimizes the risk of missing cancer cases when compared to the Brock model.
Choi et al., 2018 [5]	USA	Retrospective study	SVM-LASSO	LIDC-IDRI	87.2%	81.2%	84.6%	Cross-validation	The proposed AI model achieved 84.6% accuracy, which was 12.4% higher than accuracy for Lung-RADS.

Table 1. *Cont.*

Author and Year	Country	Study Design	AI Models Used	Source of Dataset	Key Performance Metrics of AI Models			Validation Method	Outcomes
					Sensitivity	Specificity	Accuracy		
Huang et al., 2018 [11]	Taiwan	Prospective study	ML	-	For internal validation 92.3% For external validation 83.3%	For internal validation 92.9% For external validation 86.2%	For internal validation 92.7% For external validation 85.4%	External validation/ internal validation	The integration of the sensor array technique and machine learning enables the precise identification of lung cancer with high accuracy.
Li et al., 2019 [12]	China	Retrospective study	3D deep CNN (SS-OLHF)	LIDC_IDRI	82.6%	91.3%	93.0%	Cross-validation	The proposed fusion algorithm achieved the highest specificity, sensitivity, and accuracy scores among all classification models.
Nasrullah et al., 2019 [13]	China	Retrospective study	3D CMixNet	LUNA-16 and LIDC-IDRI	94.0%	91.0%	-	-	The proposed system, evaluated on LIDC-IDRI datasets, showed better results compared to the existing methods.
Reddy et al., 2018 [14]	India	Retrospective study	GLCM	Microarray data clustering mechanisms	95.3%	100%	97.6%	Cross-validation	GLCM features predicted lung tumor with higher accuracy than histogram features.
Schwyzer et al., 2018 [15]	Switzerland	Retrospective study	ANN	Internal	93.6%	95.5%	94.7%	Cross-validation	ML algorithms may help in fully automated lung cancer detection even at a low effective radiation doses.
Ardila et al., 2019 [16]	USA	Retrospective study	3D CNN	LUNA, LIDC and NLST	64.7%	95.2%	-	Internal validation	DL models have the potential to increase consistency and accuracy and enable the adoption of lung cancer screening.

Table 1. *Cont.*

Author and Year	Country	Study Design	AI Models Used	Source of Dataset	Key Performance Metrics of AI Models			Validation Method	Outcomes
					Sensitivity	Specificity	Accuracy		
Coudray et al., 2018 [17]	USA	Retrospective study	CNN	TCGA	89%	93%	-	Internal validation	The outcomes suggest that DL can assist healthcare professionals in the detection of cancer.
Hussein et al., 2017 [18]	USA	-	3D CNN	LIDC-IDRI	-	-	91.2%	Internal validation	The proposed approach achieved cutting-edge outcomes in regressing malignancy scores.
Venkadesh et al., 2021 [19]	Denmark	Retrospective study	DL	NLST and DLCST	For subset A: 91% For subset B: 54%	90%	-	Cross-validation	This algorithm has the potential to provide reliable and reproducible malignancy risk scores for experts, which could contribute to promoting the effectiveness of lung cancer screening management.
Ciompi et al., 2017 [20]	The Netherlands	Retrospective study	CNN	MILD and DLCST	-	-	72.9%	Internal validation	The performance of the proposed DL model in classifying nodule type surpassed that of conventional machine learning approaches.
Petousis et al., 2016 [21]	USA	Retrospective study	DBN	NLST	-	-	-	Cross-validation	The lung cancer screening DBNs were reported to have high discrimination and predictive power with most of the cancer as well as non-cancer cases.
Zhang et al., 2019 [22]	China	Retrospective study	3D CNN	LUNA and Kaggle	84.4%	83%	-	Cross-validation	The 3D CNN with a DL algorithm may assist experts by providing accurate data for diagnosing pulmonary nodules.

Table 1. *Cont.*

Author and Year	Country	Study Design	AI Models Used	Source of Dataset	Key Performance Metrics of AI Models			Validation Method	Outcomes
					Sensitivity	Specificity	Accuracy		
Petousis et al., 2019 [23]	USA	Retrospective study	ML and DBN	NLST	-	-	-	Cross-validation	The proposed model reduced the FPR while maintaining TPR, and improved early prediction of cancer cases.
Huang et al., 2019 [24]	USA	Retrospective study	DL	NLST and PanCan	88%	60%	-	External validation	DL scores could be used for early detection of cancer cases.
Cui et al., 2020 [25]	China	Retrospective study	DL	LUNA	90%	85%	-	External validation	The DL system had better identification sensitivity and performance than that of the experts.
Chauvie et al., 2020 [26]	Italy	Retrospective study	Neural network	-	90%	100%	100%	Cross-validation	The utilization of visual analysis along with NN could help experts to reduce the number of false positive cases.
Tam et al., 2021 [27]	UK	Retrospective study	DNN	NHS Cancer Registry Database	80%	93%	87%	-	The proposed AI resulted in a reduction in radiologist errors and improved clinician reporting performance.
Schwyzer et al., 2020 [28]	Switzerland	Retrospective study	DL	-	For BSREM 69.2% For OSEM 66.7%	For BSREM 84.5% For OSEM 79.0%	-	Cross-validation	AI performed substantially better on images with BSREM than OSEM.
Teramoto et al., 2010 [29]	Japan	Retrospective study	CNN	-	90.1%	-	-	Cross-validation	CNN technique can be very effective for the early detection of pulmonary nodules in PET/CT images.
Kirienko et al., 2018 [30]	Italy	Retrospective study	CNN	Independent dataset	-	67%	69%	Cross-validation	CNNs can be used as a reliable tool to assist in the staging of lung cancer patients.

Table 1. *Cont.*

Author and Year	Country	Study Design	AI Models Used	Source of Dataset	Key Performance Metrics of AI Models			Validation Method	Outcomes
					Sensitivity	Specificity	Accuracy		
Sibille et al., 2020 [31]	US and Germany	Retrospective study	Deep CNN	Independent dataset	81.0%	97.3%	-	Independent internal validation	CNN achieved high diagnostic performance when both PET and CT images were utilized.
Toney et al., 2014 [32]	USA	Prospective study	ANN	Independent dataset	-	-	99.2%	-	ANNs can provide more accurate and consistent assessment of nodal stage in lung cancer patients.
Scott et al., 2019 [33]	USA	Retrospective study	ANN	Independent dataset	100%	93.1%	-	External validation	ANNs have potential to improve diagnostic certainty and can be useful to help direct clinical and imaging follow-up.
Hyun et al., 2019 [34]	Korea	Retrospective study	RF, NN, NBS, LL, SVM	Independent dataset	SVM 52.6% RF 52.6% NN 52.6% NBS 52.6% LL 52.6%	-	SVM 67.1% RF 67.1% NN 67.1% NBS 67.1% LL 67.1%	Internal validation	An ML approach with PET-based radiomics aids in early detection of lung cancer.
Jayasurya et al., 2010 [35]	Belgium and the Netherlands	Retrospective study	BN	Independent dataset	-	-	-	External validation	BN models are better at handling missing data as compared to SVM models and are more suitable for the medical domain.

Table 1. *Cont.*

Author and Year	Country	Study Design	AI Models Used	Source of Dataset	Key Performance Metrics of AI Models			Validation Method	Outcomes
					Sensitivity	Specificity	Accuracy		
Luo et al., 2018 [36]	USA	Retrospective study	BN	Independent dataset	-	-	-	Cross-validation	The proposed BN model is stable and can identify hierarchical relationships among biophysical features for the prediction of lung cancer.
Chamberlin et al., 2021 [37]	USA, Europe, and Asia	Retrospective study	CNN	Independent data set	100%	70.8%	-	Internal validation	AI software strongly agrees with expert radiologist determination of detection of both lung nodules and CACV.
Hsu et al., 2020 [6]	Taiwan	Retrospective study	ANN	Hospital-based cancer registry	75.0%	85.0%	-	Cross-validation	This study reported that ANN had better sensitivity for the detection of lung cancer than Lung-RADS.
Duan et al., 2020 [38]	China	Retrospective study	Decision tree C5.0, ANN, SVM	-	C5.0-1 61.8% SVM 59.2% ANN 81.5%	C5.0-1 73.4% SVM 68.7% ANN 65.6%	-	-	These AI models can be utilized for the on-site screening and clinical diagnosis of the high-risk population.
Silva et al., 2017 [39]	Brazil	Retrospective study	Deep CNN	LIDC-IDRI	94.6%	95.1%	94.7%	Internal validation	The proposed DL models demonstrated promising performance and avoided the need for feature extraction and selection steps.
Trajanovski et al., 2021 [40]	USA	Retrospective study	CNN	Kaggle, NLST, and UCM	93%	-	-	External validation	The proposed DL had a better sensitivity of 93% and can be used for the screening of lung cancer.

Table 1. *Cont.*

Author and Year	Country	Study Design	AI Models Used	Source of Dataset	Key Performance Metrics of AI Models			Validation Method	Outcomes
					Sensitivity	Specificity	Accuracy		
Chen et al., 2022 [41]	China	Retrospective study	CNN	Self-built data from hospital	94.1%	77.7%	87.1%	External validation	The DL-based AI film reading system has higher sensitivity for the diagnosis of NSCLC than radiologists.
Coruch et al., 2021 [42]	Turkey	Retrospective study	CNN	Hospital records	92.2%	58.7%	75.2%	Cross-validation	The performance of the observers in evaluating the risk of malignancy was slightly higher than the performance of fusion AI algorithms.

PD-L1 = programmed death-ligand 1, CNN = convolutional neural network, RNN = recurrent neural network, ML = machine learning, NLP = natural language processing, TMC = thoracic malignancies cohort, SVM-LASSO = support vector machine–least absolute linkage and selection operator, EDC = electronic data capture, RADS = Reporting and Data System, CMixNet = customized mixed link network, LIDC = lung image database consortium, IDRI = image database resource initiative. GLCM = gray level co-event matrix, NLST = National Lung Cancer Screening Trial, LUNA = lung nodule analysis, TCGA = the Cancer Genome Atlas, DLCST = Danish lung cancer screening trial, MILD = multicenter Italian lung detection, RF = random Forest, BN = Bayesian network, DNN = deep neural network, BSREM = block-sequential regularized expectation maximization, OSEM = ordered subset expectation maximization, ANN = artificial neural network, GGO = ground-glass opacities, LL = logistic regression, NBS = naïve Bayesian system, NN = neural network, UCM = University of Chicago.

Figures 2 and 3 present forest plots of the pooled sensitivity and sensitivity of AI models for the early diagnosis of lung cancer. The pooled sensitivity and specificity of AI models across the included studies were 0.87 (95% CI: 0.82–0.90) and 0.87 (95% CI: 0.80–0.91), respectively. These results indicate that AI models demonstrated a high level of accuracy in correctly identifying true positives and true negatives, showing promising results for the early diagnosis of lung cancer. However, heterogeneity was observed among the included studies. This heterogeneity may be attributed to variations in study populations, data sources, and model specifications. The results of quality assessments are presented in Figure 4.

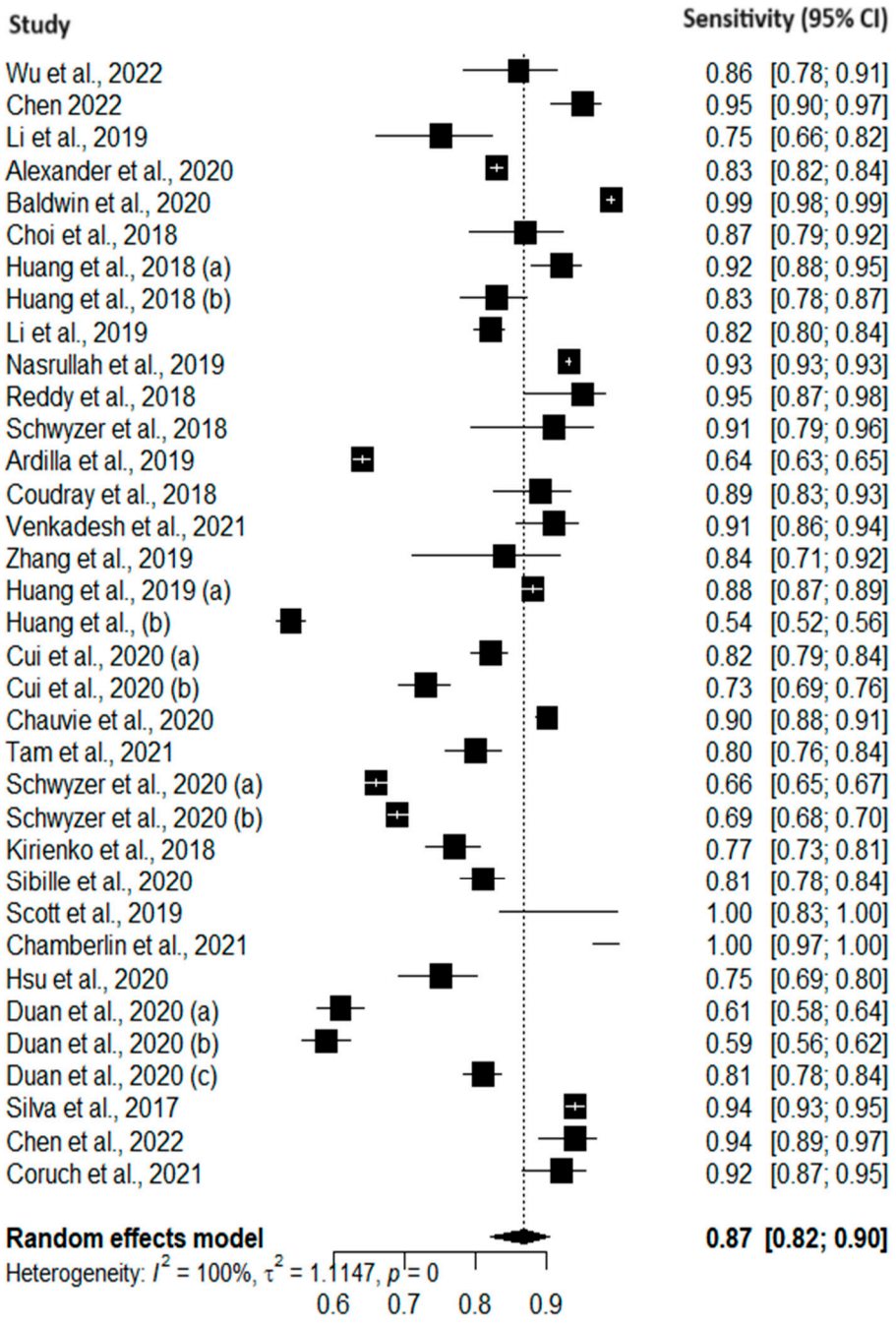

Figure 2. Forest plot of pooled sensitivity of AI models [4,5,7–42].

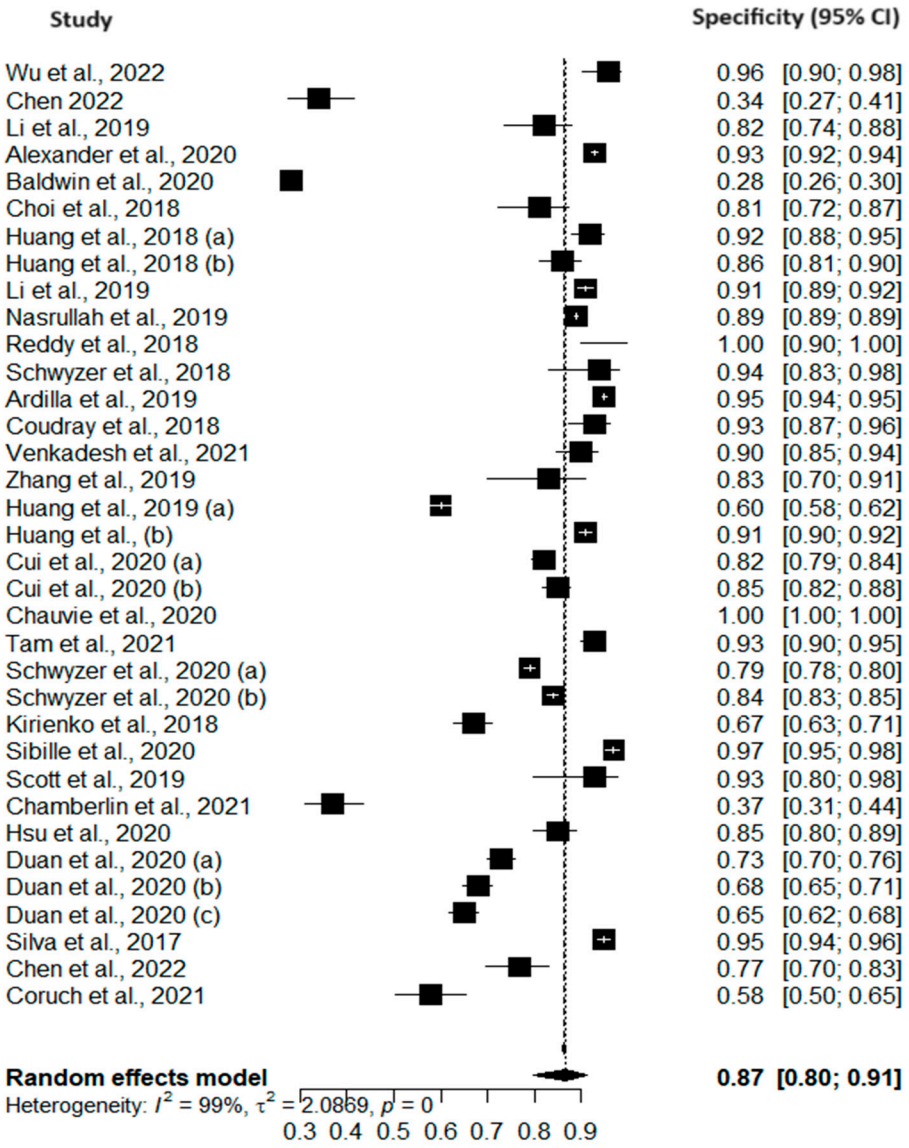

Figure 3. Forest plot of pooled specificity of AI models [4,5,7–42].

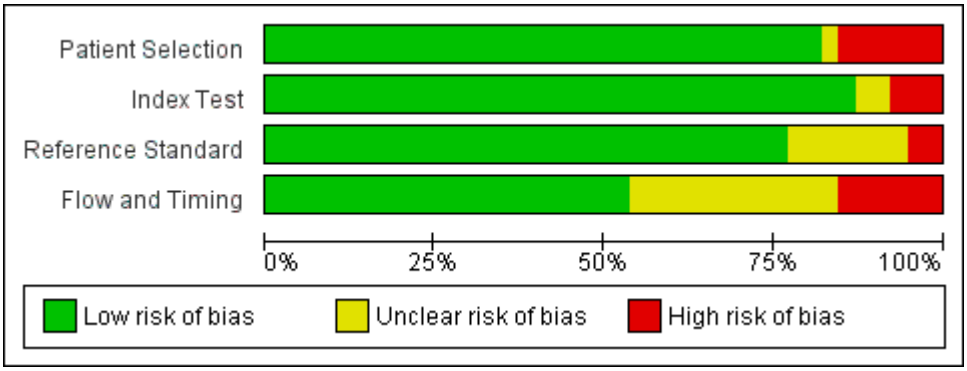

Figure 4. Quality assessment of included studies using QUADAS-2 tool [43].

4. Discussion

Lung cancer is one of the most prevalent diseases worldwide and the leading cause of cancer-associated deaths, with an estimated 2.2 million new cases and 1.8 million deaths in 2020 [44]. Currently, a CT scan of the chest is the most frequent method of lung cancer

screening. Its high resolution can elucidate the association among surrounding organs and blood vessels more clearly, and it plays a significant role in the early detection of lung cancer [45]. However, the accuracy of this method can be influenced by benign lesions such as necrosis, inflammation, tuberculosis, various textures in lung images, and several other factors like the experience of radiologists, potentially leading to misdiagnosis and omissions [46]. With the implementation of AI-assisted diagnostic systems into clinical practice, a new era has dawned in the field of lung cancer diagnosis. Recent studies have documented the growing and widespread utilization of AI models in clinical diagnosis and treatment, respectively [47–49]. AI models primarily focus on diagnosing and evaluating various medical images, including skin lesions, pathological microscopic images, and radiological data. AI models are remarkable in their ability to enhance diagnostic accuracy, stability, and work efficiency.

This review documented promising results, indicating that AI models for the early diagnosis of lung cancer demonstrated a high level of accuracy, with pooled sensitivity and specificity values of 0.87 (95% CI: 0.82–0.90) and 0.87 (95% CI: 0.80–0.91), respectively. These findings suggest that AI models exhibit significant capability in correctly identifying true positives and true negatives. Liu et al. recently conducted a systematic review and meta-analysis in which they also demonstrated the commendable performance of AI models in predicting lung cancer, with a pooled sensitivity and specificity of 89% and 87%, respectively [46]. Robust performance is sometimes crucial in terms of lung cancer, where early detections substantially impact patient outcomes. High accuracy of more than 90% was observed in this study, which aligns with the broader trend in the medical literature supporting the effectiveness of AI in diagnostic settings [7,8,10,12,14,15,26,32]. However, a lower pooled accuracy was also reported in studies mainly focused on lung cancer screening, specifically considering the results obtained across all studies [20,30,34,42] and those in which a CNN model was employed [20,30,42], ranging from 67–75%, respectively. Despite these consolidated findings, we are confident that AI models are a valuable resource for radiologists to detect lung cancer.

AI-assisted diagnostic systems result in different diagnostic outcomes. A study reported that a 3D CNN model exhibited greater advantages in detecting lung cancer as compared to improvements seen with other AI models [18]. However, in two studies, ANN achieved high diagnostic performance that could be useful for the detection of lung cancer [32,33]. Distinct algorithms demonstrate diverse diagnostic capabilities, notably in radionics and deep learning, which not only assist in predicting the benign or malignant nature of lung nodules, but also identify the prognosis of small-cell lung cancer [50,51]. The utilization of AI models in clinical practice is promising; however, validity remains a critical step for generalizability. Among 39 articles, 15 articles performed cross-validation to assess the effectiveness and reliability of AI models [5,6,12,14,15,19,21,23,26,28–30,36,42,50].

Despite the overall positive outcomes, observed heterogeneity among the included studies was identified. Therefore, future research should focus on refining AI models, considering the identified heterogeneity challenges. Collaborative efforts among researchers, clinicians, and policymakers are essential to establish guidelines and standards for the development and evaluation of AI systems in lung cancer screening. By addressing these challenges collectively, the field can progress toward the implementation of AI technologies in clinical settings, ultimately improving the early diagnosis and management of lung cancer.

The current study has certain limitations that should be addressed: (1) The exclusion of the studies lacking complete diagnostic data may have altered the results. (2) While conducting this comprehensive search, only English language articles were included, potentially introducing language bias. (3) The high heterogeneity among all included studies may be attributed to variations in study populations, data sources, and model specifications, and these results warrant further investigation. (4) The included studies were mainly designed retrospectively, which may have affected the overall quality of the systematic review and meta-analysis. Despite the initial verification of AI models' effectiveness in lung

cancer screening, most of the AI-based approaches are still in the laboratory research stage and have not yet been implemented into clinical practice. Limitations are evident in data integration, image data quality, legal liability definition, complex pathology diagnosis and cost of use. However, a huge volume of experienced healthcare professionals, especially radiologists and pathologists, are getting involved in AI-assisted lung cancer detection. It is anticipated that AI models will play a significant role in the early detection of lung cancer.

5. Conclusions

This systematic review and meta-analysis reported the promising outcomes of AI models in the early detection of lung cancer. The pooled sensitivity and specificity values of 0.87 (95% CI: 0.82–0.90) and 0.87 (95% CI: 0.80–0.91) showed the potential of AI models in identifying true positives and true negatives. Regarding the observed heterogeneity among the included studies, these findings highlight the need for standardized protocols in the development of AI models for lung cancer screening. As the medical field continues to grow, healthcare professionals and patients will benefit from the integration of AI models in clinical practice once these models have been validated in large-scale prospective studies.

Author Contributions: Conceptualization, M.K., L.A. (Leen Albraik) and H.A. (Hajar Alharbi); methodology, N.A. and L.A. (Lubna Almasuood); validation, S.A. (Shahad Aljoaid) and T.A.; formal analysis, L.A. (Lubna Almasuood) and W.A.; investigation, H.A. (Hajar Alharbi) and A.A.; resources, S.A. (Shahad Aljoaid), R.K., and R.A.; data curation, R.K., R.A. and M.A.; writing—original draft preparation, M.K., L.A. (Leen Albraik), H.A. (Hadeel Aljohani) and N.A.; writing—review and editing, S.A. (Sadeem Alqahtani), A.A. and R.K.; visualization, R.A., M.A. and A.M.; project administration, M.K., S.A. (Sadeem Alqahtani) and H.A. (Hadeel Aljohani). All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

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

References

1. Wang, Z.; Hu, L.; Li, J.; Wei, L.; Zhang, J.; Zhou, J. Magnitude, temporal trends and inequality in global burden of tracheal, bronchus and lung cancer: Findings from the Global Burden of Disease Study 2017. *BMJ Glob. Health* **2020**, *5*, e002788. [CrossRef] [PubMed]
2. Chaft, J.E.; Rimner, A.; Weder, W.; Azzoli, C.G.; Kris, M.G.; Cascone, T. Evolution of systemic therapy for stages I–III non-metastatic non-small-cell lung cancer. *Nat. Rev. Clin. Oncol.* **2021**, *18*, 547–557. [CrossRef] [PubMed]
3. Roointan, A.; Mir, T.A.; Wani, S.I.; Hussain, K.K.; Ahmed, B.; Abrahim, S.; Savardashtaki, A.; Gandomani, G.; Gandomani, M.; Chinnappan, R. Early detection of lung cancer biomarkers through biosensor technology: A review. *J. Pharm. Biomed. Anal.* **2019**, *164*, 93–103. [CrossRef]
4. Baldwin, D.R.; Gustafson, J.; Pickup, L.; Arteta, C.; Novotny, P.; Declerck, J.; Kadir, T.; Figueiras, C.; Sterba, A.; Exell, A. External validation of a convolutional neural network artificial intelligence tool to predict malignancy in pulmonary nodules. *Thorax* **2020**, *75*, 306–312. [CrossRef]
5. Choi, W.; Oh, J.H.; Riyahi, S.; Liu, C.J.; Jiang, F.; Chen, W.; White, C.; Rimner, A.; Mechalakos, J.G.; Deasy, J.O. Radiomics analysis of pulmonary nodules in low-dose CT for early detection of lung cancer. *Med. Phys.* **2018**, *45*, 1537–1549. [CrossRef]
6. Hsu, Y.-C.; Tsai, Y.-H.; Weng, H.-H.; Hsu, L.-S.; Tsai, Y.-H.; Lin, Y.-C.; Hung, M.-S.; Fang, Y.-H.; Chen, C.-W. Artificial neural networks improve LDCT lung cancer screening: A comparative validation study. *BMC Cancer* **2020**, *20*, 1023. [CrossRef]
7. Wu, J.; Liu, C.; Liu, X.; Sun, W.; Li, L.; Gao, N.; Zhang, Y.; Yang, X.; Zhang, J.; Wang, H. Artificial intelligence-assisted system for precision diagnosis of PD-L1 expression in non-small cell lung cancer. *Mod. Pathol.* **2022**, *35*, 403–411. [CrossRef]
8. Chen, S. Models of artificial intelligence-assisted diagnosis of lung cancer pathology based on deep learning algorithms. *J. Healthc. Eng.* **2022**, *2022*, 3972298. [CrossRef]
9. Li, X.; Hu, B.; Li, H.; You, B. Application of artificial intelligence in the diagnosis of multiple primary lung cancer. *Thorac. Cancer* **2019**, *10*, 2168–2174. [CrossRef]
10. Alexander, M.; Solomon, B.; Ball, D.L.; Sheerin, M.; Dankwa-Mullan, I.; Preininger, A.M.; Jackson, G.P.; Herath, D.M. Evaluation of an artificial intelligence clinical trial matching system in Australian lung cancer patients. *JAMIA Open* **2020**, *3*, 209–215. [CrossRef]
11. Huang, C.-H.; Zeng, C.; Wang, Y.-C.; Peng, H.-Y.; Lin, C.-S.; Chang, C.-J.; Yang, H.-Y. A study of diagnostic accuracy using a chemical sensor array and a machine learning technique to detect lung cancer. *Sensors* **2018**, *18*, 2845. [CrossRef]
12. Li, S.; Xu, P.; Li, B.; Chen, L.; Zhou, Z.; Hao, H.; Duan, Y.; Folkert, M.; Ma, J.; Huang, S. Predicting lung nodule malignancies by combining deep convolutional neural network and handcrafted features. *Phys. Med. Biol.* **2019**, *64*, 175012. [CrossRef]

13. Nasrullah, N.; Sang, J.; Alam, M.S.; Mateen, M.; Cai, B.; Hu, H. Automated lung nodule detection and classification using deep learning combined with multiple strategies. *Sensors* **2019**, *19*, 3722. [CrossRef]
14. Reddy, U.J.; Ramana Reddy, B.V.; Reddy, B.E. Categorization & recognition of lung tumor using machine learning representations. *Curr. Med. Imaging* **2019**, *15*, 405–413.
15. Schwyzer, M.; Ferraro, D.A.; Muehlematter, U.J.; Curioni-Fontecedro, A.; Huellner, M.W.; von Schulthess, G.K.; Kaufmann, P.A.; Burger, I.A.; Messerli, M. Automated detection of lung cancer at ultralow dose PET/CT by deep neural networks—Initial results. *Lung Cancer* **2018**, *126*, 170–173. [CrossRef]
16. Ardila, D.; Kiraly, A.P.; Bharadwaj, S.; Choi, B.; Reicher, J.J.; Peng, L.; Tse, D.; Etemadi, M.; Ye, W.; Corrado, G. End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. *Nat. Med.* **2019**, *25*, 954–961. [CrossRef]
17. Coudray, N.; Ocampo, P.S.; Sakellaropoulos, T.; Narula, N.; Snuderl, M.; Fenyö, D.; Moreira, A.L.; Razavian, N.; Tsirigos, A. Classification and mutation prediction from non–small cell lung cancer histopathology images using deep learning. *Nat. Med.* **2018**, *24*, 1559–1567. [CrossRef]
18. Hussein, S.; Cao, K.; Song, Q.; Bagci, U. *Risk Stratification of Lung Nodules Using 3D CNN-Based Multi-Task Learning*; Springer International Publishing: Berlin/Heidelberg, Germany, 2017; pp. 249–260.
19. Venkadesh, K.V.; Setio, A.A.A.; Schreuder, A.; Scholten, E.T.; Chung, K.; Wille, W.M.M.; Saghir, Z.; van Ginneken, B.; Prokop, M.; Jacobs, C. Deep learning for malignancy risk estimation of pulmonary nodules detected at low-dose screening CT. *Radiology* **2021**, *300*, 438–447. [CrossRef]
20. Ciompi, F.; Chung, K.; Van Riel, S.J.; Setio, A.A.A.; Gerke, P.K.; Jacobs, C.; Scholten, E.T.; Schaefer-Prokop, C.; Wille, M.M.W.; Marchiano, A. Towards automatic pulmonary nodule management in lung cancer screening with deep learning. *Sci. Rep.* **2017**, *7*, 46479. [CrossRef] [PubMed]
21. Petousis, P.; Han, S.X.; Aberle, D.; Bui, A.A.T. Prediction of lung cancer incidence on the low-dose computed tomography arm of the National Lung Screening Trial: A dynamic Bayesian network. *Artif. Intell. Med.* **2016**, *72*, 42–55. [CrossRef] [PubMed]
22. Zhang, C.; Sun, X.; Dang, K.; Li, K.; Guo, X.w.; Chang, J.; Yu, Z.Q.; Huang, F.Y.; Wu, Y.S.; Liang, Z. Toward an expert level of lung cancer detection and classification using a deep convolutional neural network. *Oncologist* **2019**, *24*, 1159–1165. [CrossRef]
23. Petousis, P.; Winter, A.; Speier, W.; Aberle, D.R.; Hsu, W.; Bui, A.A.T. Using sequential decision making to improve lung cancer screening performance. *Ieee Access* **2019**, *7*, 119403–119419. [CrossRef]
24. Huang, P.; Lin, C.T.; Li, Y.; Tammemagi, M.C.; Brock, M.V.; Atkar-Khattra, S.; Xu, Y.; Hu, P.; Mayo, J.R.; Schmidt, H. Prediction of lung cancer risk at follow-up screening with low-dose CT: A training and validation study of a deep learning method. *Lancet Digit. Health* **2019**, *1*, e353–e362. [CrossRef] [PubMed]
25. Cui, S.; Ming, S.; Lin, Y.; Chen, F.; Shen, Q.; Li, H.; Chen, G.; Gong, X.; Wang, H. Development and clinical application of deep learning model for lung nodules screening on CT images. *Sci. Rep.* **2020**, *10*, 13657. [CrossRef] [PubMed]
26. Chauvie, S.; De Maggi, A.; Baralis, I.; Dalmasso, F.; Berchialla, P.; Priotto, R.; Violino, P.; Mazza, F.; Melloni, G.; Grosso, M. Artificial intelligence and radiomics enhance the positive predictive value of digital chest tomosynthesis for lung cancer detection within SOS clinical trial. *Eur. Radiol.* **2020**, *30*, 4134–4140. [CrossRef]
27. Tam, M.D.B.S.; Dyer, T.; Dissez, G.; Morgan, T.N.; Hughes, M.; Illes, J.; Rasalingham, R.; Rasalingham, S. Augmenting lung cancer diagnosis on chest radiographs: Positioning artificial intelligence to improve radiologist performance. *Clin. Radiol.* **2021**, *76*, 607–614. [CrossRef]
28. Schwyzer, M.; Martini, K.; Benz, D.C.; Burger, I.A.; Ferraro, D.A.; Kudura, K.; Treyer, V.; von Schulthess, G.K.; Kaufmann, P.A.; Huellner, M.W.; et al. Artificial intelligence for detecting small FDG-positive lung nodules in digital PET/CT: Impact of image reconstructions on diagnostic performance. *Eur. Radiol.* **2020**, *30*, 2031–2040. [CrossRef]
29. Teramoto, A.; Fujita, H.; Yamamuro, O.; Tamaki, T. Automated detection of pulmonary nodules in PET/CT images: Ensemble false-positive reduction using a convolutional neural network technique. *Med. Phys.* **2016**, *43*, 2821–2827. [CrossRef]
30. Kirienko, M.; Sollini, M.; Silvestri, G.; Mognetti, S.; Voulaz, E.; Antunovic, L.; Rossi, A.; Antiga, L.; Chiti, A. Convolutional neural networks promising in lung cancer T-parameter assessment on baseline FDG-PET/CT. *Contrast Media Mol. Imaging* **2018**, *2018*, 1382309. [CrossRef]
31. Sibille, L.; Seifert, R.; Avramovic, N.; Vehren, T.; Spottiswoode, B.; Zuehlsdorff, S.; Schäfers, M. 18F-FDG PET/CT uptake classification in lymphoma and lung cancer by using deep convolutional neural networks. *Radiology* **2020**, *294*, 445–452. [CrossRef]
32. Toney, L.K.; Vesselle, H.J. Neural networks for nodal staging of non–small cell lung cancer with FDG PET and CT: Importance of combining uptake values and sizes of nodes and primary tumor. *Radiology* **2014**, *270*, 91–98. [CrossRef] [PubMed]
33. Scott, J.A.; McDermott, S.; Kilcoyne, A.; Wang, Y.; Halpern, E.F.; Ackman, J.B. Comparison of 18F-FDG avidity at PET of benign and malignant pure ground-glass opacities: A paradox? Part II: Artificial neural network integration of the PET/CT characteristics of ground-glass opacities to predict their likelihood of malignancy. *Clin. Radiol.* **2019**, *74*, 692–696. [CrossRef]
34. Hyun, S.H.; Ahn, M.S.; Koh, Y.W.; Lee, S.J. A machine-learning approach using PET-based radiomics to predict the histological subtypes of lung cancer. *Clin. Nucl. Med.* **2019**, *44*, 956–960. [CrossRef]
35. Jayasurya, K.; Fung, G.; Yu, S.; Dehing-Oberije, C.; De Ruysscher, D.; Hope, A.; De Neve, W.; Lievens, Y.; Lambin, P.; Dekker, A. Comparison of Bayesian network and support vector machine models for two-year survival prediction in lung cancer patients treated with radiotherapy. *Med. Phys.* **2010**, *37*, 1401–1407. [CrossRef]

36. Luo, Y.; McShan, D.; Ray, D.; Matuszak, M.; Jolly, S.; Lawrence, T.; Kong, F.-M.; Ten Haken, R.; El Naqa, I. Development of a fully cross-validated Bayesian network approach for local control prediction in lung cancer. *IEEE Trans. Radiat. Plasma Med. Sci.* **2018**, *3*, 232–241. [CrossRef]
37. Chamberlin, J.; Kocher, M.R.; Waltz, J.; Snoddy, M.; Stringer, N.F.C.; Stephenson, J.; Sahbaee, P.; Sharma, P.; Rapaka, S.; Schoepf, U.J.; et al. Automated detection of lung nodules and coronary artery calcium using artificial intelligence on low-dose CT scans for lung cancer screening: Accuracy and prognostic value. *BMC Med.* **2021**, *19*, 55. [CrossRef]
38. Duan, S.; Cao, H.; Liu, H.; Miao, L.; Wang, J.; Zhou, X.; Wang, W.; Hu, P.; Qu, L.; Wu, Y. Development of a machine learning-based multimode diagnosis system for lung cancer. *Aging* **2020**, *12*, 9840. [CrossRef]
39. da Silva, G.L.F.; da Silva Neto, O.P.; Silva, A.C.; de Paiva, A.C.; Gattass, M. Lung nodules diagnosis based on evolutionary convolutional neural network. *Multimed. Tools Appl.* **2017**, *76*, 19039–19055. [CrossRef]
40. Trajanovski, S.; Mavroeidis, D.; Swisher, C.L.; Gebre, B.G.; Veeling, B.S.; Wiemker, R.; Klinder, T.; Tahmasebi, A.; Regis, S.M.; Wald, C.; et al. Towards radiologist-level cancer risk assessment in CT lung screening using deep learning. *Comput. Med. Imaging Graph.* **2021**, *90*, 101883. [CrossRef]
41. Chen, Y.; Tian, X.; Fan, K.; Zheng, Y.; Tian, N.; Fan, K. The value of artificial intelligence film reading system based on deep learning in the diagnosis of non-Small-Cell lung cancer and the significance of efficacy monitoring: A retrospective, clinical, nonrandomized, controlled study. *Comput. Math. Methods Med.* **2022**, *2022*, 2864170. [CrossRef]
42. Gürsoy Çoruh, A.; Yenigün, B.; Uzun, Ç.; Kahya, Y.; Büyükceran, E.U.; Elhan, A.; Orhan, K.; Kayı Cangır, A. A comparison of the fusion model of deep learning neural networks with human observation for lung nodule detection and classification. *Br. J. Radiol.* **2021**, *94*, 20210222. [CrossRef]
43. Whiting, P.F.; Rutjes, A.W.S.; Westwood, M.E.; Mallett, S.; Deeks, J.J.; Reitsma, J.B.; Leeflang, M.M.G.; Sterne, J.A.C.; Bossuyt, P.M.M.; QUADAS-2 Group. QUADAS-2: A revised tool for the quality assessment of diagnostic accuracy studies. *Ann. Intern. Med.* **2011**, *155*, 529–536. [CrossRef]
44. Leiter, A.; Veluswamy, R.R.; Wisnivesky, J.P. The global burden of lung cancer: Current status and future trends. *Nat. Rev. Clin. Oncol.* **2023**, *20*, 624–639. [CrossRef]
45. Lauri, H. High-resolution CT of the lungs: Indications and diagnosis. *Duodecim Laaketieteellinen Aikakauskirja* **2017**, *133*, 549–556.
46. 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**, *18*, e0273445. [CrossRef]
47. Binczyk, F.; Prazuch, W.; Bozek, P.; Polanska, J. Radiomics and artificial intelligence in lung cancer screening. *Transl. Lung Cancer Res.* **2021**, *10*, 1186. [CrossRef]
48. Rabbani, M.; Kanevsky, J.; Kafi, K.; Chandelier, F.; Giles, F.J. Role of artificial intelligence in the care of patients with nonsmall cell lung cancer. *Eur. J. Clin. Investig.* **2018**, *48*, e12901. [CrossRef]
49. Espinoza, J.L.; Dong, L.T. Artificial intelligence tools for refining lung cancer screening. *J. Clin. Med.* **2020**, *9*, 3860. [CrossRef]
50. Zhang, X.; Zhang, Y.; Zhang, G.; Qiu, X.; Tan, W.; Yin, X.; Liao, L. Deep Learning With Radiomics for Disease Diagnosis and Treatment: Challenges and Potential. *Front. Oncol.* **2022**, *12*, 773840. [CrossRef] [PubMed]
51. Braghetto, A.; Marturano, F.; Paiusco, M.; Baiesi, M.; Bettinelli, A. Radiomics and deep learning methods for the prediction of 2-year overall survival in LUNG1 dataset. *Sci. Rep.* **2022**, *12*, 14132. [CrossRef] [PubMed]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.