Artificial Intelligence for Lung Cancer Detection and Characterization on Computed Tomography: A Narrative Review of Validated Performance, Clinical Implementation, and the Road to Population-Level Impact

Harshatej Simhadri[1]

Received June 20, 2026
Accepted August 11, 2026
Electronic access September 30, 2026

Background: Lung cancer is one of the leading causes of cancer deaths worldwide, responsible for approximately 1.8 million deaths each year. Although low-dose CT (LDCT) screening can reduce lung cancer mortality by 20–33%, limited radiologist capacity and low screening participation hinders the early detection. This review explores use of Artificial Intelligence (AI) to support the early detection of lung cancer on CT, focusing on nodule detection, malignancy risk stratification, temporal growth analysis, molecular phenotype prediction and current U.S. regulatory landscape.
Methods: A comprehensive review of PubMed/MEDLINE, EMBASE, and IEEE Xplore was conducted for peer-reviewed studies, randomized trials and regulatory datasets published between January 2015 to August 2025, with foundational methodological references from before 2015. In curating literature for this review, priority was given to externally validated algorithms, prospective or randomized designs, and systematic reviews with meta-analysis.
Results: Deep Learning Algorithms detected nodule sensitivity of 86–98% on LDCT versus 68–76% for unassisted radiologists. In 2025, a meta-analysis of 315 studies showed pooled sensitivity and specificity of 0.86 and an SROC-AUC of 0.92 for lung cancer detection. Malignancy risk models have shown AUCs of 0.93–0.97 and have reduced unnecessary follow-ups with radiologists by ∼20–30%. Temporal AI models improved assessment of nodule growth and identified growth patterns associated with malignancy earlier than standard radiologist assessment in selected cohorts. Radiomic models predicted EGFR mutation status with AUCs of 0.88–0.91, supporting the potential of CT-based AI for molecular phenotyping. Among the CT based AI tools reviewed, Optellum Virtual Nodule Clinic and Riverain ClearRead CT had confirmed FDA 510(k) clearance for pulmonary nodule analysis.
Conclusions: Clinical validation of AI based CT analysis for Pulmonary nodule detection, malignancy risk stratification, temporal growth and molecular phenotype prediction is considered whereas Population level deployment will require rigorous external validation and prospective outcome studies.

Keywords: Artificial Intelligence, Deep Learning, Lung Cancer, Computed Tomography, Pulmonary Nodule, Screening, Radiomics
Abbreviations: AUC = Area under the ROC Curve, CNN = Convolutional Neural Network, CT = Computed Tomography, DL = Deep Learning, EGFR = Epidermal Growth Factor Receptor, FDA = Food and Drug Administration, LDCT = Low Dose Computed Tomography, LUNA16 = Lung Nodule Analysis 2016, LUNG-RADS = Lung CT Screening Reporting and Data System, NELSON = Nederlands-Leuvens Longkanker Screenings Onderzoek, NLST = National Lung Screening Trial, NSCLC = Non Small cell Lung Cancer, SAMD = Software as a Medical Device, SROC = Summary Receiver Operating Characteristic.

Introduction

Lung cancers accounts for about 238,00 new cases each year in the United States and has a high mortality rate than breast, colorectal and prostate cancers combined[1]. Also, the overall five-year survival rate remain low at 23% compared with 61% for patients diagnosed at Stage I, highlighting the importance of early detection[2]. A Low-dose CT (LDCT) lung cancer screening, validated by the National Lung Screening Trial (NLST) and the NELSON trial, has demonstrated relative reductions in lung cancer mortality of 20% and 24–33% respectively making it one of the most effective interventions currently available for reducing lung cancer mortal-

[1] *Westford Academy, Massachusetts, USA*

ity[3,4]. However, LDCT screening is limited by substantial radiologist workload, high false-positive rates (96.4% of positive NLST screens are non-malignant) and variability among radiologists in pulmonary nodule assessment[3]. Although standardised reporting through the Lung CT Screening Reporting and Data System (Lung-RADS) has improved consistency in nodule management but the size-based categories may not fully account for variations in malignancy risk associated with nodule characteristics and individual patient factors. Artificial Intelligence (AI), especially Deep Learning (DL)-based computer vision, offers a promising approach to overcome the limitations. DL algorithms can learn from large annotated CT datasets to detect nodules, estimate the risk of malignancy and follow them over time. Several studies report similar or better performance than radiologists[5–8]. These capabilities provide a dual benefit, increasing sensitivity for early-stage cancers that are not detected by current programmes, and reducing unnecessary follow-up procedures that place a burden on patients and incur healthcare costs. This review systematically provides the evidence base for AI in LDCT based lung cancer detection across four functional domains as below

(a) Automated nodule detection and false-positive reduction

(b) Risk stratification for malignancy and integration with Lung-RADS

(c) Temporal growth analysis and sub-threshold change detection

(d) Radiomic molecular phenotyping.

We further assess the performance of AI compared with radiologists, based on the current U.S. regulatory landscape, and the practical challenges that need to be overcome before these technologies can be used at the population level. Although many reviews have evaluated the performance of AI, most focus on algorithm development or clinical validation in isolation. This review validates performance, regulatory clearance and readiness for clinical implementation. To identify the significant gap in the evidence required by healthcare systems considering the integration of AI into LDCT screening programs.

Evidence Acquisition

A structured literature review was performed in PubMed/MEDLINE, EMBASE, and IEEE Xplore for publications from January 2015 to August 2025 using the following Medical Subject Heading (MeSH) and keyword combinations (artificial intelligence OR deep learning OR convolutional neural network OR radiomics) AND (lung cancer OR pulmonary neoplasm OR pulmonary nodule) AND (computed tomography OR CT screening OR low-dose CT). Supplementary searches include FDA 510(k) regulatory clearance records, Lung-RADS guidelines (ACR 2022), and NLST/NELSON trial publications. The current study prioritised externally validated algorithms, randomised or prospective designs, and systematic reviews with meta-analysis. This review follows a narrative framework. Search strategy and study reporting are informed by PRISMA 2020 principles to ensure transparency; however, this is a narrative review, and no formal protocol registration was undertaken so, no pooled meta-analysis was conducted by the authors.

Technical Foundations

Current AI systems for lung cancer detection often use convolutional neural networks (CNNs), which can learn important visual patterns directly from CT images without requiring manually selected features[9].

Several established architectures, including ResNet, DenseNet, and U-Net, have been adapted for three-dimensional CT imaging. These networks progressively learn from the CT data, starting with basic features such as intensity changes and nodule boundaries and moving toward more complex characteristics such as shape, density, spiculation, ground-glass opacity (GGO), part-solid morphology, and lobulation[10].

Vision transformers (ViTs), which emerged after 2020, use self-attention to analyse different regions of a CT scan and capture relationships across distant areas of the lungs that conventional convolutional layers may miss. Hybrid CNN-ViT models are increasingly being explored for both nodule detection and malignancy risk assessment[11].

AI-assisted lung nodule assessment can be viewed as a six-stage workflow, beginning with CT image acquisition and progressing through nodule detection, characterization, malignancy risk estimation, temporal analysis, and clinical decision support as shown in Figure 1.

1. Input LDCT Acquisition - Standardized high-quality low-dose CT (LDCT) acquisition provides the basis for robust AI analysis. Image reconstruction, preprocessing and standardized imaging protocols are some of the techniques to minimize the variations in the image quality and improve the consistency of the downstream AI performance across scanners and institutions.

2. Nodule Detection - Algorithms like YOLO, Faster-RCNN detect candidate nodules in a CT scan and localize them, giving bounding boxes and confidence scores.

3. Nodule Characterization - Segmentation networks, e.g., 3D U-Net, can delineate nodules at voxel level. Radiomic and deep learning methods extract features that describe nodule size, shape, texture, intensity, and location to characterize appearance and structure of nodules.

4. Malignancy Risk Estimation - Risk stratification models are designed to estimate the probability of malignancy from either the whole CT scan or a cropped nodule region of interest. Typically, the models output a continuous risk score that can be compared with biopsy or follow-up outcomes [5,12].

5. Temporal Analysis - Temporal analysis compares current CT scans with prior examinations to identify changes in nodule size, volume, and morphology over time. AI models can help quantify subtle growth patterns that may be difficult to recognize visually and may provide earlier indications of clinically significant growth.

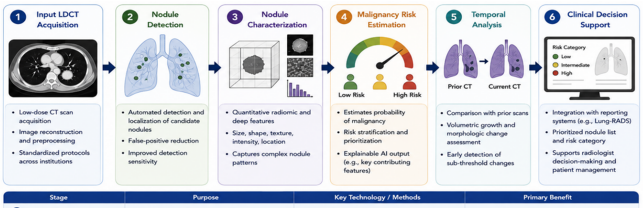
6. Clinical Decision Support - This aims to assist radiologists in prioritizing nodules, facilitating risk-based reporting and ensuring appropriate follow-up or management by incorporating AI findings with imaging and clinical data. Integrating with systems like Lung-RADS could help convert AI predictions into clinically actionable recommendations.

Training these AI systems requires large, diverse and well-annotated datasets. The Lung Image Database Consortium and Image Database Resource Initiative (LIDC-IDRI) is a database of 1,018 CT scans with expert annotations, and the National Lung Screening Trial (NLST) imaging archive is a database of more than 75,000 CT scans, which have been used as important datasets for model development and benchmarking [9,13]. Transfer learning from models pre-trained on ImageNet can reduce the amount of annotated data required for model development, while data augmentation techniques such as rotation, flipping and noise injection can help reduce overfitting [9,13].

AI-Assisted Pulmonary Nodule Detection

The LUNA16 Challenge established a benchmark for evaluating automated pulmonary nodule detection. Among 15 algorithms tested on 888 CT scans containing 1,186 nodules ≥ 3 mm, the best-performing single algorithm achieved 97.2% sensitivity at one false positive per scan, while an ensemble approach achieved more than 95% sensitivity at fewer than one false positive per scan using the Competition Performance Metric (CPM) [13]. A major challenge was reducing false positives, particularly those caused by structures such as blood vessels, airways, and pleural folds that can resemble nodules. This limitation led to the development of multi-view CNN architectures that use information from axial, coronal, and sagittal views to better distinguish true nodules from these normal anatomical structures [14].

Clinical-scale external validation of DL detection systems demonstrated consistent improvements over unassisted radiologists. A 2024 systematic review of 14 prospective and



Stage	Purpose	Key Technology / Methods	Primary Benefit
1 Input LDCT Acquisition	Obtain high-quality, standardized CT images	LDCT scanners, reconstruction, preprocessing	Consistent image quality for reliable analysis
2 Nodule Detection	Detect and localize candidate nodules	3D CNNs, object detection networks	Higher sensitivity and fewer false positives
3 Nodule Characterization	Extract features describing nodule properties	Radiomics, deep learning feature extraction	Better understanding of nodule behavior
4 Malignancy Risk Estimation	Estimate probability of malignancy	Machine learning / deep learning models	Risk-based prioritization of nodules
5 Temporal Analysis	Assess changes over time	Image registration, growth modeling	Earlier detection of clinically significant growth
6 Clinical Decision Support	Support reporting and clinical management	Decision support systems, guideline integration	Improved workflow and patient outcomes

Figure. 1 AI pipeline for early lung cancer detection on low-dose CT. Illustrating six sequential stages from image acquisition through clinical decision support. This figure is a conceptual schematic intended to summarise key workflow components; it does not present original quantitative performance data. Figure created by the author. AI=artificial intelligence. CNN=convolutional neural network. CT=computed tomography. LDCT=low-dose computed tomography. Lung-RADS=Lung CT Screening Reporting and Data System.

retrospective studies reported AI sensitivity of 86.0–98.1% versus radiologist sensitivity of 68–76% for nodule detection on LDCT [15]. The performance gap was largest for nodules smaller than 10 mm, precisely the group for which early intervention offers the greatest survival benefit [15,16]. Of particular clinical importance is AI performance on missed cancers. A retrospective analysis by Cho et al. applied DL-CAD to CT scans of NLST participants who were subsequently diagnosed with lung cancer, examining scans that radiologists had previously reported as negative or low-risk. The algorithm detected 95% of these subsequently diagnosed cancers using NLST criteria (70/74 cases) and 98% using Lung-RADS 2022 criteria (59/60 cases). Note that these cases represent false-negative reads on scans preceding diagnosis, which is distinct from interval cancers (tumours arising and presenting between scheduled screening rounds); the distinction is clinically important for interpreting the AI second-reader benefit. This finding quantifies the potential of an AI second-reader to reduce the most consequential category of false negatives in lung cancer screening programmes. Subsolid nodules—GGO and part-solid morphologies—represent a distinct challenge. They are more likely to be malignant per unit size, grow more slowly, and are substantially harder to detect reliably on standard windowing. DL systems trained specifically on subsolid morphology have demonstrated superior detection performance compared with solid-nodule models, with a published sensitivity of 91–94% for GGO nodules ≥ 5 mm [16]. Selected algorithms with training cohorts, external validation datasets, and key performance metrics are summarised in Table 1.

AUC=area under the receiver operating characteristic curve. CI=confidence interval. DLCST=Danish Lung Can-

Table 1 Selected Deep Learning Algorithms for Pulmonary Nodule Detection and Malignancy Risk Stratification on CT

Study (Year)	Algorithm / Model	Training Data	Validation Cohort	Task	Key Performance Metric
Ardila *et al.*, 2019 [3]	End-to-end 3D CNN	NLST (42,290 scans)	NLST holdout (6,716 scans)	Cancer detection	AUC 0.944; surpassed 6/6 radiologists (single CT)
Setio *et al.* LUNA16, 2017 [13]	Multi-view CNN ensemble	LIDC-IDRI (888 CTs)	LUNA16 challenge set	Nodule detection	Sensitivity 97.2% at 1 FP/scan; CPM 0.871
Venkadesh *et al.*, 2021 [12]	CNN malignancy risk	NLST (1,352 nodules)	DLCST (external validation: 883 nodules)	Malignancy probability	AUC 0.93 (vs. PanCan AUC 0.90, $p < 0.05$)
Mikhael/Sybil, 2023 [8]	Whole-CT transformer	MGH/NLST	NLST, MGH, CGMH (Taiwan)	1- to 6-yr cancer risk	1-yr AUC 0.92 (NLST); 0.94 (Taiwan); 6-yr C-index 0.75
Venkadesh *et al.*, 2023 [17]	Temporal dual-CT CNN	NLST nodule pairs	DLCST + MILD	Longitudinal malignancy risk	Significant improvement over single-CT model (both cohorts)
Huang *et al.*, 2019 [7]	3D DL risk prediction	PanCan cohort	NLST + PLCO validation	Follow-up risk prediction	1-yr AUC 0.968; 3-yr AUC 0.899

AUC=area under the receiver operating characteristic curve. CI=confidence interval. DLCST=Danish Lung Cancer Screening Trial. LDCT=low-dose computed tomography. NLST=National Lung Screening Trial. NPV=negative predictive value.

cer Screening Trial. LDCT=low-dose computed tomography. NLST=National Lung Screening Trial. NPV=negative predictive value.

Malignancy Risk Stratification and Lung-RADS Integration

Nodule detection is a necessary but insufficient precondition for effective screening. The critical clinical challenge is risk stratification: determining which detected nodules warrant immediate workup versus surveillance versus discharge. The Lung CT Screening Reporting and Data System (Lung-RADS 2022, American College of Radiology) provides a structured categorical framework based on nodule size, morphology, and temporal change, but its binary thresholds inherently sacrifice sensitivity at one extreme and specificity at the other [18]. AI-based malignancy risk models provide a continuous probability score that complements categorical Lung-RADS assignment. Venkadesh et al. trained a CNN on 1,352 NLST nodules and validated it on the Danish Lung Cancer Screening Trial (DLCST) cohort, achieving AUC 0.93 on external validation—significantly outperforming the validated PanCan model (AUC 0.90, $p < 0.05$) [12]. The model reduced unnecessary follow-up of benign nodules by 29% at equivalent sensitivity. The Sybil model (Massachusetts General Hospital/MIT) takes a different approach: it ingests an entire LDCT scan, requiring no radiologist-defined nodule annotation, and predicts the probability of lung cancer development within one to six years. Validated across three independent cohorts—the NLST (1-year AUC 0.92, 95% CI 0.88–0.95), Massachusetts General Hospital (AUC 0.86), and Chang Gung Memorial Hospital, Taiwan (AUC 0.94)—Sybil demonstrated consistent predictive performance across demographic and scanner heterogeneity. The six-year C-index of 0.75 suggests utility for identifying high-risk individuals who may benefit from escalated surveillance even when current Lung-RADS categorisation does not mandate immediate follow-up [8]. Integration of AI probability scores with Lung-RADS categories has been explicitly modelled. Tammemägi et al. modelled an AI + Lung-RADS hybrid management strategy in a simulated screening cohort of approximately 3,200 baseline participants [19]. Using AI probability scores to upgrade Lung-RADS 1/2 nodules with high malignancy probability (triggering earlier follow-up) and downgrade Lung-RADS 3/4 nodules with low probability (deferring follow-up) substantially reduced total follow-up CT examinations without a statistically significant reduction in cancer detection rate, supporting the feasibility of AI-assisted nodule management for improving programme efficiency [19]. The FDA-cleared Optellum Virtual Nodule Clinic (LCP-CNN) is the most rigorously validated commercial implementation of this approach. In a multi-reader study submitted to support 510(k) clearance (K203037, March 2021), the AI tool improved clinician AUC by a mean of 6.85 percentage points (range 2.4–12.1; $p < 0.001$) across all participating radiologists and pulmonologists, with consistent benefit regardless of reader experience level [20]. Key malignancy risk stratification models evaluated in this review, including Sybil, the Venkadesh CNN, and the Optellum LCP-CNN, are listed in Table 1.

Temporal AI Analysis and Sub-threshold Growth Detection

The most underutilised capability of AI in lung cancer screening is longitudinal temporal analysis. Radiologists reviewing annual LDCT scans assess nodule change qualitatively or using caliper-based diameter measurements, a method with inter-reader coefficient of variation (CV) of 14–18% for subcentimetre nodules [21]. AI volumetric measurement achieves CV of 2–4%, approximately five-fold more reproducible [21,22]. This precision advantage translates directly into earlier detection of growth. Venkadesh et al. demonstrated that incorporating a prior year CT as a secondary input, the malignancy

risk CNN significantly improved performance over the single-CT model across both the DLCST and MILD validation cohorts[17]. The temporal model can detect subtle changes in density, volume, and morphology over time that may be too small or complex to be reliably identified through visual assessment alone. For 4–6 mm nodules, generally classified as Lung-RADS 3 and managed with follow-up rather than immediate intervention, temporal AI models may help identify which nodules are likely to show malignant growth within three or six months and which are likely to remain stable, supporting more targeted surveillance[7,17]. Deep learning growth prediction models for subsolid nodules extend this capability further. CNNs trained on serial CT pairs accurately classified whether part-solid nodules would demonstrate clinically significant growth within 12 months, outperforming radiologist assessment based on Lung-RADS criteria alone[23]. Separately, a volumetric AI system applied to ultra-LDCT screening data achieved a negative predictive value exceeding 99% for malignancy at baseline, identifying a large proportion of participants for whom annual follow-up could safely be extended to biennial surveillance, reducing overall programme cost and radiation exposure[24]. The principle that temporal AI can advance the diagnostic window is supported by analogous findings in breast cancer screening, where retrospective deep learning analysis identified malignant lesions on prior mammograms preceding clinical diagnosis by one to two years[25]. This breast cancer analogy is presented explicitly as evidence from a different modality and tumour type, not as direct lung cancer data. In the lung cancer context, NLST-based retrospective analyses demonstrate that AI can identify actionable CT findings on scans preceding clinical diagnosis[26], suggesting that prospective AI-assisted temporal surveillance could facilitate meaningful stage downshift for a subset of screen-detectable cancers. These findings collectively support the principle that AI-based temporal surveillance, if deployed prospectively, could advance the diagnostic window by one or more screening cycles as shown in Figure 2.

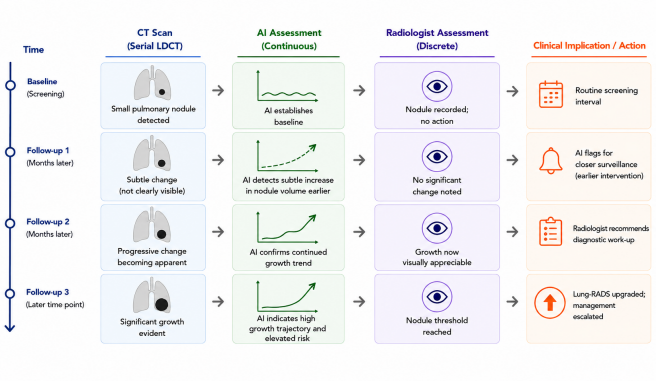


Figure. 2 Conceptual illustration of AI-assisted longitudinal nodule assessment on serial LDCT across four time points (Baseline through Follow-up 3). AI volumetric assessment detects subtle growth earlier than discrete radiologist visual assessment, potentially enabling earlier clinical intervention. No quantitative data are presented. AI = artificial intelligence; LDCT = low-dose computed tomography; Lung-RADS = Lung CT Screening Reporting and Data System.

Radiomic Molecular Phenotyping

Radiomics extends AI-based CT analysis beyond morphological characterisation to the extraction of high-throughput quantitative imaging features that encode information invisible to the human observer. A single CT nodule region-of-interest yields 150–1,000+ quantitative features encompassing first-order intensity statistics, second-order texture matrices (grey-level co-occurrence matrix [GLCM], grey-level run-length matrix [GLRLM]), higher-order wavelet transforms, and shape descriptors[27]. In non-small-cell lung cancer (NSCLC), radiomic signatures have demonstrated consistent value for molecular phenotype prediction. A 2024 multicentre study using deep learning–radiomic integration predicted EGFR mutation status with AUC 0.88–0.91 from standard-of-care CT[28], offering a non-invasive alternative to tissue biopsy for patients in whom tissue sampling is high-risk or insufficient. KRAS mutation status prediction achieved AUC 0.82–0.94 following scanner harmonisation[29]. PD-L1 expression level—a key immunotherapy eligibility biomarker—was predicted with AUC 0.81, providing a potential imaging surrogate for checkpoint inhibitor response stratification[30]. For overall survival prediction, a deep learning–radiomics model validated on contrast-enhanced CT from a multicentre NSCLC cohort achieved C-index 0.74–0.75 and AUC of 0.73–0.76 at 8, 12, and 24 months post-diagnosis[31]. A 2022 meta-analysis of 19 studies reported pooled AUC of 0.83 for lung cancer diagnosis and staging using combined radiomics and deep learning approaches[32]. The clinical translational pathway for radiomics-based molecular phenotyping is clearer than for purely detection-oriented AI: it addresses a specific unmet need (non-invasive molecular characterisation), has a definable comparator (biopsy), and a defined clinical decision point (treatment selection). For patients with incidentally detected or screen-detected nodules awaiting biopsy, a CT-derived radiomic probability of EGFR mutation could support earlier initiation of targeted therapy or guide biopsy site selection in heterogeneous tumours as shown in Figure 3.

AI Versus Radiologist Performance

Systematic evaluation of AI versus radiologist performance on LDCT interpretation has produced a consistent and nuanced picture. The foundational observation—established by Ardila

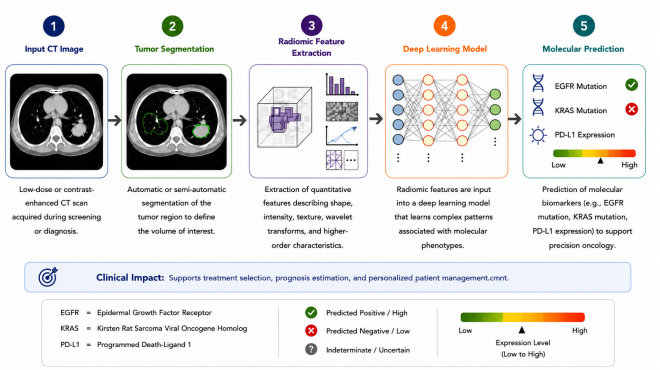


Figure. 3 Conceptual five-stage workflow of radiomics-based molecular phenotype prediction from CT imaging, from tumour segmentation through molecular biomarker probability estimation for EGFR, KRAS, and PD-L1. No patient-specific data or quantitative performance results are presented. CT = computed tomography; EGFR = epidermal growth factor receptor; KRAS = Kirsten Rat Sarcoma Viral Oncogene Homolog; PD-L1 = Programmed Death-Ligand 1

et al. on 6,716 NLST CT scans—is that a single DL model surpassed all six participating radiologists in cancer detection when only a single CT scan was available (AUC 0.944 vs. mean radiologist AUC 0.93) and achieved AUC 0.959 in the full analysis[5]. When prior CT scans were provided to radiologists (standard clinical practice), the gap narrowed, but AI performance remained non-inferior. A 2021 observer study by Yoo et al. enrolled 11 radiologists from diverse practice settings and evaluated them alongside 10 competing DL algorithms on the same LDCT dataset[6]. Top algorithms (AUC 0.900–0.902) were statistically non-inferior to the radiologist group (mean AUC 0.917), demonstrating that no algorithm individually outperformed the best radiologists, but the best AI equalled the best human readers—a meaningful threshold for a screening-adjunct application. For nodules ¡10 mm, AI sensitivity exceeded mean radiologist sensitivity by 18–24 percentage points[15]. The most clinically relevant performance metric is not standalone AI accuracy but AI-assisted radiologist performance. A 2024 systematic review reported that with AI assistance, radiologist sensitivity for nodule detection increased by 24.3 percentage points without a meaningful increase in false-positive rate[15]. This represents the core value proposition: AI as a consistent, fatigue-independent second reader that adds most value precisely when radiologist performance is most vulnerable—during high-volume sessions, after-hours reads, and in settings without subspecialty thoracic radiology coverage. However, AI performance is not uniformly superior. In complex scenarios—including unusual nodule morphologies, post-treatment changes, and CT studies with dense parenchymal disease—AI systems trained predominantly on screening-eligible populations show decreased performance. Furthermore, the inter-platform variability among commercial AI tools is substantial: in head-to-head comparisons, sensitivity ranged from 77% to 98% across evaluated products, underscoring the importance of independent validation before clinical adoption[15,33]. A summary of pooled diagnostic performance estimates from four systematic reviews and meta-analyses are presented in Table 2.

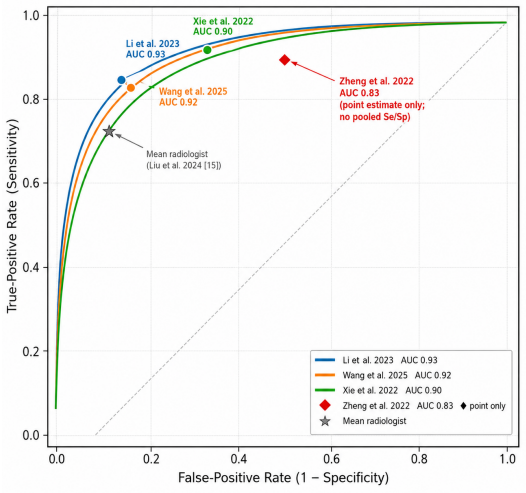


Figure. 4 Illustrative comparison of pooled diagnostic performance from four meta-analyses of AI-based lung cancer detection on CT (2022–2025). SROC curves are approximations based on published pooled estimates; Zheng et al. reported only AUC and therefore is shown as a point estimate. Overlap among primary studies should be considered when interpreting the AUC range (0.83–0.93). Author-created figure; no new experimental data.

US Regulatory Landscape and FDA-Cleared Tools

The US FDA has applied an adaptive regulatory framework to AI-based lung CT tools. Most are classified as Software as a Medical Device (SaMD) and reviewed under the 510(k) pathway as Class II devices, requiring demonstration of substantial equivalence to a predicate device. As of 2024, five AI tools for lung CT nodule analysis or related imaging applications are described in this review (Table 3). Of these, Optellum Virtual Nodule Clinic (K203037) and Riverain ClearRead CT (K161468) have confirmed FDA 510(k) clearance for CT-based pulmonary nodule analysis. Lunit INSIGHT CXR has FDA clearance for chest X-ray triage (not CT) and is included in Table 3 for completeness with explicit CXR designation. LungQ 4 has a 2023 listing but its 510(k) number remains unconfirmed. Aidence Veye Lung Nodules holds European CE marking with FDA clearance pending. Optellum Virtual Nod-

Table 2 Systematic Reviews & Meta-Analyses of AI-Based Lung Cancer Detection & Classification on CT (2022–2025)

Study (Year)	Number of Studies	Task	Pooled Sensitivity (95% CI)	Pooled Specificity (95% CI)	SROC-AUC (95% CI)
Xie *et al.*, 2022 [32]	6 studies	Diagnosis / cancer detection	0.93 (0.85–0.98)	0.68 (0.49–0.84)	0.90 (0.86–0.92)
Li *et al.*, 2023 [34]	Multiple cohorts	AI-aided diagnosis	0.87 (0.82–0.90)	0.87 (0.82–0.91)	0.93 (0.91–0.95)
Zheng *et al.*, 2022 [35]	19 studies (radiomics + DL)	Staging and diagnosis	—	—	0.83 (0.78–0.88)
npj Precis Oncol, 2025 [36]	315 studies total; diagnosis subset: 209 studies (pooled values apply to this subset only)	Diagnosis and prognosis	0.86 (0.84–0.87)	0.86 (0.84–0.87)	0.92 (0.90–0.94)
Liu *et al.* (AI vs. radiologist), 2024 [15]	14 studies	Standalone AI vs. radiologists	AI: 86.0–98.1% vs. radiologist: 68–76%	AI: 77.5–87% vs. radiologist: 87–91.7%	AI superior accuracy in most settings

AUC=Area Under the receiver operating characteristic Curve. CaI=Confidence Interval. SROC=Summary Receiver Operating Characteristic. Pooled estimates calculated by bivariate random-effects model in each source meta-analysis.

ule Clinic (K203037, cleared March 2021) is the first FDA-cleared AI clinical decision support tool specifically for early lung cancer diagnosis, not merely detection. Its clearance was supported by a prospective multi-reader study demonstrating statistically significant improvement in clinician diagnostic accuracy [20,33]. Riverain ClearRead CT (cleared 2016) was among the earliest FDA-cleared AI tools for CT nodule detection, demonstrating a 29% reduction in missed actionable nodules in a multi-reader multi-case study [37]. Lunit INSIGHT CXR (K213696, cleared November 2021) is a chest X-ray (CXR) triage tool and not a CT product; its FDA clearance and published performance metrics pertain exclusively to chest X-ray interpretation (Project AIR study, Radboud University Medical Center) [38]. Its AUC of 0.93 for lung nodule detection reflects CXR performance only and must not be compared with CT-based tool performance. Lunit INSIGHT CXR is clearly identified as a CXR application in Table 3 and is excluded from CT performance comparisons. The FDA has published a predetermined change control plan (PCCP) framework enabling AI developers to prospectively define performance boundaries within which algorithm updates do not require new submissions. This approach is specifically relevant to temporal AI systems that learn from accumulating longitudinal data and allow performance improvement while maintaining regulatory accountability. The FDA's action plan for AI/ML-based SaMD 2021 further establishes principles of transparency, real-world performance monitoring, and algorithmic accountability that frame the current regulatory environment [39].

Limitations and Implementation Challenges

While the impressive benchmark performance of AI for lung cancer screening is promising, there are several limitations that preclude its immediate population-level deployment. Dataset bias and generalisability represent the primary challenge. The majority of high-performing algorithms were trained on NLST or LIDC-IDRI data derived predominantly from high-income, predominantly White American populations using modern CT scanner generations. Performance degrades when algorithms are applied to data from different scanner manufacturers, reconstruction kernels, and slice thicknesses [9,40]. A DL algorithm achieving AUC 0.93 on a US screening cohort may perform substantially worse in a community setting with heterogeneous scanner protocols. It is important to note that the headline performance figures in Tables 1 and 2 derive predominantly from single-study or vendor-sponsored validations, and benchmark results from standard test datasets (LUNA16, NLST holdouts) tend to overstate real-world performance in heterogeneous clinical populations. Systematic external validation across diverse populations including racially and ethnically diverse US cohorts is a prerequisite for equitable deployment.

Prospective outcome data remain limited. Most published validation studies use nodule detection or malignancy probability as surrogate outcomes rather than clinical endpoints (stage at detection, treatment initiation delay, lung cancer mortality). The NLST-AI sub-study and ongoing National Cancer Institute-funded prospective implementation trials will provide the first randomised evidence on whether AI-assisted screening reduces mortality compared with standard screening data currently unavailable [41]. Explainability and the black-box problem persist for deployment contexts requiring regulatory and clinical accountability. Gradient-weighted class activation mapping (Grad-CAM) and attention-map visualisation techniques provide post-hoc spatial explanations, but the fidelity of these explanations to the model's actual decision pathway remains contested, and there is no established standard for explainability validation in the FDA review process [40]. Automation bias over-reliance on AI outputs with reduced indepen-

Table 3 AI-Based Lung CT Analysis Tools with US FDA Clearance (as of 2024)

Product	Company	FDA 510(k)	Clearance Year	Application	Key Performance Data
Virtual Nodule Clinic (LCP-CNN)	Optellum	K203037	2021	Malignancy risk stratification and clinical decision support	Mean AUC improvement +6.85 points across all clinician readers ($p < 0.001$); first FDA-cleared AI for early lung cancer diagnosis (20,33)
ClearRead CT	Riverain Technologies	K161468	2016	Nodule detection and concurrent reading	29% reduction in missed actionable nodules; detection of small nodules improved from 13% to 24.2% (34)
INSIGHT CXR (triage)	Lunit	K213696	2021	Lung nodule triage on chest X-ray (CXR) ONLY — NOT a CT product. FDA clearance K213696 is for CXR triage, not CT analysis.	Nodule detection on CHEST X-RAY is AUC 0.93 (Project AIR study, Radboud UMC) — CXR performance only; not applicable to CT comparisonead-to-head; sensitivity 89% vs. mean radiologist 81% (35)
LungQ 4	Thirona	Not publicly confirmed	2023	Quantitative nodule volumetry and density classification	Automated volume measurement for Lung-RADS and NELSON protocol compliance
Veye Lung Nodules	Aidence	CE-marked; FDA pending	—	Nodule detection and volumetric follow-up	AUC 0.900 in DSB2017/RSNA observer study (6); real-world NPV >99% at ultra-LDCT (25)

510(k)=premarket notification pathway; CAD=computer-aided detection; FDA=Food and Drug Administration; FP=false positive; SaMD=software as a medical device.

dent assessment is a recognised risk in clinical AI applications Gaube et al. demonstrated that clinicians exhibit susceptibility to AI-generated decision aids, with AI recommendations influencing diagnostic decisions regardless of their correctness[41]. Primary prospective data quantifying automation bias specifically in CT lung cancer screening workflows remains limited, and dedicated human-factors research in this context is needed. Training in AI-specific clinical decision calibration is not yet systematically incorporated into radiology residency curricula.

Future Directions

Foundation models and multimodal AI represent the next generation of capability. Large vision–language models trained on hundreds of thousands of CT-report pairs can generate structured radiology reports from raw CT data with acceptable accuracy for common pathologies. Multimodal AI systems integrating CT imaging features with electronic health record data, smoking history, spirometry, and liquid biopsy biomarkers outperform any single-modality model for cancer risk stratification, and several such systems are in clinical validation[8]. Federated learning offers a technically feasible pathway to training AI models across distributed institutional datasets without sharing patient-level data, directly addressing the dataset diversity problem[40]. Consortia such as the National Cancer Institute's Imaging Data Commons provide open, harmonised CT archives that can support federated model development at scale.

AI for incidental nodule management represents a large underserved clinical need. Over 1.5 million incidental pulmonary nodules are detected annually on non-screening CT studies in the United States. Applying validated LDCT-trained AI to this population with appropriate re-calibration for the different prevalence of malignancy in non-screening populations could standardise management and reduce the substantial variation in follow-up practice currently observed across institutions. Ultra-low-dose CT protocols (effective dose ¡0.1 mSv) made practically feasible by AI-driven noise reduction algorithms (deep learning image reconstruction, DLIR) may enable annual screening at significantly reduced radiation burden, improving the risk–benefit ratio for younger high-risk individuals and facilitating extension of screening eligibility criteria to populations currently excluded from USPSTF recommendations.

Conclusions

AI-based LDCT analysis has demonstrated clinically validated performance in the detection of pulmonary nodules, malignancy risk stratification, temporal growth assessment, and molecular phenotype prediction, with several systems showing performance comparable to radiologists. However, high diagnostic performance in itself does not demonstrate reduction in lung cancer mortality. Widespread implementation will require prospective outcome studies, external validation in heterogeneous US populations and clinical settings, and safe incorporation into clinical workflows. Equitable deployment will be just as critical to ensure that the populations most affected by lung cancer, in particular underserved and underscreened populations, can access the potential advantages of AI-assisted early detection.

References

1 Siegel, R. L., Miller, K. D., Wagle, N. S. Jemal, A. Cancer statistics, 2023. CA Cancer J Clin. Vol. 73, pg. 17–48, 2023.

2 Howlader, N., Noone, A. M., Krapcho, M. et al. SEER cancer statistics review, 1975–2019. National Cancer Institute, Bethesda, MD. 2022. Available at: [https://seer.cancer.gov/csr/1975_2019/.

3 Aberle, D. R., Adams, A. M., Berg, C. D. et al. Reduced lung-cancer mortality with low-dose computed tomographic screening. N Engl J Med. Vol. 365, pg. 395–409, 2011, [https://doi.org/10.1056/NEJMoa1102873.

4 de Koning, H. J., van der Aalst, C. M., de Jong, P. A. et al. Reduced lung-cancer mortality with volume CT screening in a randomized trial. N Engl J Med. Vol. 382, pg. 503–513, 2020, [https://doi.org/10.1056/NEJMoa1911793.

5 Ardila, D., Kiraly, A. P., Bharadwaj, S. et al. End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. Nat Med. Vol. 25, pg. 954–961, 2019, [https://doi.org/10.1038/s41591-019-0447-x.

6 Yoo, H., Gichoya, J. W., Woo, J. H. et al. Deep learning for lung cancer detection on screening CT scans: results of a large-scale public competition and an observer study with 11 radiologists. Radiol Artif Intell. Vol. 3, pg. e210027, 2021, [https://doi.org/10.1148/ryai.2021210027.

7 Huang, P., Lin, C. T., Li, Y. et al. 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. Vol. 1, pg. e353–e362, 2019, [https://doi.org/10.1016/S2589-7500(19)30159-1.

8 Mikhael, P. G., Wohlwend, J., Yala, A. et al. Sybil: a validated deep learning model to predict future lung cancer risk from a single low-dose chest computed tomography. J Clin Oncol. Vol. 41, pg. 2191–2200, 2023, [https://doi.org/10.1200/JCO.22.01345.

9 Litjens, G., Kooi, T., Bejnordi, B. E. et al. A survey on deep learning in medical image analysis. Med Image Anal. Vol. 42, pg. 60–88, 2017, [https://doi.org/10.1016/j.media.2017.07.005.

10 LeCun, Y., Bengio, Y. Hinton, G. Deep learning. Nature. Vol. 521, pg. 436–444, 2015, [https://doi.org/10.1038/nature14539.

11 Dosovitskiy, A., Beyer, L., Kolesnikov, A. et al. An image is worth 16×16 words: transformers for image recognition at scale. In: International Conference on Learning Representations (ICLR). 2021. Available at: [https://openreview.net/forum?id=YicbFdNTTy.

12 Venkadesh, K. V., Setio, A. A. A., Schreuder, A. et al. Deep learning for malignancy risk estimation of pulmonary nodules detected at low-dose screening CT. Radiology. Vol. 300, pg. 438–447, 2021, [https://doi.org/10.1148/radiol.2021204433.

13 Setio, A. A. A., Traverso, A., de Bel, T. et al. Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in CT images: the LUNA16 challenge. Med Image Anal. Vol. 42, pg. 1–13, 2017, [https://doi.org/10.1016/j.media.2017.06.015.

14 Setio, A. A. A., Ciompi, F., Litjens, G. et al. Pulmonary nodule detection in CT images: false positive reduction using multi-view convolutional networks. IEEE Trans Med Imaging. Vol. 35, pg. 1160–1169, 2016, [https://doi.org/10.1109/TMI.2016.2536809.

15 Liu, M. C., Hsieh, M. H., Hsu, Y. H. et al. Standalone deep learning versus experts for diagnosis of lung cancer on chest computed tomography: a systematic review. Eur Radiol. Vol. 34, pg. 3860–3872, 2024, [https://doi.org/10.1007/s00330-023-10418-w.

16 Schreuder, A., Jacobs, C., van Ginneken, B. Prokop, M. Artificial intelligence for detection and characterization of pulmonary nodules in lung cancer CT screening: ready for practice? Transl Lung Cancer Res. Vol. 10, pg. 2378–2388, 2021, [https://doi.org/10.21037/tlcr-20-1123.

17 Venkadesh, K. V., Schreuder, A., Scholten, E. T. et al. Prior CT improves deep learning for malignancy risk estimation of screening-detected pulmonary nodules. Radiology. Vol. 308, pg. e223308, 2023, [https://doi.org/10.1148/radiol.223308.

18 Lung-RADS version 2022. American College of Radiology. 2025. Available at: [https://www.acr.org/Clinical-Resources/Reporting-and-Data-Systems/Lung-Rads.

19 Tammemägi, M. C., Ruparel, M., Tremblay, A. et al. Development and cost analysis of a lung nodule management strategy combining artificial intelligence and Lung-RADS for baseline lung cancer screening. Chest. Vol. 160, pg. 214–225, 2021, [https://doi.org/10.1016/j.chest.2021.01.046.

20 Optellum Ltd. Clinical utility of an artificial intelligence radiomics-based tool for risk stratification of pulmonary nodules. [online ahead of print]. Thorax. 2024, [https://doi.org/10.1136/thorax-2023-220488.

21 Zhao, B., James, L. P., Moskowitz, C. S. et al. Evaluating variability in tumor measurements from same-day repeat CT scans of patients with non-small cell lung cancer. Radiology. Vol. 252, pg. 263–272, 2009, [https://doi.org/10.1148/radiol.2522081593.

22 de Hoop, B., Gietema, H., van Ginneken, B., Zanen, P., Groen, H. Prokop, M. A comparison of six software packages for evaluation of solid lung nodules using semi-automated volumetry: what is the minimum increase in size to detect growth in repeated CT examinations. Eur Radiol. Vol. 19, pg. 800–808, 2009, [https://doi.org/10.1007/s00330-008-1229-x.

23 Peng, H., Chan, H. P., Hadjiiski, L. M. et al. Deep learning growth prediction for sub-solid pulmonary nodules on CT images. Front Oncol. Vol. 12, pg. 1002809, 2022, [https://doi.org/10.3389/fonc.2022.1002809.

24 Chamberlin, J., Kocher, M. R., Waltz, 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. Vol. 19, pg. 55, 2021, [https://doi.org/10.1186/s12916-021-01928-3.

25 Buda, M., Saha, A., Walsh, R. et al. A deep learning model for detecting malignant lesions in the breast using prior mammograms. Nat Commun. Vol. 12, pg. 4330, 2021, [https://doi.org/10.1038/s41467-021-24544-0.

26 Ardila, D., Kiraly, A. P., Bharadwaj, S. et al. End-to-end lung cancer screening: deep learning-based detection of missed lung cancers in the NLST. Radiology. Vol. 298, pg. 224–232, 2021, [https://doi.org/10.1148/radiol.2020202326.

27 Gillies, R. J., Kinahan, P. E. Hricak, H. Radiomics: images are more than pictures, they are data. Radiology. Vol. 278, pg. 563–577, 2016, [https://doi.org/10.1148/radiol.2015151169.

28 Zhang, L., Zhang, Z., Zhao, Y. et al. Deep learning-radiomics integrated noninvasive detection of epidermal growth factor receptor mutations in non-small cell lung cancer patients. Sci Rep. Vol. 14, pg. 2167, 2024, [https://doi.org/10.1038/s41598-024-52667-3.

29 Kirienko, M., Sollini, M., Corbetta, M. et al. Impact of feature harmonization on radiogenomics analysis: prediction of EGFR and KRAS mutations from non-small cell lung cancer PET/CT images. Comput Biol Med. Vol. 142, pg. 105230, 2022, [https://doi.org/10.1016/j.compbiomed.2021.105230.

30 Tian, P., He, B., Mu, W. et al. Assessing PD-L1 expression in non-small cell lung cancer and predicting responses to immune checkpoint inhibitors using deep learning on computed tomography images. Theranostics. Vol. 11, pg. 2098–2107, 2021, [https://doi.org/10.7150/thno.48027.

31 Liang, W., Zhang, L., Fang, M. et al. Radiomics-based deep learning prediction of overall survival in non-small-cell lung cancer using contrast-enhanced computed tomography. Cancers. Vol. 14, pg. 3798, 2022, [https://doi.org/10.3390/cancers14153798.

32 Xie, Y., Meng, W. Y., Li, R. Z. et al. Deep learning algorithms for diagno-

sis of lung cancer: a systematic review and meta-analysis. Cancers. Vol. 14, pg. 3914, 2022, [https://doi.org/10.3390/cancers14163914.

33 Chetan, M. R., Dowson, N., Price, N. W. et al. Developing an understanding of artificial intelligence lung nodule risk prediction using insights from the Brock model. Eur Radiol. Vol. 32, pg. 5437–5447, 2022, [https://doi.org/10.1007/s00330-022-08598-8.

34 Li, Y., Zhang, Z., Dai, C. et al. The value of artificial intelligence in the diagnosis of lung cancer: a systematic review and meta-analysis. PLoS One. Vol. 18, pg. e0282970, 2023, [https://doi.org/10.1371/journal.pone.0282970.

35 Zheng, X., He, B., Hu, Y. et al. Diagnostic accuracy of deep learning and radiomics in lung cancer staging: a systematic review and meta-analysis. Front Public Health. Vol. 10, pg. 938113, 2022, [https://doi.org/10.3389/fpubh.2022.938113.

36 Wang, X., Li, W., Li, Z. et al. Systematic review and meta-analysis of artificial intelligence for image-based lung cancer classification and prognostic evaluation. NPJ Precis Oncol. Vol. 9, pg. 300, 2025, [https://doi.org/10.1038/s41698-025-01095-1.

37 Riverain Technologies. ClearRead CT: FDA 510(k) clearance documentation. US FDA database, K161468. 2016.

38 van Leeuwen, K. G., Schalekamp, S., Rutten, M. J. C. M. et al.; for the Project AIR Working Group. Comparison of commercial AI software performance for radiograph lung nodule detection and bone age prediction. Radiology. Vol. 310, pg. e230981, 2024, [https://doi.org/10.1148/radiol.230981.

39 US Food and Drug Administration. Artificial intelligence and machine learning (AI/ML)-based software as a medical device (SaMD) action plan. Silver Spring, MD: FDA. 2021. Available at: [https://www.fda.gov/media/145022/download.

40 Hosny, A., Parmar, C., Quackenbush, J., Schwartz, L. H. Aerts, H. J. W. L. Artificial intelligence in radiology. Nat Rev Cancer. Vol. 18, pg. 500–510, 2018, [https://doi.org/10.1038/s41568-018-0016-5.

41 Gaube, S., Suresh, H., Raue, M., Merali, A., Pronovost, P. J., Weibel, M. et al. Do as AI say: susceptibility in deployment of clinical decision-aids. NPJ Digit Med. Vol. 4, pg. 31, 2021, [https://doi.org/10.1038/s41746-021-00385-9.