세브란스 SEVERANCE | 용인세브란스병원

2026 동계 분자폐암연구회 임상연구 워크숍

The Adaptation of Artificial Intelligence in Lung Cancer Clinical Trials

Eun Hye Lee

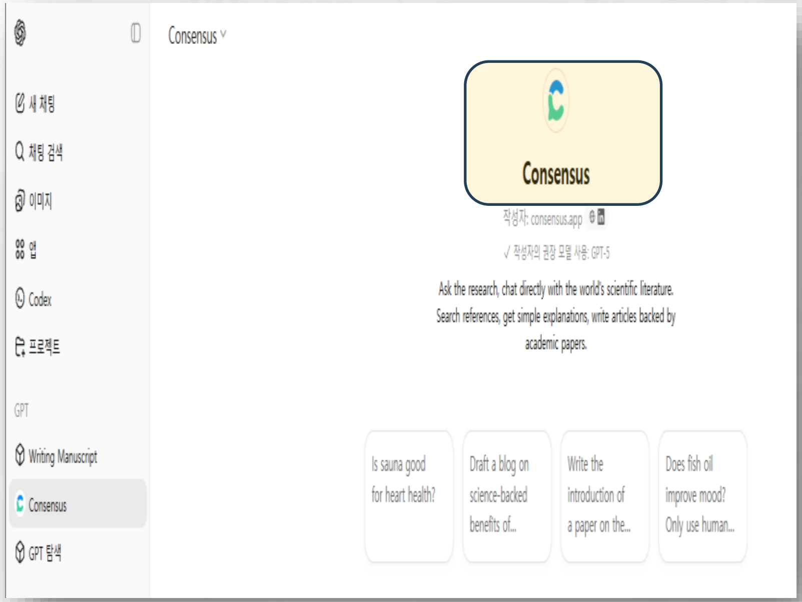
Yongin Severance Hospital, Yonsei University College of Medicine

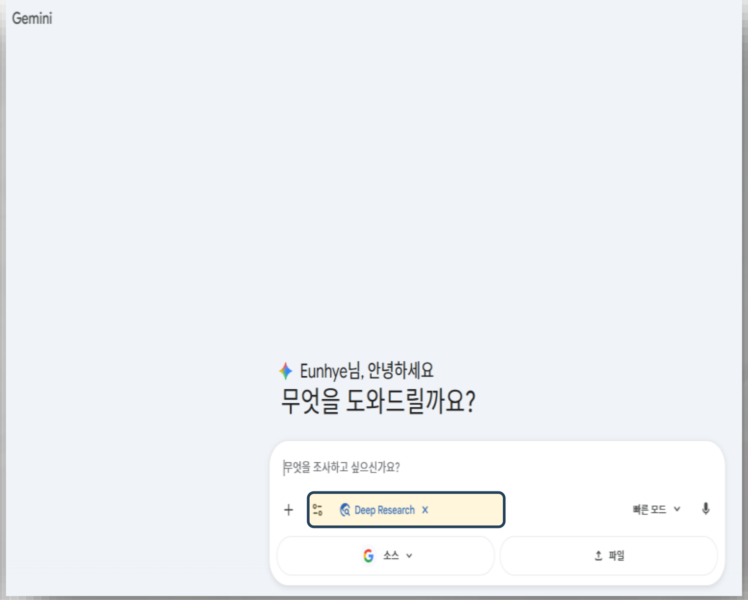
Acknowledgement

AI

Generative artificial intelligence tools (e.g., ChatGPT 5.2, Gemini 3.0, Genspark AI) were used to assist in organizing content and drafting preliminary slide materials

The author takes full responsibility for the accuracy, interpretation, and final content of this presentation





AI

Agenda

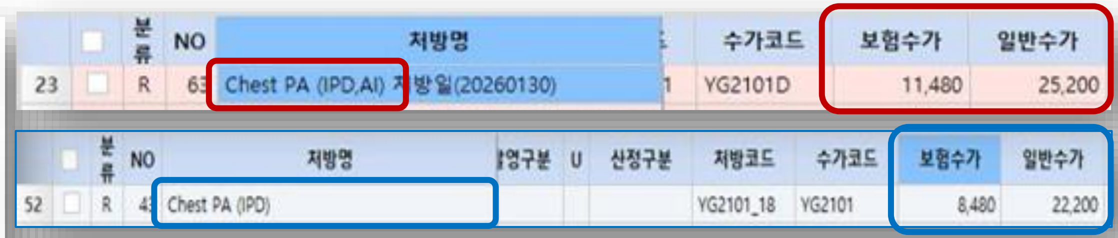
Overview of Topics

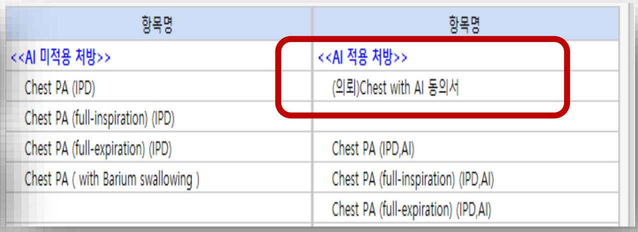
Current AI Solution at Our institution

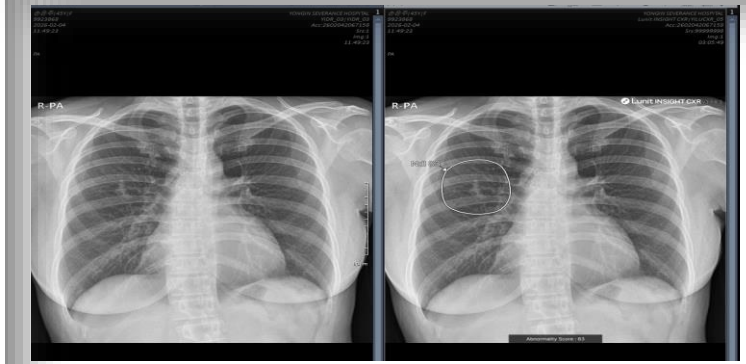
Table 1. Current State of Digital Solutions of Yongin Severance Hospital

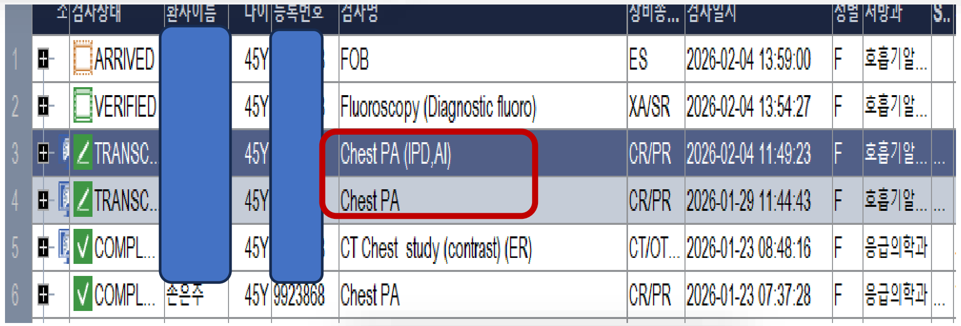
Target	Digital solution
Patient safety	- Integration and response space - Real-time location system - AI facial recognition solution for preventing patient swap - Analysis solution for infection epidemiology
Patient service	- Intelligent mobile application for patients - Bedside monitor - Prescription video in animations
Treatment efficiency	- Y-Talk (messenger application for cooperative treatment of doctors) - Automatic measuring kiosk for height/weight/blood pressure - Smart ID card for beds - 5G MEC infrastructure construction
Treatment accuracy	- AI voice recognition documenting system - AI decision-making solution

AI, artificial intelligence; MEC, mobile edge cloud









Current State and Strategy for Establishing a Digitally Innovative Hospital; Yonsei Med J 2020 Aug;61(8):647-651

AI in Lung Cancer Now

Introduction

Retrospective to Prospective

Moving from initial algorithm development on historical datasets to rigorous prospective validation in Randomized Controlled Trials (RCTs). Verifying real-world clinical utility beyond theoretical performance.

Full Care Continuum

Deep learning models now span the entire patient journey: from early detection via Radiomics (screening) to diagnosis, treatment response prediction (Pathomics), and personalized therapy selection.

Maturation of Deep Learning

Advanced architectures (CNNs, Transformers) demonstrating superior performance on complex medical imaging (CT, CXR) and digital pathology data compared to traditional machine learning methods.

Regulatory Frameworks

Emergence of standardized reporting guidelines like CONSORT-AI ensuring transparency, reproducibility, and rigorous evaluation of AI interventions in clinical settings.

Evidence Landscape

Radiomics (Imaging)

Quantitative extraction of features from medical images (LDCT, CXR) to uncover patterns invisible to the human eye. Focus on risk stratification and early detection.

Key Studies: ①Sybil, ②AI-CXR RCT, ③CTRS

Pathomics (Tissue)

Deep learning analysis of digitized whole-slide images (WSI) for automated grading, spatial phenotyping, and treatment response prediction.

Key Studies: ④ANORAK, ⑤Deep-IO

Multimodal Integration

Synthesizing radiologic, pathologic, and genomic data for holistic patient profiling and precision treatment selection.

Key Study: ⑥ A-STEP

AI Integration Across Care Continuum

Radiomics

- Early Detection
- Risk Prediction
- Nodule Classification

Pathomics

- Auto-Grading
- Spatial Analysis
- ICI Response

Multimodal Integration

Genomics + Clinical Data

Ongoing Prospective Validation

Section I

Radiomics

From risk prediction to treatment response biomarkers in CT/CXR

Radiomics Overview

Radiomics Overviews

AI

Quantitative Imaging Features

Radiomics extracts high-dimensional data from LDCT and Chest X-rays (CXR) beyond human visual perception. It transforms medical images into mineable data, quantifying tumor heterogeneity, texture, and shape.

Risk Stratification

AI models like Sybil predict future lung cancer risk (1-6 years) from a single scan, potentially expanding screening eligibility to non-smokers or lower-risk groups currently excluded by guidelines.

Detection & Diagnosis

AI-assisted detection significantly improves sensitivity and reduces reading time in screening populations. Validated in RCTs (e.g., Nam et al., Radiology 2023) showing improved nodule detection rates.

Treatment Response

Developing radiomic biomarkers (e.g., CTRS) to predict immunotherapy (ICI) response. Moving towards non-invasive virtual biopsies to monitor treatment efficacy and guide therapy selection.

① Sybil: Single LDCT to Predict Lung Cancer Risk

Radiomics; Sybil (JCO 2023)

AI

Sybil: A Validated Deep Learning Model to Predict Future Lung Cancer Risk From a Single Low-Dose Chest Computed Tomography

Peter G. Mikhael, BSc[1,2]; Jeremy Wohlwend, ME[1,2]; Adam Yala, PhD[1,2]; Ludvig Karstens, MSc[1,2]; Justin Xiang, ME[1,2]; Angelo K. Takigami, MD[3,4]; Patrick P. Bourgouin, MD[3,4]; PuiYee Chan, PhD[5]; Sofiane Mrah, MSc[4]; Wael Amayri, BSc[4]; Yu-Hsiang Juan, MD[6,7]; Cheng-Ta Yang, MD[6,8]; Yung-Liang Wan, MD[6,7]; Gigin Lin, MD, PhD[6,7]; Lecia V. Sequist, MD, MPH[3,5]; Florian J. Fintelmann, MD[3,4]; and Regina Barzilay, PhD[1,2]

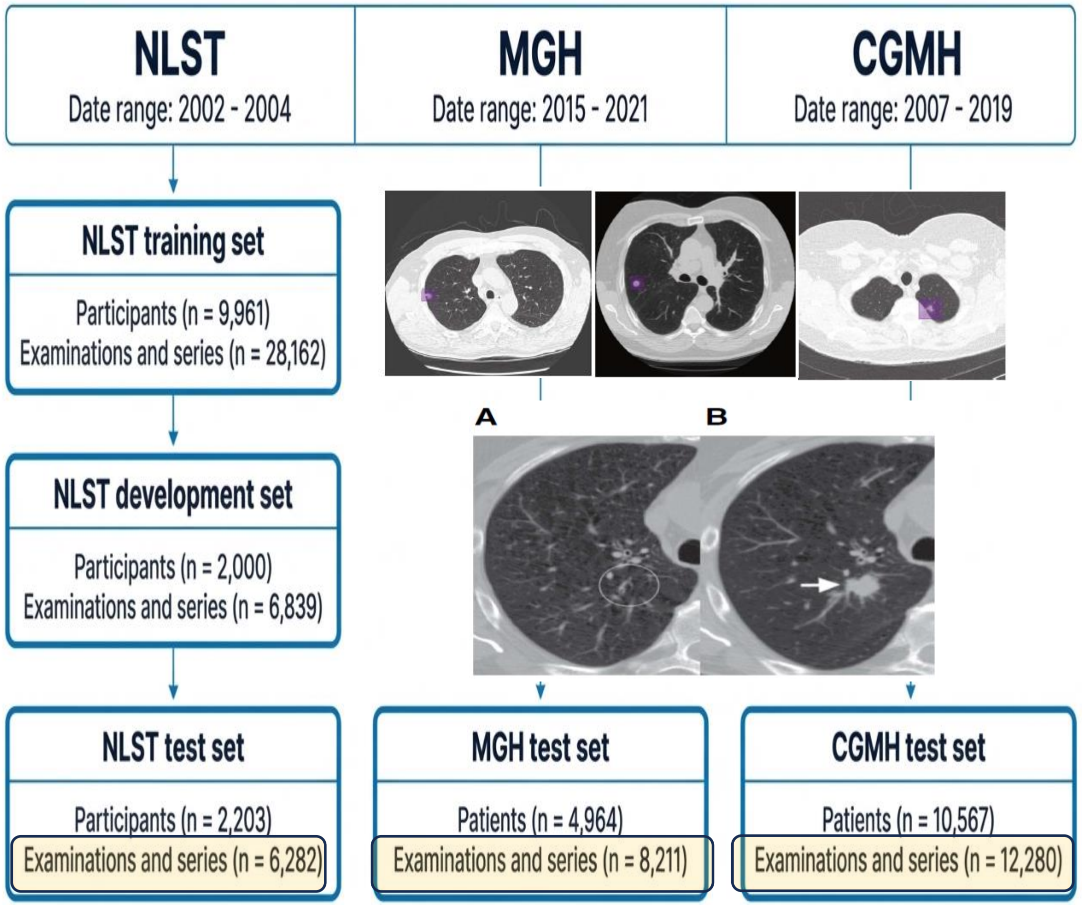
PURPOSE Low-dose computed tomography (LDCT) for lung cancer screening is effective, although most eligible people are not being screened. Tools that provide personalized future cancer risk assessment could focus approaches toward those most likely to benefit. We hypothesized that a deep learning model assessing the entire volumetric LDCT data could be built to predict individual risk without requiring additional demographic or clinical data.

METHODS We developed a model called Sybil using LDCTs from the National Lung Screening Trial (NLST). Sybil requires only one LDCT and does not require clinical data or radiologist annotations; it can run in real time in the background on a radiology reading station. Sybil was validated on three independent data sets: a heldout set of 6,282 LDCTs from NLST participants, 8,821 LDCTs from Massachusetts General Hospital (MGH), and 12,280 LDCTs from Chang Gung Memorial Hospital (CGMH, which included people with a range of smoking history including nonsmokers).

RESULTS Sybil achieved area under the receiver-operator curves for lung cancer prediction at 1 year of 0.92 (95% CI, 0.88 to 0.95) on NLST, 0.86 (95% CI, 0.82 to 0.90) on MGH, and 0.94 (95% CI, 0.91 to 1.00) on CGMH external validation sets. Concordance indices over 6 years were 0.75 (95% CI, 0.72 to 0.78), 0.81 (95% CI, 0.77 to 0.85), and 0.80 (95% CI, 0.75 to 0.86) for NLST, MGH, and CGMH, respectively.

CONCLUSION Sybil can accurately predict an individual's future lung cancer risk from a single LDCT scan to further enable personalized screening. Future study is required to understand Sybil's clinical applications. Our model and annotations are publicly available.

J Clin Oncol 41:2191-2200. © 2023 by American Society of Clinical Oncology



Modified from Mikhael PG et al. JCO 2023, Fig 1.

Mikhael PG et al. *Journal of clinical oncology* 2023; 41(12): 2191-200

① Sybil: Single LDCT to Predict Lung Cancer Risk

Radiomics; Sybil (JCO 2023)

AI

Study Overview (Mikhael et al.)

Deep learning model predicting future lung cancer risk from a single low-dose CT scan without requiring clinical data or radiologist annotation.

Key Performance Results

0.92
AUC (1-Year Prediction)

0.75
AUC (6-Year Prediction)

3
Validation Cohorts

0
Clinical Inputs Required

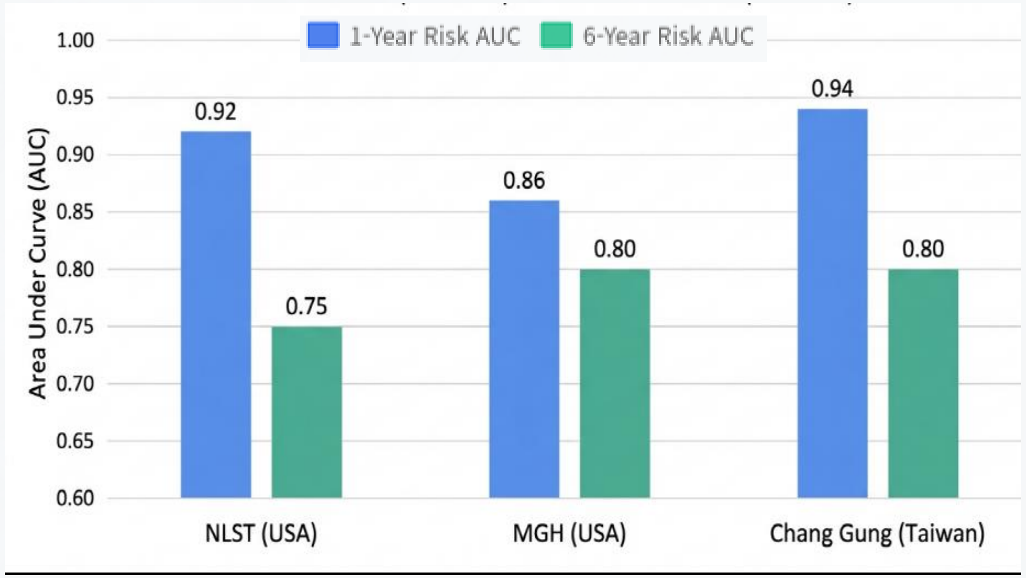
Clinical Implication

Enables opportunistic screening and risk stratification beyond current smoking-history criteria (USPSTF), potentially identifying non-smokers at high risk.

Validation Across Diverse Cohorts

Consistent performance across institutions

NLST (n=6,282) | MGH (n=8,821) | Chang Gung (n=12,280)



Sybil outperforms traditional risk models (plco2012) which rely on demographic data, solely using 3D volume imaging features.

Mikhael PG et al. *Journal of clinical oncology* 2023; 41(12): 2191-200

② AI-Assisted CXR Screening RCT

Radiomics; Nam et al. Radiology 2023

Radiology
ORIGINAL RESEARCH • THORACIC IMAGING

AI Improves Nodule Detection on Chest Radiographs in a Health Screening Population: A Randomized Controlled Trial

Ju Gang Nam, MD, PhD • Eui Jin Hwang, MD, PhD • Jayoun Kim, PhD • Nanhee Park, MS • Eun Hee Lee, MD, PhD • Hyun Jin Kim, MD, PhD • Miyeon Nam, MD • Jong Hyuk Lee, MD, PhD • Chang Min Park, MD, PhD • Jin Mo Goo, MD, PhD

Background: The impact of artificial intelligence (AI)–based computer-aided detection (CAD) software has not been prospectively explored in real-world populations.

Purpose: To investigate whether commercial AI-based CAD software could improve the detection rate of actionable lung nodules on chest radiographs in participants undergoing health checkups.

Materials and Methods: In this single-center, pragmatic, open-label randomized controlled trial, participants who underwent chest radiography between July 2020 and December 2021 in a health screening center were enrolled and randomized into intervention (AI group) and control (non-AI group) arms. One of three designated radiologists with 13–36 years of experience interpreted each radiograph, referring to the AI-based CAD results for the AI group. The primary outcome was the detection rate, that is, the number of true-positive radiographs divided by the total number of radiographs, of actionable lung nodules confirmed on CT scans obtained within 3 months. Actionable nodules were defined as solid nodules larger than 8 mm or subsolid nodules with a solid portion larger than 6 mm (Lung Imaging Reporting and Data System, or Lung-RADS, category 4). Secondary outcomes included the positive-report rate, sensitivity, false-referral rate, and malignant lung nodule detection rate. Clinical outcomes were compared between the two groups using univariable logistic regression analyses.

Results: A total of 10 476 participants (median age, 59 years [IQR, 50–66 years]; 5121 men) were randomized to an AI group (n = 5238) or non-AI group (n = 5238). The trial met the predefined primary outcome, demonstrating an improved detection rate of actionable nodules in the AI group compared with the non-AI group (0.59% [31 of 5238 participants] vs 0.25% [13 of 5238 participants], respectively; odds ratio, 2.4; 95% CI: 1.3, 4.7; P = .008). The detection rate for malignant lung nodules was higher in the AI group compared with the non-AI group (0.15% [eight of 5238 participants] vs 0.0% [0 of 5238 participants], respectively; P = .008). The AI and non-AI groups showed similar false-referral rates (45.9% [56 of 122 participants] vs 56.0% [56 of 100 participants], respectively; P = .14) and positive-report rates (2.3% [122 of 5238 participants] vs 1.9% [100 of 5238 participants]; P = .14).

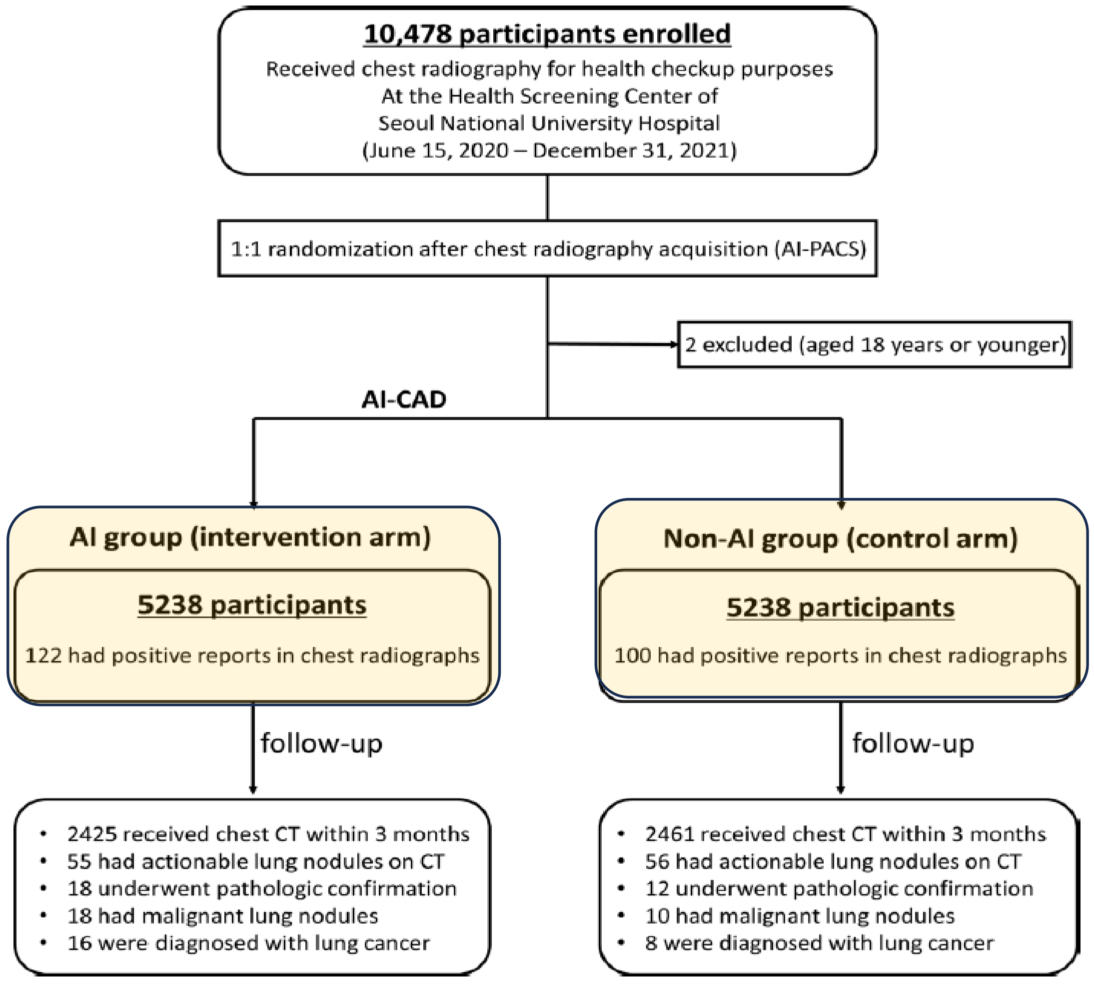


Figure 1: Participant flow diagram. AI = artificial intelligence, CAD = computer-aided detection, PACS = picture archiving and communication system.

Nam et al. *Radiology* 2023;307(2):e221894.

AI

② AI-Assisted CXR Screening RCT

Radiomics; Nam et al. Radiology 2023

Study Design & Population

Randomized Controlled Trial (RCT) in a real-world health screening setting in South Korea.

n = 10,479 participants (health screening population)
AI-Assisted Reading vs. Standard Reading (1:1)

Key Findings

AI assistance significantly improved detection rate of actional nodules (solid nodules≥ 8 mm or subsolid nodules (solid portion ≥6 mm) without increasing false referrals excessively.

0.59%
AI-Assisted Reading

0.25%
Standard Reading

Odds ratio, 2.4, *P*=0.008

Workflow Impact

Reduced reading time per case, suggesting efficiency gains alongside diagnostic accuracy improvements.

Table 3: Diagnostic Performance of Chest Radiography for Detecting Actionable Lung Nodules in the 4886 Participants Who Underwent Chest CT

Parameter	All Participants with Chest CT (*n* = 4886)	AI Group with Chest CT (*n* = 2425)	Non-AI Group with Chest CT (*n* = 2461)	*P* Value
Actionable nodules on CT scans	111	55	56	...
Positive report at chest radiography	156	87	69	...
True positive	44	31	13	...
False positive	112	56	56	...
Negative report at chest radiography	4730	2338	2349	...
True negative	4663	2314	2349	...
False negative	67	24	43	...
Diagnostic performance of chest radiography*				
Sensitivity	39.6 (30.5, 48.7)	56.4 (43.3, 69.5)	23.2 (12.2, 34.3)	<.001†
Specificity	97.7 (97.2, 98.1)	97.6 (97.0, 98.3)	97.7 (97.1, 98.3)	.94
Positive predictive value	28.2 (21.1, 35.3)	35.6 (25.6, 45.7)	18.8 (9.6, 28.1)	.02†
Negative predictive value	98.6 (98.3, 98.9)	99.0 (98.6, 99.4)	98.2 (97.7, 98.7)	.03†

Note.—Except where indicated, data are numbers of participants. These results were analyzed among 4886 participants who underwent chest CT for any purpose within 3 months after chest radiography. Actionable lung nodules were defined as solid nodules larger than 8 mm or subsolid nodules with solid a portion larger than 6 mm on CT scans (Lung Imaging Reporting and Data System, or Lung-RADS, category 4). AI = artificial intelligence.

* Diagnostic performance measures of chest radiography are presented as percentages, with 95% CIs in parentheses. Each measure was compared using the χ^2 test between the AI and non-AI groups.

† Statistically significant.

Nam et al. *Radiology* 2023;307(2):e221894.

③ CT Response Score (CTRS)

Radiomics; CTRS (JCO Clin Cancer Inform 2024)

Original Reports | Artificial Intelligence

Real-World and Clinical Trial Validation of a Deep Learning Radiomic Biomarker for PD-(L)1 Immune Checkpoint Inhibitor Response in Advanced Non–Small Cell Lung Cancer

Chiharu Sako, PhD; Chong Duan, PhD; Kevin Maresca, PhD; Sean Kent, PhD; Taly Gilat Schmidt, PhD; Hugo J.W.L. Aerts, PhD; Ravi B. Parikh, MD, MPP; George R. Simon, MD; and Petr Jordan, PhD

ABSTRACT

PURPOSE This study developed and validated a novel deep learning radiomic biomarker to estimate response to immune checkpoint inhibitor (ICI) therapy in advanced non–small cell lung cancer (NSCLC) using real-world data (RWD) and clinical trial data.

MATERIALS AND METHODS Retrospective RWD of 1,829 patients with advanced NSCLC treated with PD-(L)1 ICIs were collected from 10 academic and community institutions in the United States and Europe. The RWD included data sets for discovery (Data Set A-Discovery, n = 1,173) and independent test (Data Set B, n = 458). A radiomic pipeline, containing a deep learning feature extractor and a survival model, generated the computed tomography (CT) response score (CTRS) applied to the pretreatment routine CT/positron emission tomography (PET)-CT scan. An enhanced CTRS (eCTRS) also incorporated age, sex, treatment line, and lesion annotations. Performance was evaluated against progression-free survival (PFS) and overall survival (OS). Biomarker generalizability was further evaluated using a secondary analysis of a prospective clinical trial (ClinicalTrials.gov identifier: NCT02573259) evaluating the PD-1 inhibitor sasanlimab in second or later line of treatment (Data Set C, n = 54).

RESULTS In RWD Test Data Set B, the CTRS identified patients with a high probability of response to ICI with a PFS hazard ratio (HR) of 0.46 (95% CI, 0.26 to 0.82) and an OS HR of 0.50 (95% CI, 0.28 to 0.92) in the first-line ICI monotherapy cohort, after adjustment for baseline covariates including the PD-L1 tumor proportion score. In Clinical Trial Data Set C, the CTRS demonstrated an adjusted PFS HR of 1.03 (95% CI, 0.43 to 2.47) and an OS HR of 0.33 (95% CI, 0.14 to 0.91). The CTRS and eCTRS outperformed traditional imaging biomarkers of lesion size in PFS and OS for RWD Test Data Set B and in OS for the Clinical Trial Data Set.

CONCLUSION The study developed and validated a deep learning radiomic biomarker using pretreatment routine CT/PET-CT scans to identify ICI benefit in advanced NSCLC.

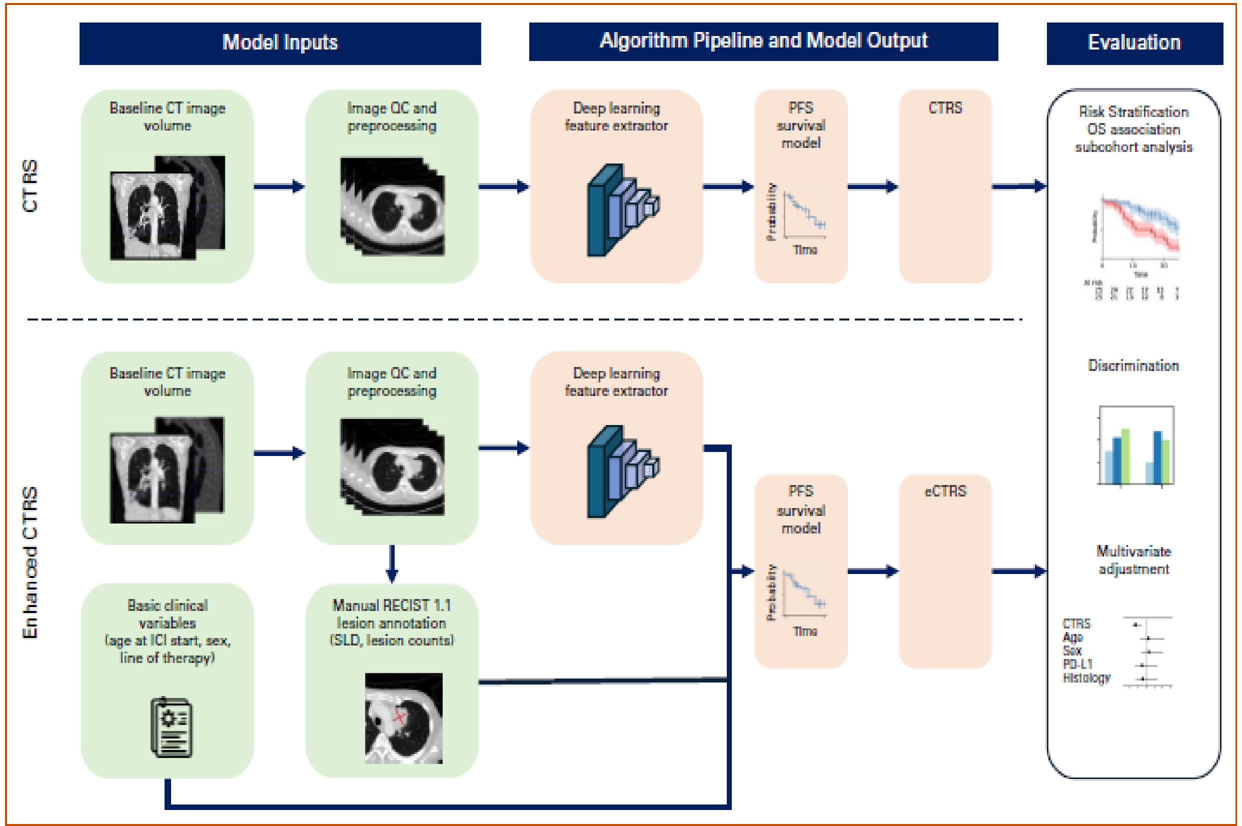


FIG 1. Diagram of the radiomic biomarker validation pipeline which includes image preprocessing, deep learning feature extractor, and survival model. This study evaluated one implementation of the CTRS only on the basis of imaging features and another implementation that combined the CTRS score with age, sex, treatment line, and manual lesion measurements as inputs (eCTRS). CT, computed tomography; CTRS, CT response score; eCTRS, enhanced CTRS; ICI, immune checkpoint inhibitor; OS, overall survival; PFS, progression-free survival; QC, quality check; SLD, sum of longest diameters.

Sako et al. *JCO Clin Cancer Inform.* 2024;8:e2400133.

③ CT Response Score (CTRS)

Radiomics; CTRS (JCO Clin Cancer Inform 2024)

Study Overview

" Real-World and Clinical Trial Validation of a Deep Learning Radiomic Biomarker for PD-(L)1 ICI Response"

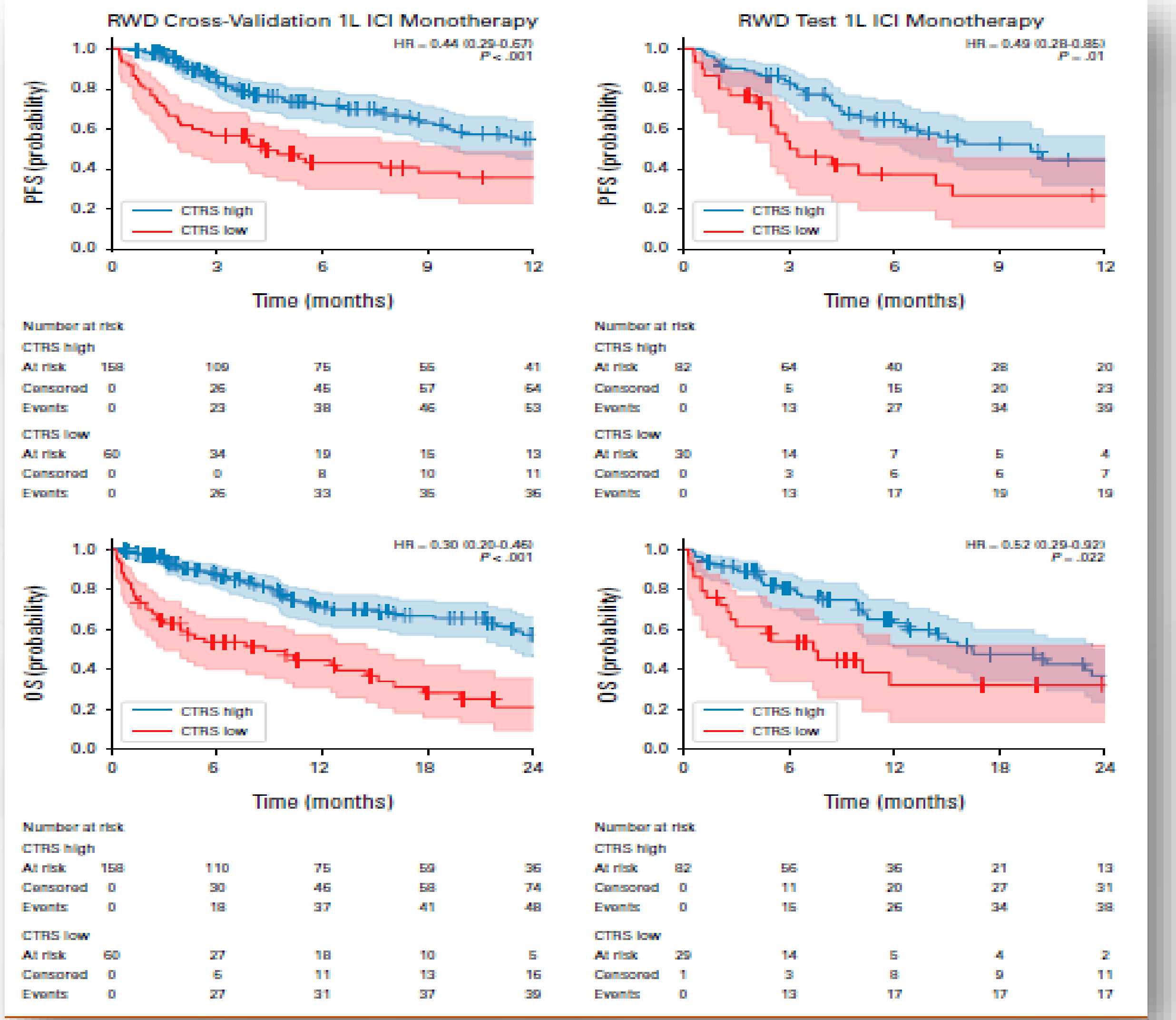
N = 1,829

10 Institutions (US, Europe)

Real-World Data

Key Findings

- Predictive Power: Deep learning radiomic biomarker (CTRS) successfully predicted response to ICI therapy in advanced NSCLC.
- Outcome Stratification: Significantly stratified Progression-Free Survival (PFS) and Overall Survival (OS) across multiple validation cohorts.
- Independence: Performance was independent of PD-L1 status, suggesting complementary value.



Sako et al. *JCO Clin Cancer Inform.* 2024;8:e2400133.

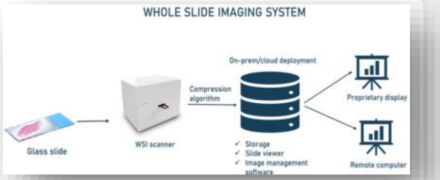
Section II

Pathomics

AI from whole-slide images for grading and immunotherapy prediction

Pathomics Overview

Pathomics Overviews



Digital Pathology Revolution

Transition from glass slides to Whole Slide Images (WSI) enables pixel-level phenotyping. AI algorithms can extract quantitative features from H&E stained slides that are imperceptible to the human eye.

Spatial Heterogeneity

Quantification of tumor microenvironment (TME) and spatial architecture. Assessing intratumoral heterogeneity and immune cell distribution patterns to predict disease progression and recurrence.

Standardized Grading

Automated histopathological grading systems (e.g., ANORAK) reduce inter-observer variability among pathologists. Providing consistent, objective scoring for complex patterns like lepidic, acinar, or micropapillary.

Predictive Biomarkers

Direct prediction of molecular alterations and treatment responses from routine histology. AI models like Deep-IO predicting Immunotherapy (ICI) response and gene mutations without additional genomic testing.

④ ANORAK: AI based histopathological grading of LUAD

Pathomics; ANORAK (Nature Cancer 2024)

nature cancer

Technical Report

https://doi.org/10.1038/s43018-023-00694-w

The artificial intelligence-based model ANORAK improves histopathological grading of lung adenocarcinoma

Received: 6 December 2022

Accepted: 21 November 2023

Published online: 10 January 2024

Check for updates

Xiaoxi Pan[1,2,19], Khalid AbduJabbar[1,2,108], Jose Coelho-Lima[3,4,108], Anca-Ioana Grapa[1,2], Hanyun Zhang[1,2], Alvin Ho Kwan Cheung[5], Juvenal Baena[6,20], Takahiro Karasaki[5,7], Claire Rachel Wilson[6,8], Marco Sereno[9], Selvaraju Veeriah[5,7], Sarah J. Aitken[3,4], Allan Hackshaw[10], Andrew G. Nicholson[11,12], Mariam Jamal-Hanjani[7,13,14], TRACERx Consortium*, Charles Swanton[5,7,14], Yinyin Yuan[1,2,19,109], John Le Quesne[15,16,17,109] & David A. Moore[5,7,18,109]

The introduction of the International Association for the Study of Lung Cancer grading system has furthered interest in histopathological grading for risk stratification in lung adenocarcinoma. Complex morphology and high intratumoral heterogeneity present challenges to pathologists, prompting the development of artificial intelligence (AI) methods. Here we developed ANORAK (pyrAmid pooliNg crOss stReam Attention networK), encoding multiresolution inputs with an attention mechanism, to delineate growth patterns from hematoxylin and eosin-stained slides. In 1,372 lung adenocarcinomas across four independent cohorts, AI-based grading was prognostic of disease-free survival, and further assisted pathologists by consistently improving prognostication in stage I tumors. Tumors with discrepant patterns between AI and pathologists had notably higher intratumoral heterogeneity. Furthermore, ANORAK facilitates the morphological and spatial assessment of the acinar pattern, capturing acinus variations with pattern transition. Collectively, our AI method enabled the precision quantification and morphology investigation of growth patterns, reflecting intratumoral histological transitions in lung adenocarcinoma.

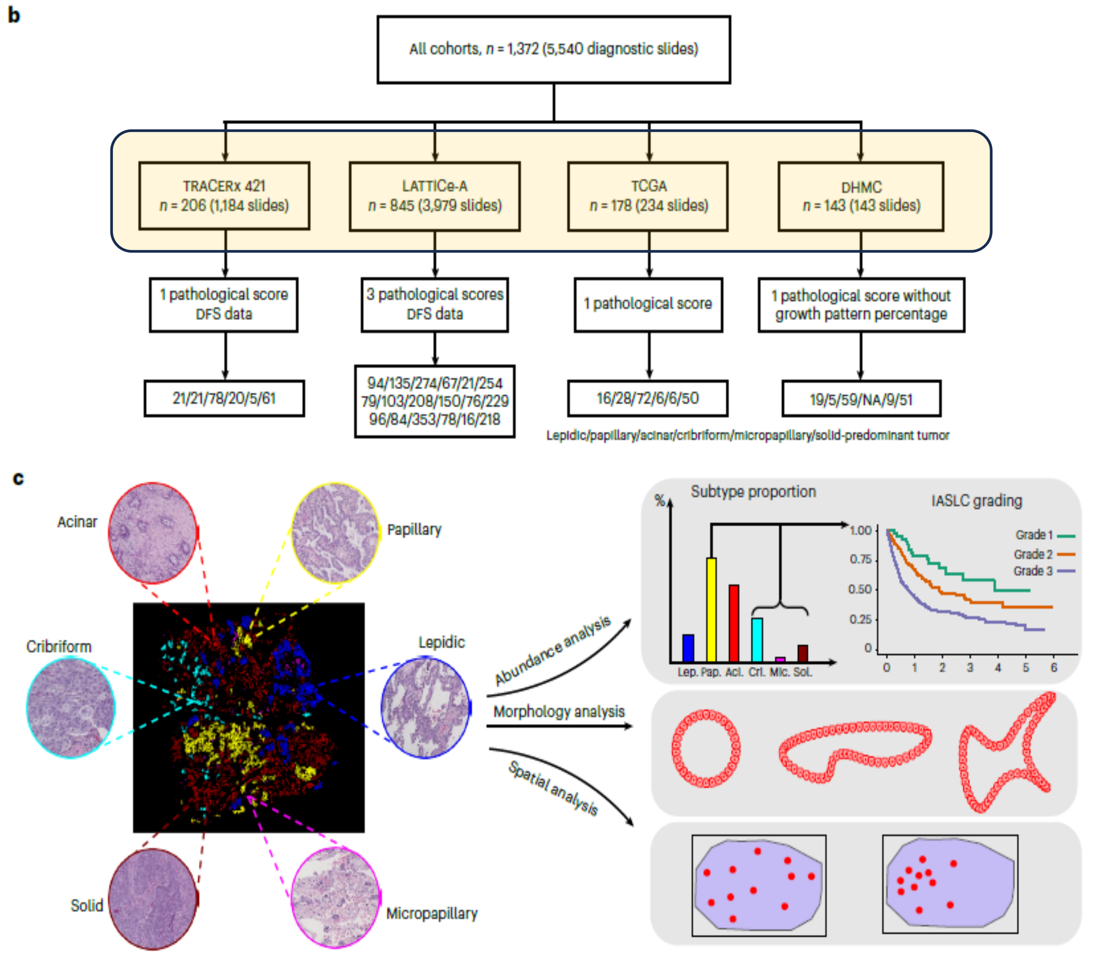


Fig. 1 | Proposed computational pipeline for precision mapping and spatial heterogeneity analyses. a, The deep learning network architecture for growth pattern segmentation integrating inputs over multiple spatial resolutions and delivering pixel-wise delineations. **b**, Overview of all the cohorts and available data. **c**, Downstream analyses enabled by the AI method, including abundance quantification, risk stratification and morphological and spatial heterogeneity analyses. PPM, Pyramid Pooling Module; Lep., lepidic; Pap., papillary; Aci., acinar; Cri., cribriform; Mic., micropapillary; Sol., solid; NA, not applicable.

Pan X et al. *Nature Cancer* 2024;5(2):347-363

AI

④ ANORAK: AI based histopathological grading of LUAD

Pathomics; ANORAK (Nature Cancer 2024)

Pan X, et al. TRACERx Consortium

Pixel-Wise Phenotyping

The ANORAK model performs precise segmentation of 6 distinct lung adenocarcinoma growth patterns (lepidic, papillary, acinar, cribriform, micropapillary, solid) at the pixel level, moving beyond simple patch classification.

Study Scale & Validation

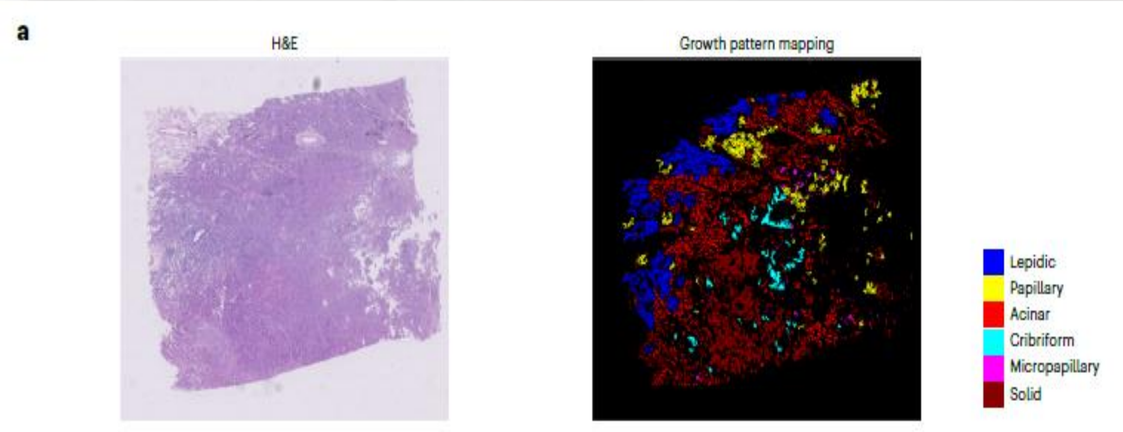
Developed on TRACERx multi-region data and validated across 4 independent cohorts (n=1,372 patients). It captures spatial heterogeneity often missed by human grading.

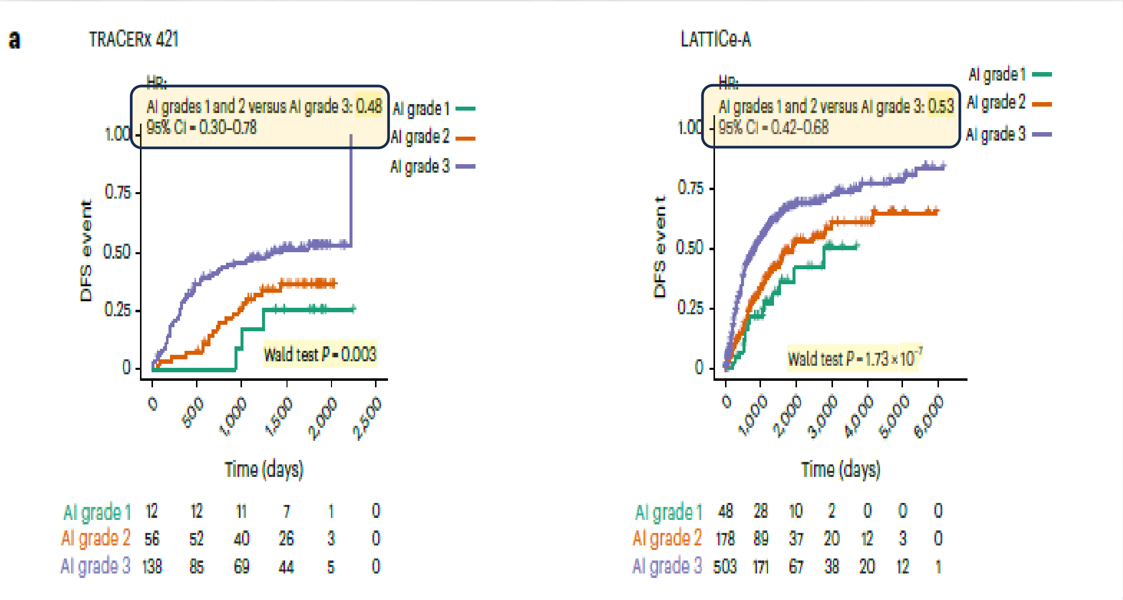
Key Findings

ANORAK-derived grades showed superior prognostic stratification compared to pathologist consensus. It remained an independent predictor of disease-free survival (DFS) in multivariate analysis.

HR: 0.48, *P*=0.003
(AI grade 1–2 vs grade 3)

DFS Hazard Ratio (TRACERx)





Pan X et al. *Nature Cancer* 2024;5(2):347-363

⑤ Deep-IO: H&E to ICI Response

Pathomics; Deep-IO (JAMA Oncol 2025)

JAMA Oncology | Original Investigation

Deep Learning Model for Predicting Immunotherapy Response in Advanced Non–Small Cell Lung Cancer

Mehrdad Rakaee, PhD; Masoud Tafavvoghi, MSc; Biagio Ricciuti, MD; Joao V. Alessi, MD; Alessio Cortellini, MD, PhD; Fabrizio Citarella, MD; Lorenzo Nibid, MD; Giuseppe Perrone, MD; Elio Adib, MD; Claudia A. M. Fulgenzi, MD; Cassio Murilo Hidalgo Filho, MD; Alessandro Di Federico, MD; Falah Jabar, PhD; Sayed Hashemi, MD; Ilias Houda, MD; Elin Richardsen, MD, PhD; Lill-Tove Rasmussen Busund, MD, PhD; Tom Donnem, MD, PhD; Idris Bahce, MD, PhD; David J. Pinato, MD, PhD; Åslaug Helland, MD, PhD; Lynette M. Sholl, MD; Mark M. Awad, MD, PhD; David J. Kwiatkowski, MD, PhD

IMPORTANCE Only a small fraction of patients with advanced non–small cell lung cancer (NSCLC) respond to immune checkpoint inhibitor (ICI) treatment. For optimal personalized NSCLC care, it is imperative to identify patients who are most likely to benefit from immunotherapy.

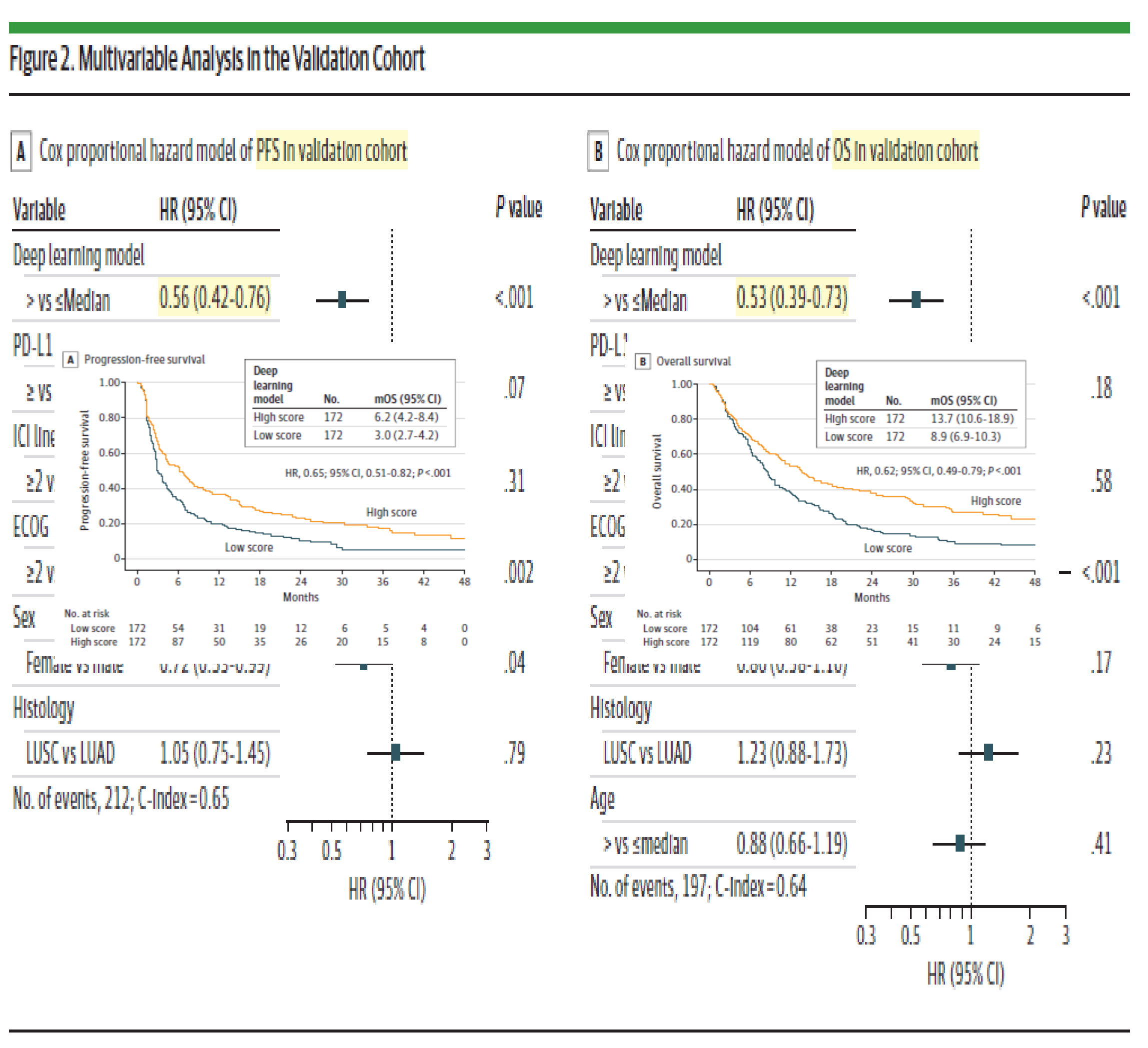
OBJECTIVE To develop a supervised deep learning–based ICI response prediction method; evaluate its performance alongside other known predictive biomarkers; and assess its association with clinical outcomes in patients with advanced NSCLC.

DESIGN, SETTING, AND PARTICIPANTS This multicenter cohort study developed and independently validated a deep learning–based response stratification model for predicting ICI treatment outcome in patients with advanced NSCLC from whole slide hematoxylin and eosin–stained images. Images for model development and validation were obtained from 1 participating center in the US and 3 in the European Union (EU) from August 2014 to December 2022. Data analyses were performed from September 2022 to May 2024.

EXPOSURE Monotherapy with ICIs.

MAIN OUTCOMES AND MEASURES Model performance measured by clinical end points and objective response rate (ORR) differentiation power vs other predictive biomarkers, ie, programmed death-ligand 1 (PD-L1), tumor mutational burden (TMB), and tumor-infiltrating lymphocytes (TILs).

RESULTS A total of 295 581 image tiles from 958 patients (mean [SD] age, 66.0 [10.6] years; 456 [48%] females and 502 [52%] males) treated with ICI for NSCLC were included in the analysis. The US-based development cohort consisted of 614 patients with median (IQR) follow-up time of 54.5 (38.2-68.1) months, and the EU-based validation cohort, 344 patients with 43.3 (27.4-53.9) months of follow-up. The ORR to ICI was 26% in the developmental cohort and 28% in the validation cohort. The deep learning model's area under the receiver operating characteristic curve (AUC) for ORR was 0.75 (95% CI, 0.64-0.85) in the internal test set and 0.66 (95% CI, 0.60-0.72) in the validation cohort. In a multivariable analysis, the deep learning model's score was an independent predictor of ICI response in the validation cohort for both progression-free (hazard ratio, 0.56; 95% CI, 0.42-0.76; $P < .001$) and overall survival (hazard ratio, 0.53; 95% CI, 0.39-0.73; $P < .001$). The tuned deep learning model achieved a higher AUC than TMB, TILs, and PD-L1 in the internal set; in the validation cohort, it was superior to TILs and comparable with PD-L1 (AUC, 0.67; 95% CI, 0.60-0.74), with a 10-percentage point improvement in specificity. In the validation cohort, combining the deep learning model with PD-L1 scores achieved an AUC of 0.70 (95% CI, 0.63-0.76), outperforming either marker alone, with a response rate of 51% compared to 41% for PD-L1 (≥50%) alone.



Rakaee M et al. *Jama Oncol.* 2025;11(2):109-118.

AI

⑤ Deep-IO: H&E to ICI Response

Pathomics; Deep-IO (JAMA Oncol 2025)

Multi-Center Validation Study

Deep learning model predicting immunotherapy response directly from H&E-stained slides. Multicenter cohort (N=958 total, validation n=344) across US and Europe.

Superior Predictive Performance

Deep-IO demonstrated independent predictive value for survival outcomes.

- PFS Prediction: HR 0.56 (95% CI: 0.42-0.76, p<0.001)
- OS Prediction: HR 0.53 (95% CI: 0.39-0.73, p<0.001)

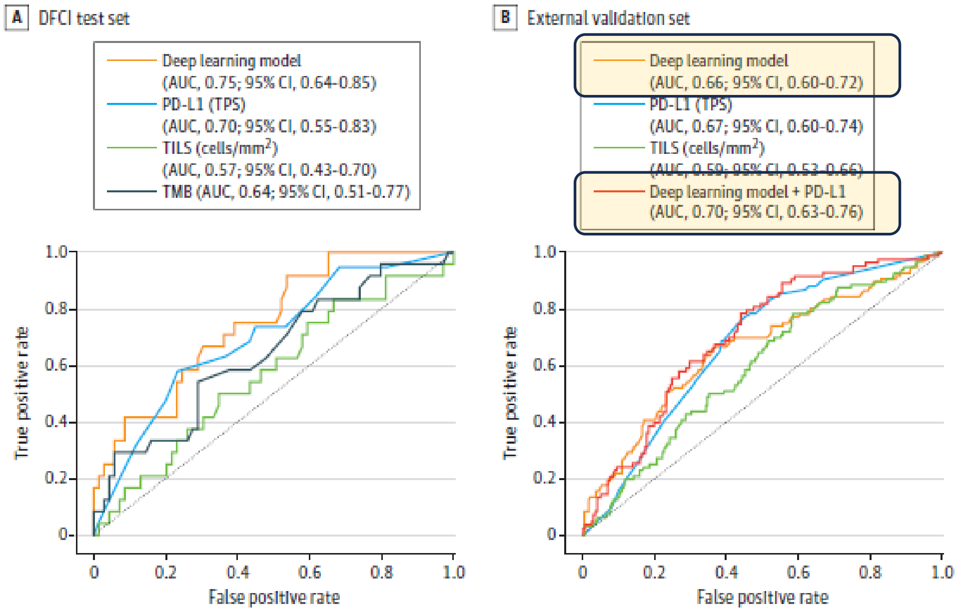
Synergy with PD-L1

Combining Deep-IO with PD-L1 improved response prediction accuracy compared to either biomarker alone.

Validation Cohort Performance (n=344)

0.66	0.70
AUC (Deep-IO)	AUC (Combined with PD-L1)

Figure 3. Performance of Immune Checkpoint Inhibitors (ICIs) Biomarkers vs the Deep-IO Model



Rakaee M et al. *Jama Oncol.* 2025;11(2):109-118.

Section III

Multimodal AI

Integrating radiomics, pathomics, and clinicogenomic information

Why Multimodal AI?

Overcoming Limitations

Single-modality approaches often face ceiling effects and confounding variables. Integrating diverse data streams provides a more holistic view of tumor biology that no single test can capture alone.

Synergistic Power

Combines macroscopic radiographic burden (CT), microscopic tissue architecture (Pathology), and molecular landscape (Genomics). The whole is greater than the sum of its parts for predictive accuracy.

Precision Selection

Moving beyond simple prognostication to actual treatment selection (e.g., Mono-ICI vs. Chemo-ICI) and dynamic monitoring of therapeutic response through longitudinal data integration.

"The next frontier involves systems that synthesize radiomics, pathomics, and clinicogenomic information to guide personalized therapeutic strategies."

⑥ A-STEP: Clinicogenomic Treatment Selection

Multimodal AI; A-STEP (Nature Communications,2025)

nature communications

Article

https://doi.org/10.1038/s41467-025-61823-w

Machine-learning driven strategies for adapting immunotherapy in metastatic NSCLC

Received: 2 February 2024

Accepted: 2 July 2025

Published online: 24 July 2025

Check for updates

A list of authors and their affiliations appears at the end of the paper

Immune checkpoint inhibitors (ICIs), either as monotherapy (ICI-Mono) or combined with chemotherapy (ICI-Chemo), improves survival in advanced non-small cell lung cancer (NSCLC). However, prospective guidance for choosing between these options remains limited, and single-feature biomarkers like PD-L1 prove inadequate. We develop a machine learning model using clinicogenomic data from four cohorts (MD Anderson $n = 750$; Mayo Clinic $n = 80$; Dana-Farber $n = 1077$; Stand Up To Cancer $n = 393$) to predict individual benefit from adding chemotherapy. Benefit scores are calculated using five distinct functions derived from 28 genomic and 6 clinical features. Our integrated model, A-STEP (Attention-based Scoring for Treatment Effect Prediction), estimates heterogeneous treatment effects and achieves the largest reduction in 3-month progression risk, improving weighted risk reduction by 13–23% over stand-alone models. A-STEP recommends treatment changes for over 50% of patients, most often favoring ICI-Chemo. In simulation on external cohort, patients treated in accordance with A-STEP recommendations show improved 2-year progression-free survival (HR = 0.60 for ICI-Mono treatment arm; HR = 0.58 for ICI-Chemo treatment arm). Predictive features include FBXW7, APC, and PD-L1. In this study, we demonstrate how machine learning can fill critical gaps in immunotherapy selection for NSCLC, by modeling treatment heterogeneity with real-world clinicogenomic data, driving precision medicine beyond conventional biomarker boundaries.

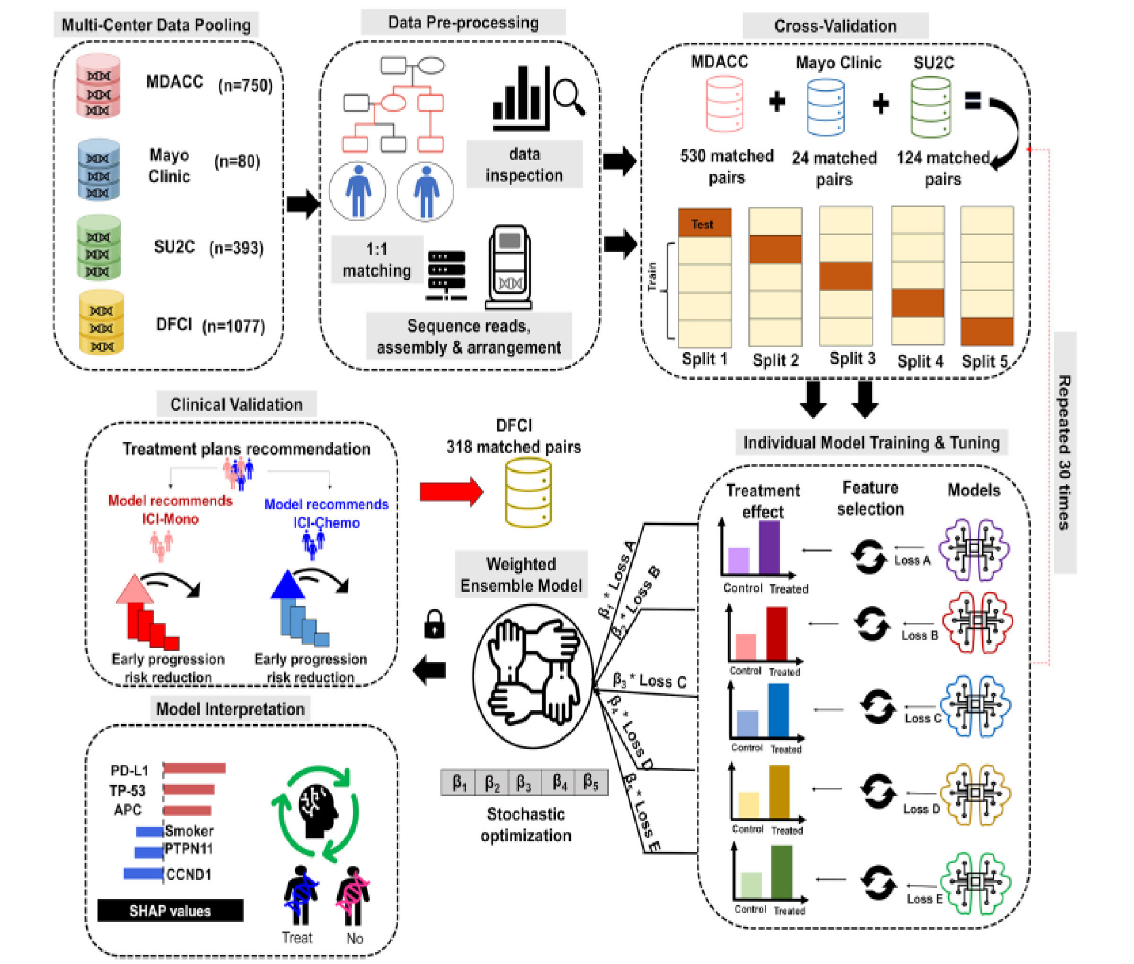


Fig. 1 | Overview of the A-STEP system from data collection to clinical translation. Schematic diagram illustrating the A-STEP workflow, including data collection, preprocessing, modeling, and translation to clinical practice. Institutions involved include MD Anderson Cancer Center (MDACC), Stand Up To Cancer Consortium (SU2C), and Dana-Farber Cancer Institute (DFCI).

Saad MB et al. *Nature Communications* (2025) 16:6828

⑥ A-STEP: Clinicogenomic Treatment Selection

Multimodal AI; A-STEP (Nature Communications,2025)

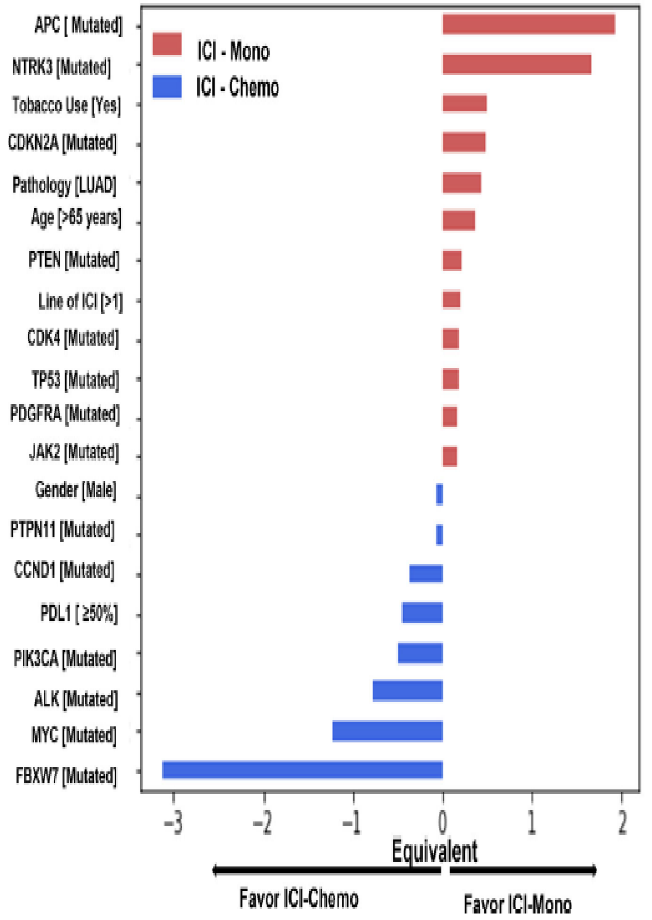


Fig. 5 | SHAP-based interpretation of the A-STEP model. SHAP interpretation of the A-STEP model showing the top 20 features ranked by their importance, reflecting both the direction and magnitude of their effects on model predictions. Source data are provided as a Source Data file.

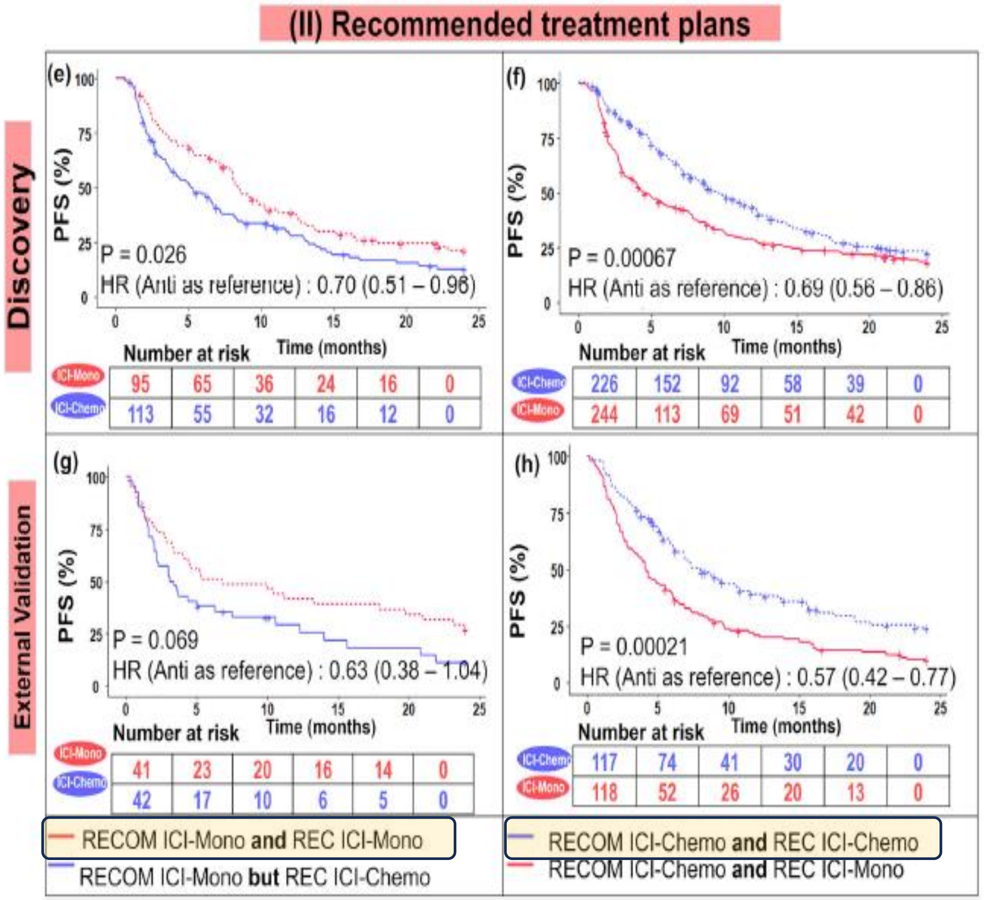


Fig. 3 | Impact of A-STEP Recommendations on 2-Year Progression-Free Survival. Comparison of 2-year progression-free survival (PFS) stratified by (I) actual treatment plans (received, REC) and (II) treatment plans recommended by A-STEP (RECOM). Subfigures (a–d) and (e–h) represent the discovery and external validation cohorts, respectively. *P*-values were calculated using the two-sided log-rank test. Hazard ratios (HRs) and 95% confidence intervals (CIs) are shown, with the anti-recommendation group used as the reference. Source data are provided as a Source Data file.

Saad MB et al. *Nature Communications* (2025) 16:6828

⑥ A-STEP: Clinicogenomic Treatment Selection

Multimodal AI; A-STEP (Nature Communications,2025)

Study Objective

To develop a machine learning-driven strategy for selecting the optimal first-line therapy (ICI Monotherapy vs. Chemo-Immunotherapy) in metastatic NSCLC.

Multimodal Methodology

Integration of clinical variables (age, sex, ECOG, histology) and genomic features (molecular alterations from panel sequencing).

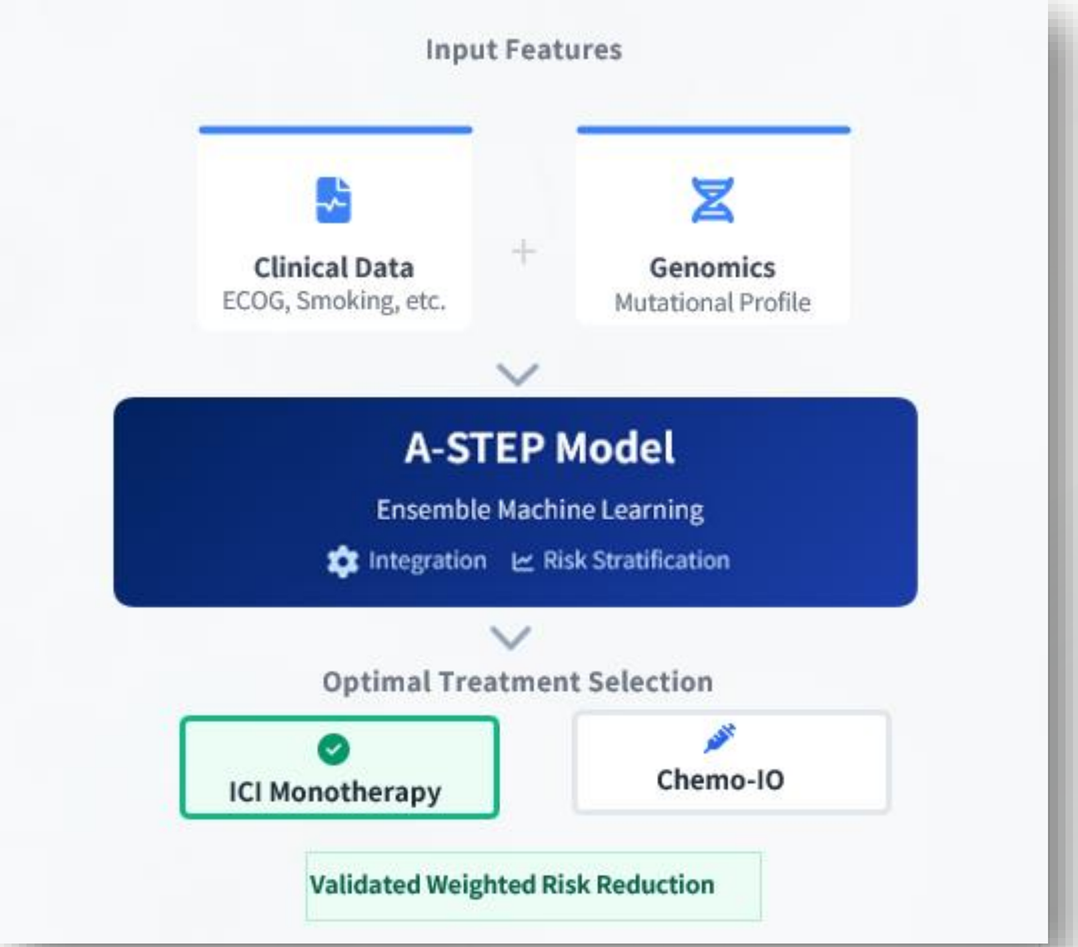
Discovery cohort: 678 matched cases; Validation cohort: 318 cases

Key Findings

The A-STEP ensemble model successfully identified patient subgroups benefiting from specific regimens, demonstrating significant Weighted Risk Reduction (WRR) compared to standard of care selection.

Implication

Moving beyond single biomarkers (like PD-L1) to comprehensive clinicogenomic profiling for precise therapeutic decision-making



Saad MB et al. *Nature Communications* (2025) 16:6828

Section IV

Research Design & Methodology

From reporting standards to AI-specific trial designs

AI Trial Design Innovations

Research Methodology

AI

Synthetic Control Arms

Concept

Generating control groups using historical trial data and Real-World Data (RWD) via machine learning matching.

Key Application

Replacing placebo arms in rare mutations (e.g., ROS1, RET) where recruitment is challenging.

- Reduces patient burden (fewer on placebo)
- Accelerates trial timeline
- Lowers trial costs

Digital Twins

Concept

Creating computational patient avatars to simulate treatment responses in silico before actual dosing.

Key Application

Precision cohort enrichment by predicting likelihood of response/toxicity for inclusion criteria.

- Optimized inclusion/exclusion
- Personalized dosing simulation
- Reduced failure rates in Phase II/III

Federated Learning

Concept

Training algorithms across institutions without sharing raw patient data (model travels, data stays).

Key Application

Building robust, generalizable AI models across multi-center global networks while preserving privacy.

- Privacy-preserving collaboration
- Access to diverse datasets
- Reduced data silo effects

CONSORT-AI: Reporting Guideline

Research Methodology

AI Algorithm Specification

Explicitly state the AI model version, input data requirements (e.g., image resolution, stain normalization), and pre-processing steps. Ensure reproducibility by detailing software dependencies and availability.

Training & Validation Datasets

Describe the source of training data and how it differs from the trial population. Report demographics to assess potential algorithmic bias. Define strict separation between development, internal validation, and test sets.

Human-AI Interaction

Define the protocol for AI intervention: Is it a 'second reader', a 'triage tool', or a 'concurrent aid'? Specify how clinicians should resolve disagreements with AI predictions and any required training.

Performance & Failure Modes

Prespecify clinical endpoints (not just AUC) and error analysis plans. Report performance across subgroups to detect bias. Establish protocols for handling AI failures or unavailable outputs (e.g., poor image quality).

Liu et al., *Nature Medicine* 2020

Section V

Ongoing Prospective Trials

Screening, treatment selection, and workflow evaluations

Ongoing Prospective Trials

From Detection to Risk Prediction: AI in Lung Cancer Screening

Ongoing Trials

Trial Name / Region	Design / Scale	Target Population	AI Application	Primary Endpoint
DACAPO NCT05704920 (France)	RCT N = 2,722	Heavy Smokers High-risk (50-80y)	AI-integrated LDCT workflow (CAD)	Detection Rate Cost-effectiveness
Taiwan AI-LDCT NCT07280559 (Taiwan)	Pragmatic RCT N ≈ 3,000	General Screening Smokers & Non-smokers	AI-assisted Interpretation Support	Diagnostic Accuracy Workflow Efficiency
LC-SHIELD NCT06295497 (Hong Kong)	Prospective Cohort N = 5,000	Never-Smokers High-risk (Family Hx)	AI as First Reader for Triage	Early Detection Rate In Asian Non-smokers
Sybil Validation USA (MGH, BWH)	Prospective Validation Multi-site	Incidental & Screening Diverse Cohorts	6-Year Risk Prediction Model	AUC / Calibration Real-world Utility

Data from ClinicalTrials.gov and ASCO/ESMO abstracts (2024-2025

AI

Ongoing Prospective Trials

Ongoing Trials

Beyond Diagnosis: AI for Response Prediction, Genomics, and Cost-Effectiveness

Trial Name / ID	Design & Region	AI Application	Primary Endpoint	Key Goal
DeepGEM NCT07110259	China Cluster RCT (N=1,200)	Pathology → Mutation Predicting EGFR/ALK/ROS1 from H&E images	Concordance with NGS, Turnaround Time (TAT)	Pre-screening (No Sequencer)
AEGEAN Substudy NCT03800134	Global Phase III RCT (N=802)	Multimodal Monitor Radiomics + ctDNA dynamics integration	Pathological Complete Response (pCR)	Non-invasive Response Tracking
China Cost RCT NCT06988579	China Cluster RCT (N=7,294)	Radiologist Assist AI-aided reading workflow optimization	Healthcare Costs, Diagnostic Accuracy	Economic Viability
Wuhan Diagnosis NCT04000620	China Prospective (Diagnostic)	Deep Learning Staging Automated TNM staging from CT imaging	Staging Accuracy vs. Pathological Staging	Treatment Planning

Data from ClinicalTrials.gov and ASCO/ESMO abstracts (2024-2025

Beyond Accuracy

Newer pragmatic RCTs (e.g., DACAPO, Taiwan) are evaluating cost-effectiveness and workflow efficiency rather than just diagnostic sensitivity.

AI Role Evolution

AI is evolving from a concurrent CAD "second opinion" to an independent "first reader" for triage (LC-SHIELD) or a long-term risk predictor (Sybil).

Asian Pragmatism

Large-scale cluster RCTs (>1,000 patients) are predominantly led by Asian centers (China, Taiwan), focusing on practical implementation in high-volume settings.

AI

Challenges & Futures

Generalizability Gap

Ensuring consistent model performance across diverse patient populations (e.g., Asian vs. Western), varying scanner manufacturers (GE, Siemens, Philips), and different imaging protocols. Overcoming the "domain shift" problem.

Calibration over Discrimination

Moving beyond simple AUC metrics to rigorous risk calibration. A model must not only rank patients correctly (discrimination) but also provide accurate probability estimates (calibration) to be clinically actionable.

Cost-Effectiveness & Reimbursement

Demonstrating economic value is crucial for adoption. Does AI reduce total costs (e.g., fewer biopsies, earlier stage detection) to justify reimbursement? Trials like the China Cost RCT (NCT06988579) are addressing this.

Governance & Liability

Establishing clear audit trails for AI-assisted decisions. Defining liability when human and AI disagree. Protocols for continuous monitoring and "version control" as algorithms update over time.

Conclusions and Take-Home Messages

Summary

AI

From Research to Practice

AI in lung cancer is rapidly transitioning from retrospective experimental utility to clinical necessity. Prospective trials (DACAPO, Taiwan AI-LDCT) are now validating real-world effectiveness.

Value of Integrated AI

While radiomics and pathomics offer independent value, multimodal integration (clinicogenomics + imaging) demonstrates superior performance for precision treatment selection and risk stratification.

Methodological Rigor

Standardized reporting (CONSORT-AI) and pragmatic trial designs are critical enablers. Focus is shifting towards demonstrating cost-effectiveness, workflow efficiency, and robust external validity.

Upcoming Milestones

The completion of ongoing large-scale trials (2024–2026) will provide definitive evidence to shape future clinical guidelines, reimbursement policies, and widespread adoption in lung cancer care.

Success usually has a lot of
failure wrapped up in it.
We only truly fail when we shrink away from
challenges to stay in a comfort zone

Stephen Curry

The Greatest Shooter • Golden State Warriors

Thank You

for your attention