Deep learning using chest x-rays to identify high-risk candidates for lung cancer screening CT

Vineet K Raghu, PhD; Anika S Walia, BA; Aniket N Zinzuwadia, BA; Hugo Aerts, PhD; Inga T Lennes, MD, MBA; and Michael T Lu, MD, MPH
Cardiovascular Imaging Research Center, Massachusetts General Hospital and Harvard Medical School, Boston, MA
Program for Artificial Intelligence in Medicine, Brigham and Women's Hospital, Boston, MA

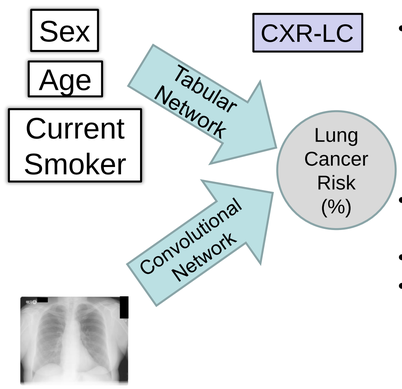
Background

- Lung cancer screening (LCS) with CT reduces lung cancer death by 20-25%
- LCS eligibility is set by 2022 Centers for Medicare and Medicaid Services (CMS) / USPSTF criteria:
  - Age 50-77 years
  - ≥20 pack-year smoking history (packs per day x years of smoking)
  - Currently smoking or quit within the past 15 years
- In the US, ~5-10% of eligible persons are screened
- Automated electronic medical record (EMR) approaches may improve uptake, but pack-year input for CMS eligibility often not in the EMR

Purpose

We tested whether an AI model (CXR-LC) can identify persons at high lung cancer risk using a chest x-ray image and basic data

Model Development



- An open-source neural network (CXR-LC) was developed using chest x-ray and basic data from 33,485 participants from the Prostate, Lung, Colorectal, and Ovarian (PLCO) cancer screening trial (in *Annals of Internal Medicine*[1])
- **Inputs**: age, sex, current smoking, and a chest x-ray image
- **Output**: 12-year lung cancer risk
- Discretized output to "CXR-LC Eligible" vs. "CXR-LC Ineligible" using a 3.297% risk threshold (matched the size of the CMS eligible population in PLCO)

Results

Table 2: 6-year lung cancer rates by CXR-LC and 2022 CMS Eligibility across validation cohorts.

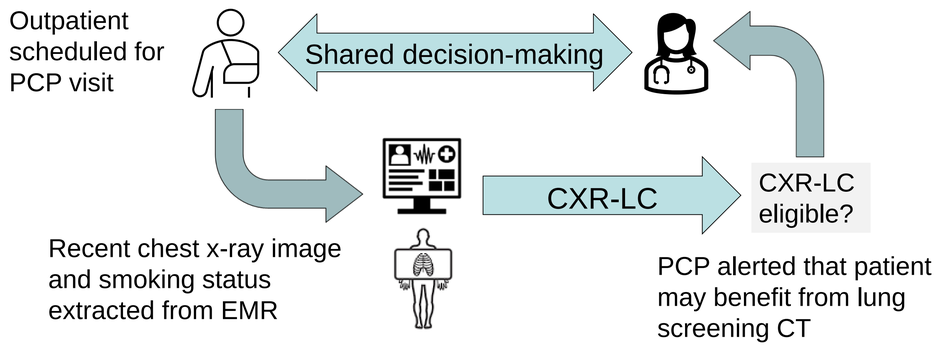
	CMS Eligible		CMS Ineligible		CMS Eligibility Unknown	
Cohort	**CXR-LC Eligible**	**CXR-LC Ineligible**	**CXR-LC Eligible**	**CXR-LC Ineligible**	**CXR-LC Eligible**	**CXR-LC Ineligible**
PLCO	**3.1% (43 / 1376)**	1.6% (17 / 1092)	**2.9% (22 / 750)**	0.4% (9 / 2397)	-	-
NLST	**5.1% (172 / 3343)**	1.6% (34 / 2150)	**-**	-	-	-
SNUH (5-year rates)	**2.2% (33 / 1520)**	0.2% (1 / 452)	**1.3% (20 / 1490)**	0.3% (2 / 730)	1.7% (8 / 482)	0.0% (0 / 200)
MGB	**8.5% (83 / 974)**	2.8% (5 / 177)	**3.3% (121 / 3703)**	0.5% (7 / 1423)	2.5% (127 / 5177)	0.5% (18 / 3283)

From a routine chest x-ray image, **CXR-LC identified persons at high risk of lung cancer**, beyond the Medicare lung screening CT eligibility criteria, in 4 validation cohorts (>30k individuals).

A pilot clinical trial will test whether CXR-LC-based EMR reminders **can improve lung cancer screening CT participation.**

CXR-LC open source model

Implementation Pilot



- 6-month period over 10 primary care clinics in Boston
- Control arm – no alerts, regardless of CXR-LC predicted risk
- Primary outcome is lung cancer screening CT within 6 months of PCP visit
- Grants from Natl Academy of Medicine/J&J and CRICO

Cohorts

Table 1: Characteristics of model development and validation cohorts

	PLCO Development (N=22711)	PLCO Validation (N=5615)	NLST (N=5493)	SNUH (N=4874)	MGB (N=14737)
Age, mean (SD), y	62.2 (5.3)	62.1 (5.3)	61.7 (5.0)	57.0 (6.0)	62.6 (6.8)
Male Sex (%)	60.1	60.1	55.3	96.0	48.5
Current Smoking (%)	19.3	20.2	49.6	29.0	26.8
Self-Reported Race					
Asian	3.7	3.5	0.7	100.0	1.4
Black	6.2	6.3	4.1	-	7.3
White	87.2	87.1	94.2	-	87.2
Other	2.8	3.0	1.0	-	3.2
Pack-years, mean (SD)	35.2 (29.0)	35.4 (29.0)	55.7 (23.5)	-	18.6 (23.5)
6-year lung cancer (%)	1.8	1.6	3.8	1.3	3.4

CXR-LC was tested in 4 cohorts with no history of lung cancer or LCS

- PLCO, smokers 55-74 years of age, 10 US sites, 1993-2001
- NLST (N=5,493); ≥30 pack-year smokers, 33 US sites, 2002-2004
- SNUH (N=4874) health checkup smokers in Seoul 50-80 years
- MGB (N=14737), smokers 50-80 years and had an outpatient CXR in 2013-2014. Data from the EMR

Conclusion

A deep learning model, CXR-LC, can accurately estimate lung cancer risk using basic data available in the EMR across US and Korean validation cohorts and in subgroups defined by sex and race

Publications

1. Lu, MT*, Raghu, VK* et al. Deep learning using chest radiographs to identify high-risk smokers for lung cancer screening: Development and Validation of a Prediction Model. *Annals of Intern Med* 2020
2. Lee JH, et al. Deep learning to optimize candidate selection for lung cancer CT screening: Advancing the 2021 USPSTF Recommendations. *Radiology* 2022
3. Raghu VK, et al. Validation of a deep learning-based model to predict lung cancer risk using chest radiographs and electronic medical record data. *JAMA Network Open* 2022