Deep learning to assess long-term mortality from chest radiographs (x-rays)

Michael T Lu[1], Alexander Ivanov[1], Thomas Mayrhofer[1,2], Ahmed Hosny[3], Hugo JWL Aerts[3], Udo Hoffmann[1]. *JAMA Open* 2019;2(7):e197416.

[1]Cardiovascular Imaging Research Center (CIRC), MGH Department of Radiology; [2]Straslund University of Applied Sciences; [3]Program for AI in Medicine (AIM), BWH

Objective

The chest radiograph (x-ray or CXR) is the most common diagnostic imaging test in medicine; it may also provide a window into longevity and prognosis. We developed and tested a convolutional neural network (CNN) to predict long-term (12-year) mortality, including non-cancer death, from chest radiographs.

Datasets

1. Development (CNN training + tuning) in the Prostate, Lung, Colorectal, and Ovarian screening trial **(PLCO, n=41,856)**
- chest radiograph screening arm
- asymptomatic nonsmokers and smokers aged 55-74
- 10 US sites from 1993-2001.

2. Testing in held out PLCO dataset (**n=10,464**)

3. External testing in the **National Lung Screening Trial (NLST, n=5,493)**
- chest radiograph screening arm
- asymptomatic heavy smokers (≥30 pack-year) aged 55-74
- 21 US sites from 2002-2004.

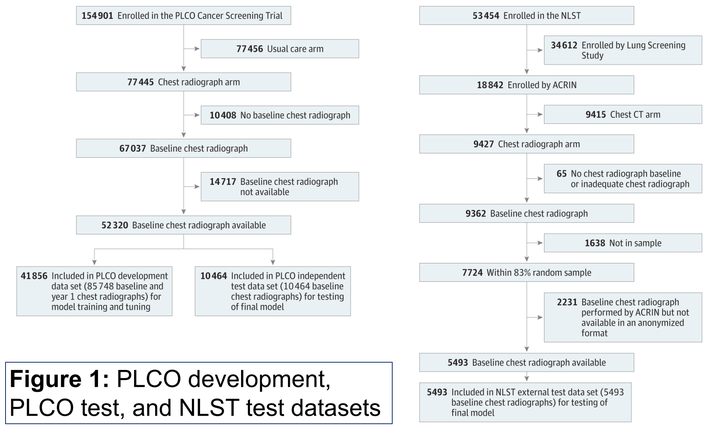


Figure 1: PLCO development, PLCO test, and NLST test datasets

CXR-Risk CNN

- Input is chest radiograph image
- Output is probability of 12-year mortality
- Inception v4 w/ transfer learning from ImageNet
- Geometric and mixup data augmentation
- Progressive resizing to 224 pixels
- One cycle w/ ADAM. Fastai & PyTorch

Figure 2: Kaplan-Meier survival by CXR-Risk score

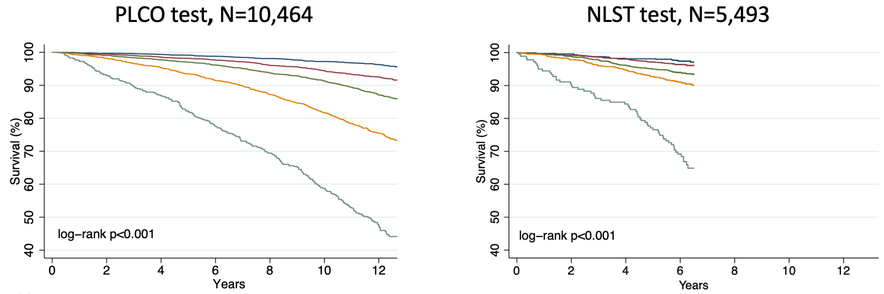


Table: Mortality by CXR-Risk score

CXR-Risk Score	Mortality, No./Total No. (%)	Deaths per 1000 Person-Years (95% CI)	Unadjusted HR (95% CI)	Unadjusted *P* Value	Adjusted HR (95% CI)[a]	Adjusted *P* Value
PLCO Test Data Set (12-y Follow-up)						
Very low	97/2543 (3.8)	3.3 (2.7-4.1)	1 [Reference]	NA	1 [Reference]	NA
Low	216/2769 (7.8)	6.8 (5.9-7.7)	2.0 (1.6-2.6)	<.001	1.4 (1.1-1.8)	.003
Moderate	339/2674 (12.7)	11.1 (10.0-12.4)	3.3 (2.7-4.2)	<.001	1.7 (1.3-2.2)	<.001
High	500/2006 (24.9)	23.0 (21.1-25.1)	7.0 (5.6-8.6)	<.001	2.6 (2.1-3.4)	<.001
Very high	250/472 (53.0)	57.4 (50.8-65.0)	18.3 (14.5-23.2)	<.001	4.8 (3.6-6.4)	<.001
Total	1402/10 464 (13.4)	11.9 (11.3-12.6)	NA	NA	NA	NA
NLST Test Data Set (6-y Follow-up)						
Very low	20/752 (2.7)	4.2 (2.7-6.6)	1 [Reference]	NA	1 [Reference]	NA
Low	64/1679 (3.8)	6.1 (4.8-7.8)	1.4 (0.9-2.4)	.16	1.2 (0.7-1.9)	.56
Moderate	115/1723 (6.7)	10.9 (9.1-13.1)	2.6 (1.6-4.1)	<.001	1.7 (1.0-2.8)	.03
High	114/1159 (9.8)	16.4 (13.6-20.0)	3.9 (2.4-6.3)	<.001	2.3 (1.4-3.7)	.002
Very high	61/180 (33.9)	62.8 (48.8-80.7)	15.2 (9.2-25.3)	<.001	7.0 (4.0-12.1)	<.001
Total	374/5493 (6.8)	11.1 (10.0-12.3)	NA	NA	NA	NA

Abbreviation: HR, hazard ratio; NA, not applicable; NLST, National Lung Screening Trial; PLCO, Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial.

[a] Hazard ratios are adjusted for 9 chest radiograph findings (lung nodule, major atelectasis, pleural plaque or effusion, lymphadenopathy, chest wall or bony lesion, chronic obstructive pulmonary disease or emphysema, lung opacity, cardiomegaly or other cardiovascular abnormality, and lung fibrosis) and 10 risk factors (age, sex, smoking category, diabetes, hypertension, obesity, underweight, and previous myocardial infarction, stroke, and cancer).

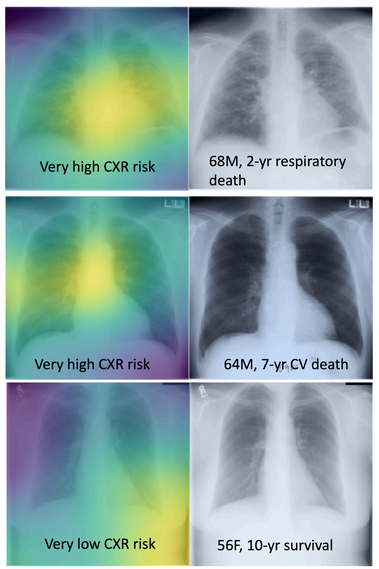
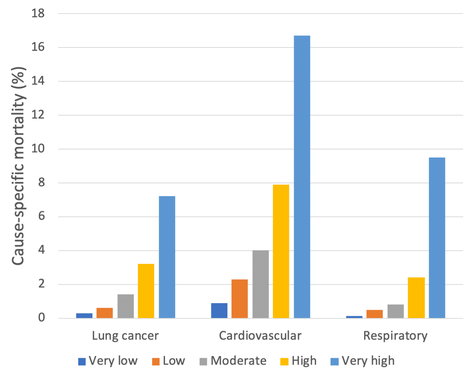


Figure 3: Gradient-weighted Class Activation Maps (Grad-CAM) localize anatomy contributing to CXR-Risk

Figure 4: Cause-specific mortality by CXR-Risk



Statistical Analysis

- Primary outcome: mortality over 12 (PLCO) and 6 (NLST) years.
- Secondary: cause-specific mortality
- Ordinal CXR-Risk score (very low, low, moderate, high, very high) based on probability centile in development dataset.
- Survival analysis: KM curves and hazard ratios
- Adjustment for radiologists' findings (lung nodule) and risk factors (age, sex, smoking)
- Test-retest reliability in 573 with repeat CXR

Results

- Very high risk had 53.0% (PLCO) and 33.9% (NLST) mortality, 18- and 15- fold higher than the very low risk (unadjusted hazard ratios PLCO: 18.3 (14.5-23.2); NLST: 15.2 (9.2-25.3), both p<0.001 **(Fig 2)**.
- CXR-Risk robust to adjustment for radiologists' findings and risk factors (adjusted hazard ratios PLCO: 4.8 (95% CI 3.6-6.4); NLST: 7.0 (4.0-12.1), p<0.001 **Table)**.
- Excellent test-retest reliability with ICC 0.89.
- Class activation maps highlight relevant anatomy: heart, sternal wires, breast and waist (**Fig 3**).
- Comparable results across age/sex strata, for lung cancer death and for non-cancer cardiovascular and respiratory death (**Fig 4).**

Conclusion

Based on a single chest radiograph, the CXR-Risk CNN stratified the risk of long-term mortality. High risk individuals may benefit from prevention, screening and lifestyle interventions.

Lu MT et al. *JAMA Open* 2019;2(7):e197416.
https://t.co/Vz6O3kGgxm