Novel Machine Learning Algorithm Predicts All-cause Mortality in the National Lung Screening Trial

Jennifer M. Wang[1], Swaraj Bose[2], Susan Murray[2], Wassim W. Labaki[1], Kevin R. Flaherty[1], MeiLan K. Han[1], Charles R. Hatt[3], Justin M. Oldham[1]

1. Division of Pulmonary and Critical Care Medicine, University of Michigan, Ann Arbor, MI, 2. Department of Biostatistics, University of Michigan, 3. Imbio Inc., Minneapolis, Minnesota

Rationale

- There has been widespread interest in applying machine learning algorithms to chest computed tomography (CT) to improve early detection of interstitial lung disease (ILD) in at-risk populations.
- Computer-Aided Lung Informatics for Pathology Evaluation and Ratings (CALIPER) quantifies key features of fibrotic lung disease using traditional machine learning techniques (i.e. nearest neighbor).
- We compared the ability of **CALIPER** and a novel deep-learning based software, **DeepLTA** (Imbio Inc) to assess risk of mortality in current and former smokers in the National Lung Screening Trial (NLST).

Study Population and Methods

- NLST enrolled 12,247 current and former smokers 55-74 years old with a minimum 30-pack-year smoking history at 33 US sites and available data for multivariable modeling over 6 years of follow-up
- Combined fibrosis score: reticular opacity and honeycombing
- Total % ground glass and reticular opacity, honeycombing and combined fibrosis score were generated using CALIPER and DeepLTA
- The outcomes studied included incident lung cancer, lung cancer mortality, and all-cause mortality

Results

Table 1: Subject Demographics

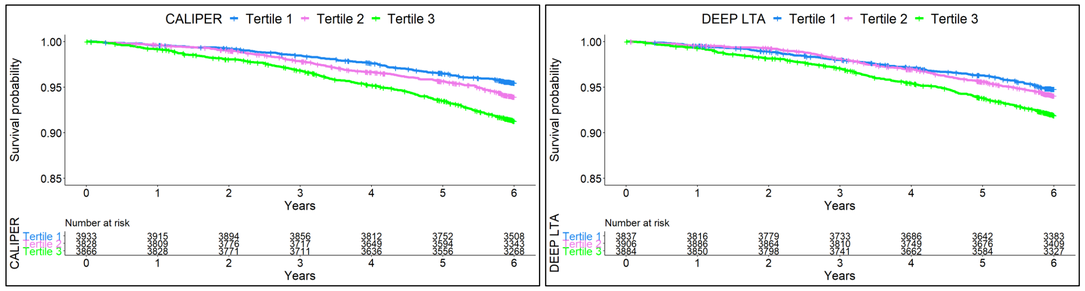
	Overall (N=11627)
Age	61.4 (5.0)
Race	
Hispanic or other	408 (3.5%)
Non-Hispanic Black	631 (5.4%)
Non-Hispanic White	10588 (91.1%)
Male	6861 (59.0%)
BMI, kg/m2	28.0 (5.0)
Smoking pack years	56.4 (23.9)
Quit years	3.8 (5.0)
Personal history of cancer	476 (4.1%)
Family history of lung cancer	2591 (22.3%)

Data are presented as means (standard deviation) for continuous variables and counts (percentages) for categorical variables.

Results

Figure 1: Kaplan Meier Curves of All-cause Mortality by Tertiles of Combined Fibrosis Score

CALIPER **DeepLTA**



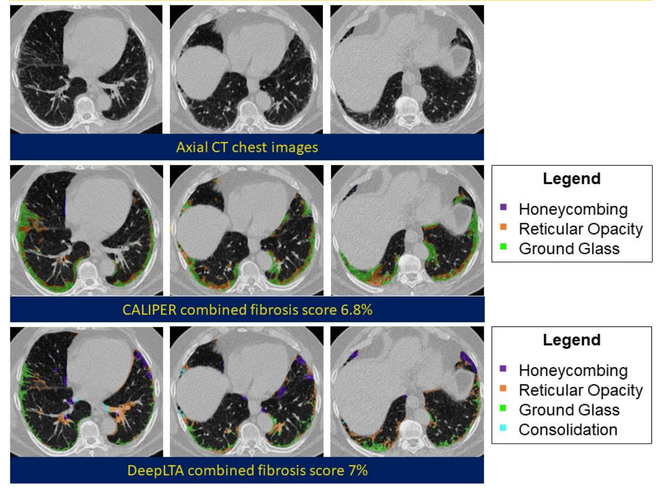
Multivariable Cox Regression Models of All-cause Mortality

- Model was adjusted for: age, race, highest education attained, gender, BMI, tobacco use (intensity and duration), time since smoking cessation, percent low attenuation area, and personal history of cancer and family history of lung cancer.

All-Cause Mortality								
	DeepLTA				CALIPER			
	Hazard Ratio, HR	95% Confidence Interval, CI	p value	C statistic	Hazard Ratio, HR	95% Confidence Interval, CI	p value	C statistic
Ground Glass Opacity	1.03	1.01 – 1.06	< 0.01	0.68	1.02	1.01 - 1.03	< 0.01	0.68
Reticular Opacity	1.17	1.10 – 1.25	< 0.01	0.68	1.05	1.03 - 1.08	< 0.01	0.68
Honeycombing	1.08	1.01 – 1.15	0.02	0.68	2.93	1.86 - 4.62	< 0.01	0.68
Combined Fibrosis Score	1.10	1.06 – 1.15	< 0.01	0.68	1.05	1.03 – 1.07	< 0.01	0.68

Results

Figure 2: CALIPER vs DeepLTA Images



Discussion

- For both DeepLTA and CALIPER an increased risk of death was observed for each percent increase in ground glass opacity, reticular opacity, honeycombing and the combined fibrosis score.
- Associations with lung cancer mortality did not reach statistical significance. Quantitative measures of fibrosis were not associated with incident lung cancer.
- Early, automated detection of ILD in a large at-risk screening population allows for timely follow-up and consideration of anti-fibrotic treatment

Conclusions

Both CALIPER and the novel deep-learning algorithm DeepLTA identified features of ILD on LDCT that were associated with risk of all-cause mortality in an at-risk screening population of current and former smokers.

Acknowledgements

The authors thank the National Cancer Institute for access to data collected by the NLST. The statements contained herein are solely those of the authors and do not represent or imply concurrence or endorsement by the National Cancer Institute.