

# Integration of Clinical Risk Factors Improves Lung Cancer Risk Stratification by Sybil Radiomic Machine Learning with Screening CT



Benjamin Zollinger, M.D.,<sup>1</sup> Annie Son,<sup>2</sup> Sriya Yalamanchili,<sup>1</sup> Jared Miller,<sup>2</sup> Rithvik Gabbireddy,<sup>1</sup> Kimberly Pullen,<sup>1</sup> Hongkun Wang, Ph.D.,<sup>3</sup> Melanie Subramanian, M.D.,<sup>1</sup> Simran Randhawa, M.D.,<sup>1</sup> Michael J. Weyant, M.D.,<sup>1</sup> Amit K. Mahajan, M.D.,<sup>1</sup> Kei Suzuki, M.D.<sup>1</sup>

<sup>1</sup> Division of Thoracic Surgery, Inova Schar Cancer Institute, Fairfax, VA; <sup>2</sup> University of Virginia School of Medicine, Charlottesville, VA; <sup>3</sup> Department of Biostatistics, Bioinformatics and Biomathematics, Georgetown University School of Medicine, Washington, DC.

## Background

- Lung cancer (LC) screening is performed with annual low-dose CT (LDCT) scan for those meeting current USPSTF guideline criteria
- Further workup for potential LC is guided by the Lung-RADS system; however, the respective Lung-RADS classifications can be inconsistent in their likelihood predictions for cancer.<sup>1-3</sup>
- Sybil is an open source, radiomic-based deep learning software developed in 2023 to generate individualized annual LC risk scores from a single LDCT scan.<sup>4</sup>
- Sybil has been trained and validated on LDCT imaging, and it generates risk scores for the succeeding six years from one imaging study.<sup>5</sup>
- However, its risk modeling only uses radiomic imaging features and does not incorporate other clinical factors or context

## Objective

- The study investigates whether the integration of selected clinical risk factors can improve the LC risk prediction of Sybil when compared to the radiomic risk score alone

## Methods

### Study design and population:

- Retrospective cohort study of patients undergoing guideline-directed LC screening with LDCT from 1/1/21 to 12/31/22 at across our single health system
- Patients were classified as LC-positive (LC+) or LC-negative (non-LC)

### Data collection:

- Clinical data were extracted from the electronic medical record, including demographics, known LC risk factors (smoking status, pack-years, COPD, emphysema, personal cancer history, family history of LC), and initial Lung-RADS score at their first LDCT scan
- Radiomic risk scores were generated using Sybil for all LDCT images

### Statistical Analyses

- Sybil risk scores and clinical factors were compared between non-LC and LC+ patients using two sample t-tests, Chi-squared, or Fischer's Exact tests
- Multivariable logistic regression was performed to determine LC predictors
- The areas under the receiver operating characteristics curve (AUC of ROC) were generated to assess model discrimination
- Two models were compared:
  - Sybil alone
  - Sybil + clinical risk factors
- DeLong's test was used to compare model performance

**Table 1.** Univariate comparison of clinical risk factors between LC+ and non-LC patients

	Non- LC (N = 1462)	LC+ (N = 24)	p-value
<b>Mean Age (years)</b>	64.0	67.2	<b>0.024</b>
<b>Pack-year history (pack-years)</b>	44.2	52.8	0.095
<b>Sex, n (%)</b>			0.054
Male	835 (57.1%)	9 (37.5%)	
Female	627 (42.9%)	15 (62.5%)	
<b>Current v. Former Smoker, n (%)</b>			0.63
Current	680 (46.5%)	9 (40.0%)	
Former	774 (53.0%)	15 (56.7%)	
<b>Emphysema, n (%)</b>			<b>&lt; 0.0001</b>
Not present	990 (62.2%)	4 (16.7%)	
Present	553 (37.8%)	20 (83.3%)	
<b>COPD, n (%)</b>			<b>0.0003</b>
No diagnosis	993 (67.9%)	8 (33.3%)	
Diagnosed	469 (32.1%)	16 (66.7%)	
<b>Family Hx of Lung Cancer, n (%)</b>			0.11
No family history	1306 (89.3%)	19 (79.2%)	
Family history of lung cancer	157 (10.7%)	5 (20.8%)	

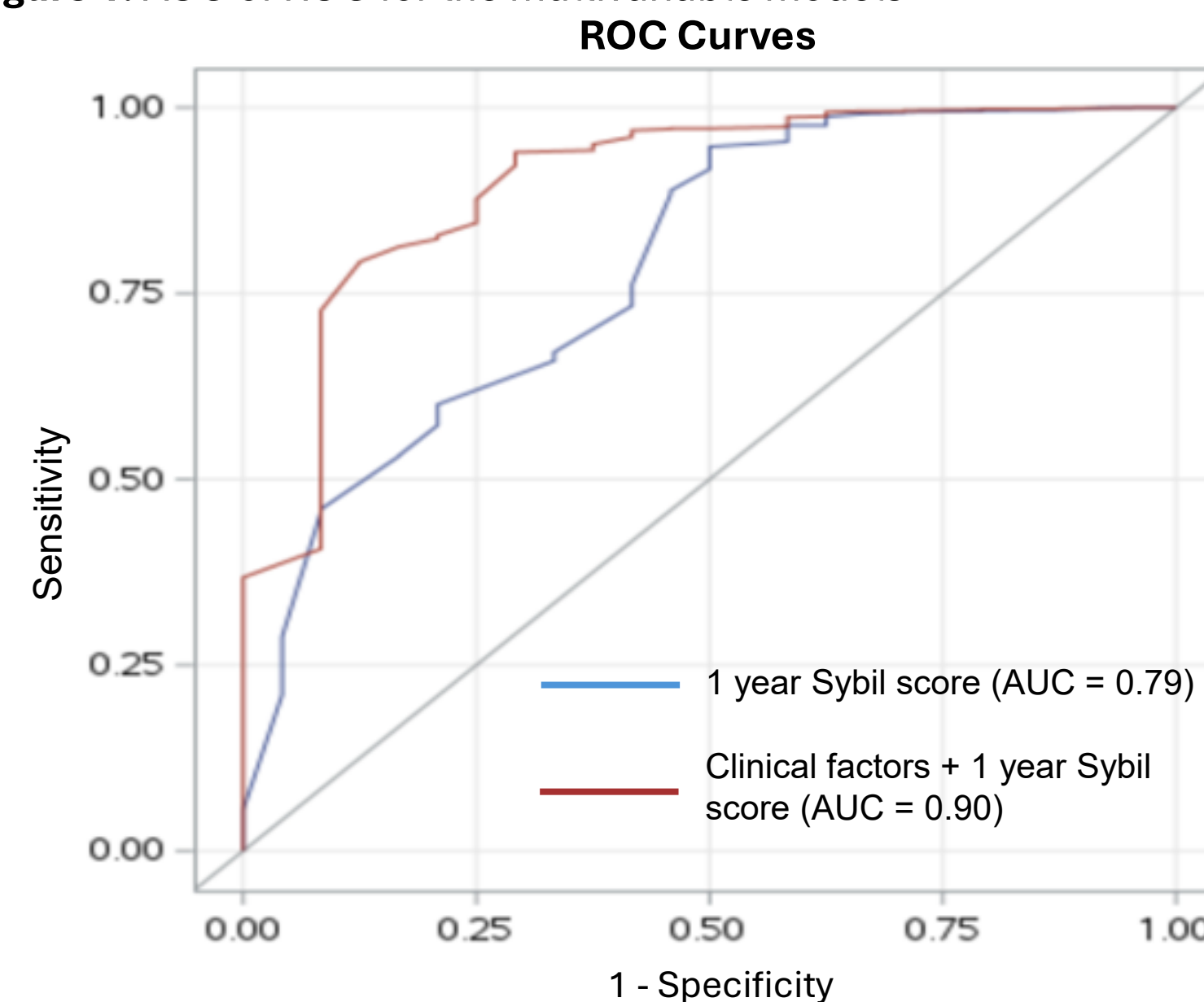
**Table 2.** Mean yearly Sybil scores for non-LC patients versus LC+ patients

	Non- LC (N = 1462)	LC+ (N = 24)	p-value
<b>Sybil score 1-year, Mean (SD)</b>	0.01 (0.03)	0.12 (0.16)	<b>&lt;0.0001</b>
<b>Sybil score 2-year, Mean (SD)</b>	0.02 (0.04)	0.16 (0.22)	<b>&lt;0.0001</b>
<b>Sybil score 3-year, Mean (SD)</b>	0.03 (0.05)	0.19 (0.22)	<b>&lt;0.0001</b>

**Table 3.** Significant predictors of LC on multivariable logistic regression

	Odds Ratio	95% Confidence Limits	p-value
Presence of emphysema	23.6	4.80 — 116.6	0.0001
Family history of lung cancer	3.98	1.34 — 11.8	0.013
Sybil score 1-year	1.018	1.013 — 1.023	<0.0001

**Figure 1.** AUC of ROC for the multivariable models



## Results

- Cohort: N = 1486 ; LC+: 24 patients (1.6%)
- Lung-RADS of LC+ patients:
  - Lung-RADS 1-3: 7 (29.2%)
  - Lung-RADS 4: 17 (70.8%)
- On univariate analysis, LC+ patients were older and more likely to have emphysema and COPD (**Table 1**)
- Sybil risk scores were significantly higher in LC+ patients across all six years ( $p < 0.0001$  for all years) (**Table 2**)
- On multivariable logistic regression, emphysema (OR 23.6), a family history of LC (OR 3.98), and Sybil score (OR 1.02) were independent predictors of LC+ (**Table 3**)
- Model performance significantly improved with clinical integration:
  - Sybil alone: AUC 0.79
  - Sybil + clinical factors: AUC 0.90 ( $p = 0.026$ ) (**Figure 1**)

## Conclusions

- A substantial proportion of LC occurred in non-Lung-RADS 4 patients, highlighting limitations of current radiographic stratification.
- Clinical factors such as emphysema and family history of LC can enhance identification of high-risk patients.
- Integration of certain clinical risk factors with Sybil risk scores significantly improves lung cancer risk prediction beyond using only radiomic data alone
- Integrated risk modeling may enable more personalized screening and follow-up strategies, particularly for patients with lower-risk imaging findings.

## References

- American College of Radiology Committee on Lung-RADS®. Lung-RADS assessment categories 2022.
- Chung K, Jacobs C, Scholten ET, et al. Malignancy estimation of lung-RADS criteria for subsolid nodules on CT: Accuracy of low and high risk spectrum when using NLSST nodules. *Eur Radiol.* 2017;27(11):4672–4679.
- Silvestri GA, Goldman L, Tanner NT, et al. Outcomes from more than 1 million people screened for lung cancer with low-dose CT imaging. *Chest.* 2023;164(1):241–251.
- Mikhael PG, Wohltend 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.* 2023;41(12):2191–2200.
- Lee JH, Chae KJ, Lu MT, et al. External testing of a deep learning model for lung cancer risk from low-dose chest CT. *Radiology.* 2025;316(2):e243393.

## Study Contact:

- Kei Suzuki, MD; kei.suzuki@inova.org