Sybil:
A model for lung cancer risk prediction

Mark Kenny and Adam Field

Lung cancer screening criteria

Age

Smoking history

Introduction

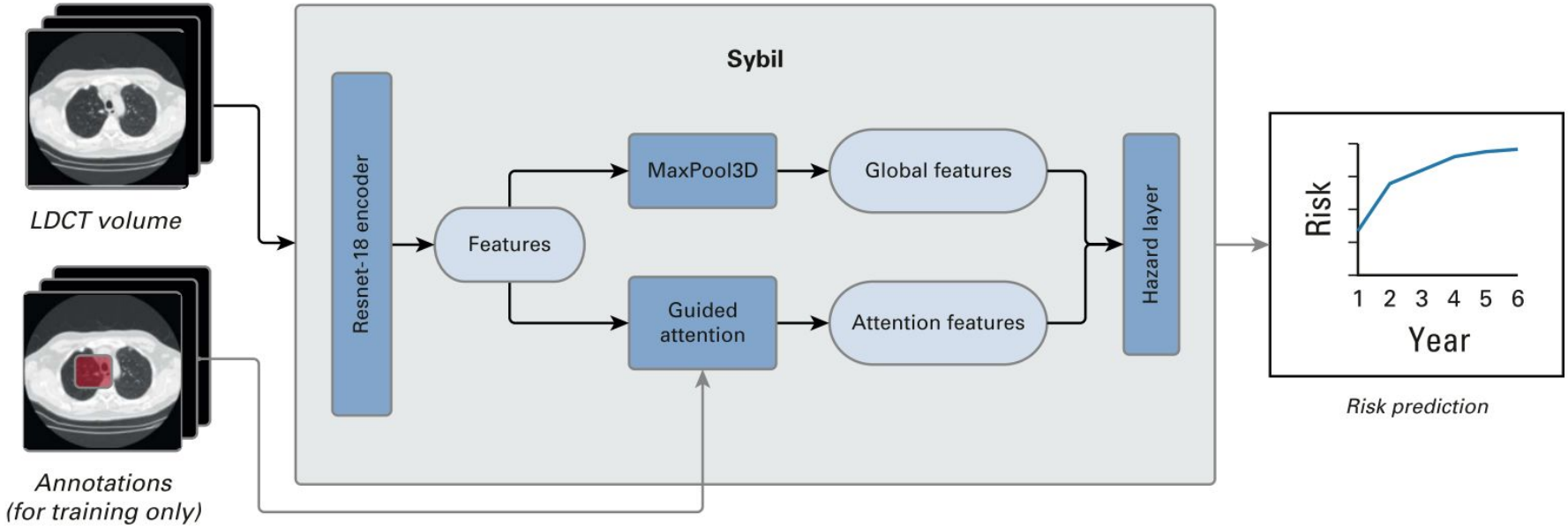
Previous Models

- 1-2 year risk prediction
- Require extra clinical data
- Require patient demographic data

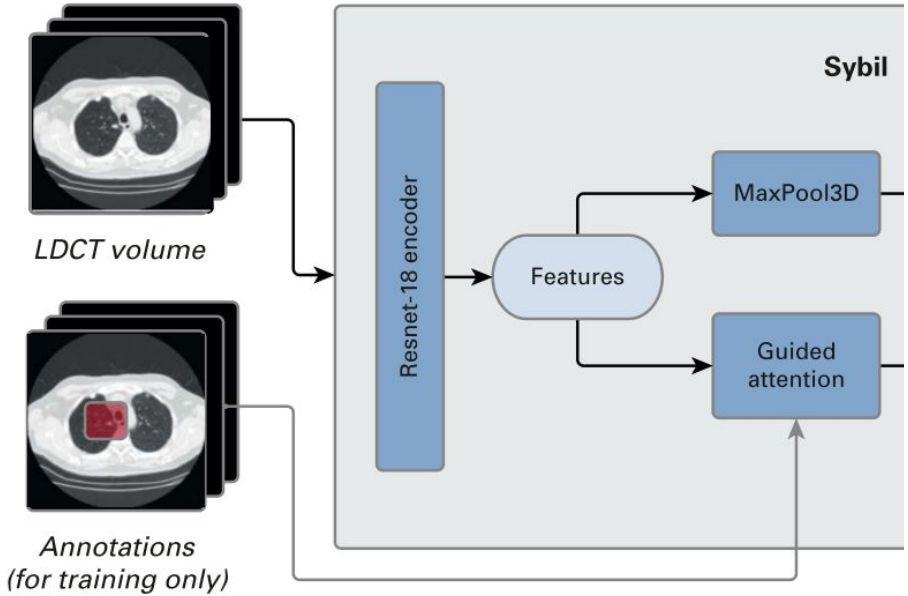
Sybil

- 1-6 year risk prediction
- No extra data needed

Architecture



Architecture



Max Pooling Layer

- Global feature recognition

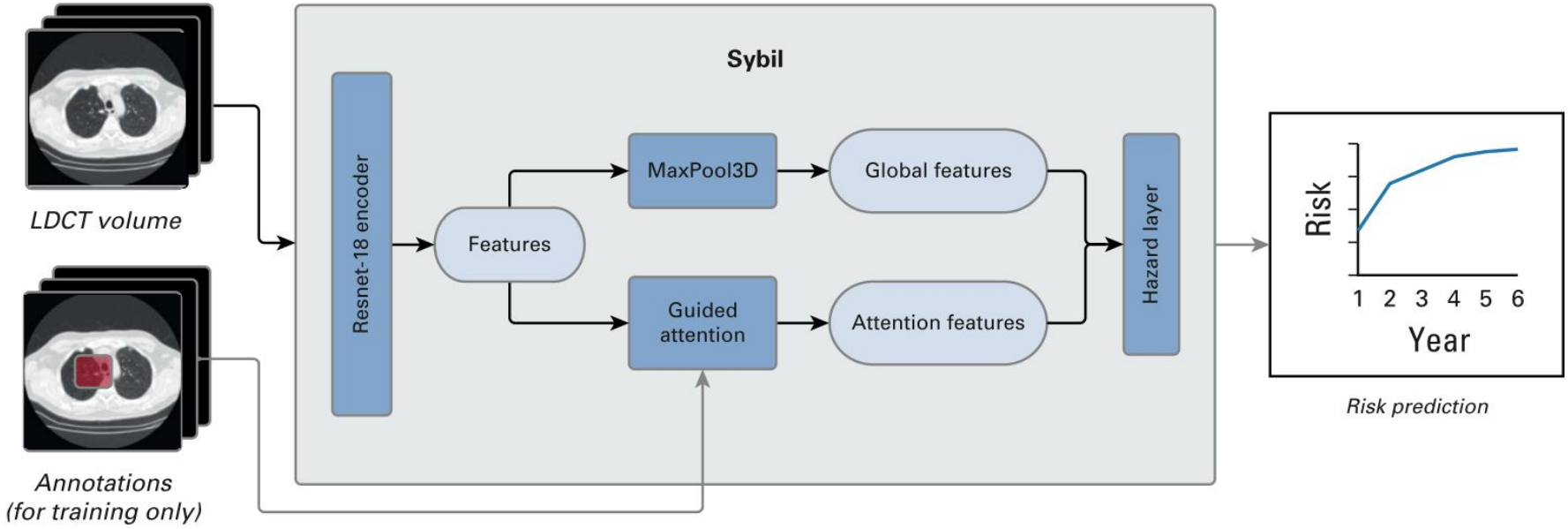
Guided Attention Layer

- Local feature recognition

Guided by

- Left vs. Right lung
- Bounding-box annotations

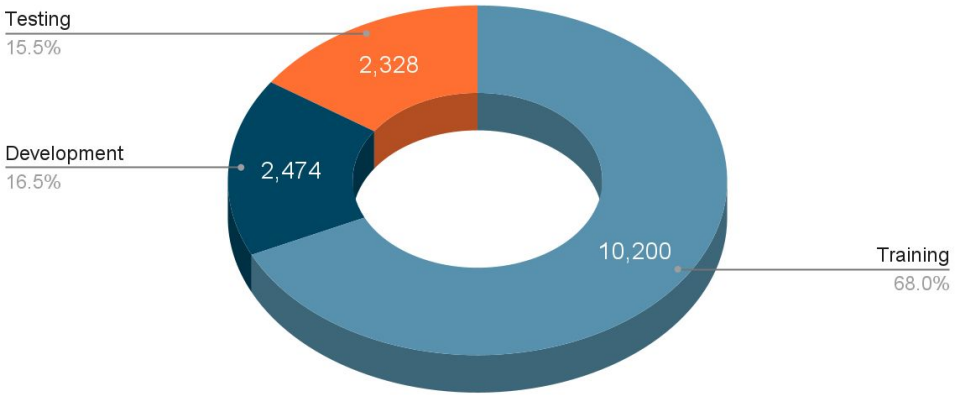
Architecture



Network Training

Dataset

15,000 Patients



- Up to 3 scans per patient
- All scans randomized
- “Positive” = lung cancer confirmed within 6 years of chest scan

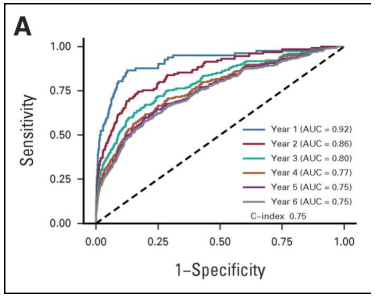
Results

- The NLST testing dataset had 6282 LDCTs in the test set
- No image annotation or clinical information was provided
- The model maintained performance across sex, age and smoking history subgroups

TABLE A2. Sybil's Future Lung Cancer Predictions Per Year in the National Lung Screening Trial Test Set, by Clinical Subgroups

Patient Groups	1-Year Risk, AUC (95% CI)	2-Year Risk, AUC (95% CI)	3-Year Risk, AUC (95% CI)	4-Year Risk, AUC (95% CI)	5-Year Risk, AUC (95% CI)	6-Year Risk, AUC (95% CI)	C-Index (95% CI)
Age, years							
50-60	0.92 (0.87 to 0.99)	0.88 (0.83 to 0.94)	0.79 (0.72 to 0.87)	0.74 (0.66 to 0.82)	0.70 (0.62 to 0.79)	0.70 (0.61 to 0.79)	0.70 (0.62 to 0.79)
60-70	0.92 (0.87 to 0.99)	0.86 (0.80 to 0.92)	0.82 (0.76 to 0.88)	0.80 (0.75 to 0.86)	0.80 (0.75 to 0.85)	0.78 (0.73 to 0.83)	0.78 (0.73 to 0.83)
Sex							
Male	0.94 (0.91 to 0.97)	0.86 (0.81 to 0.91)	0.80 (0.75 to 0.86)	0.77 (0.72 to 0.82)	0.75 (0.70 to 0.80)	0.74 (0.69 to 0.80)	0.74 (0.69 to 0.79)
Female	0.88 (0.80 to 0.99)	0.86 (0.78 to 0.94)	0.79 (0.71 to 0.87)	0.77 (0.69 to 0.85)	0.76 (0.69 to 0.83)	0.75 (0.68 to 0.83)	0.75 (0.68 to 0.82)
Race[a]							
White	0.91 (0.87 to 0.96)	0.86 (0.81 to 0.90)	0.80 (0.75 to 0.85)	0.77 (0.72 to 0.81)	0.75 (0.71 to 0.80)	0.74 (0.70 to 0.79)	0.74 (0.70 to 0.79)
Black or African American	0.99 (0.98 to 1.0)	0.95 (0.89 to 1.0)	0.93 (0.85 to 1.0)	0.84 (0.67 to 1.0)	0.83 (0.64 to 1.0)	0.83 (0.65 to 1.0)	0.83 (0.66 to 1.0)
Asian	0.97 (0.94 to 1.0)	0.95 (0.91 to 1.0)	0.77 (0.55 to 1.0)	0.77 (0.55 to 1.0)	0.74 (0.54 to 1.0)	0.70 (0.49 to 0.97)	0.71 (0.51 to 0.95)
Current smoker							
Yes	0.89 (0.82 to 0.99)	0.84 (0.78 to 0.92)	0.77 (0.70 to 0.85)	0.75 (0.68 to 0.81)	0.72 (0.66 to 0.79)	0.71 (0.65 to 0.77)	0.71 (0.65 to 0.77)
No	0.93 (0.90 to 0.97)	0.87 (0.83 to 0.92)	0.82 (0.77 to 0.88)	0.79 (0.73 to 0.85)	0.79 (0.73 to 0.85)	0.78 (0.72 to 0.85)	0.78 (0.72 to 0.84)
Smoking duration, years							
< 40	0.96 (0.94 to 0.99)	0.89 (0.84 to 0.94)	0.84 (0.79 to 0.90)	0.80 (0.73 to 0.87)	0.79 (0.72 to 0.86)	0.78 (0.71 to 0.86)	0.78 (0.72 to 0.85)
> 40	0.88 (0.82 to 0.96)	0.83 (0.77 to 0.90)	0.76 (0.70 to 0.83)	0.73 (0.67 to 0.79)	0.71 (0.66 to 0.77)	0.70 (0.65 to 0.76)	0.70 (0.65 to 0.76)

Abbreviations: AUC, area under the curve; C-index, concordance index.
[a]Results for the race categories American Indian or Alaskan Native and Native Hawaiian or other Pacific Islander are omitted as they did not contain enough cancers to provide CIs.



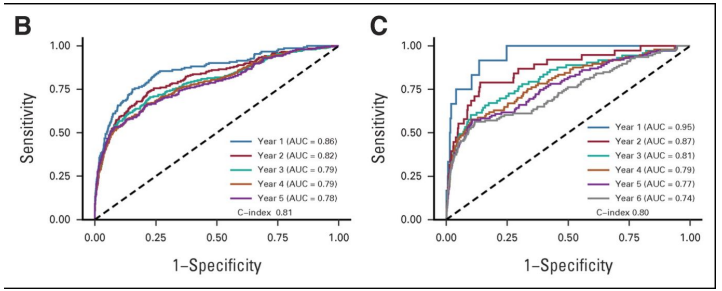
Data Set	1-Year Risk, AUC (95% CI)	2-Year Risk, AUC (95% CI)	3-Year Risk, AUC (95% CI)	4-Year Risk, AUC (95% CI)	5-Year Risk, AUC (95% CI)	6-Year Risk, AUC (95% CI)	C-Index (95% CI)
NLST	0.92 (0.88 to 0.95)	0.86 (0.82 to 0.90)	0.80 (0.77 to 0.84)	0.77 (0.73 to 0.81)	0.75 (0.72 to 0.79)	0.75 (0.72 to 0.78)	0.75 (0.72 to 0.78)

Results

- The model was then applied to MGH and CGMH testing datasets
- MGH had 8821 LCDTs, 169 confirmed cancers
- CGMH had 12280 LCDTs, 101 confirmed cancers
- CGMH did not require a positive smoking history to access LCDTs
- Sybil had a similar risk prediction in the two sets like in the NLST test set

TABLE A3. Demographics of Independent External Validation Data Sets From MGH (n = 4,954 patients) and CGMH (n = 10,567 patients)

	MGH Test Set		CGMH Test Set	
Patient Group	**Total, No. (%)**	**Future Cancers Diagnosed, No. (%)**	**Total, No. (%)**	**Future Cancers Diagnosed, No. (%)**
No. of examinations	8,821 (100.0)	255 (100.0)	12,280 (100.0)	126 (100.0)
Age cohort, years				
< 50	9 (0.1)	NA	4,296 (35.0)	24 (19.1)
50-60	2,044 (23.2)	63 (24.7)	4,258 (34.7)	42 (33.3)
60-70	4,563 (51.7)	139 (54.5)	2,878 (23.4)	33 (26.2)
70-80	2,155 (24.4)	52 (20.4)	722 (5.9)	19 (15.1)
> 80	49 (0.6)	1 (0.4)	126 (1.0)	8 (6.3)
Sex				
Female	4,159 (47.1)	151 (59.2)	5,146 (41.9)	67 (53.2)
Male	4,662 (52.9)	104 (40.8)	7,134 (58.1)	59 (46.8)
Race				
White	6,696 (75.9)	215 (84.3)	0 (0.0)	0 (0.0)
Black or African American	262 (3.0)	10 (3.9)	0 (0.0)	0 (0.0)
Asian	175 (2.0)	6 (2.4)	12,280 (100.0)	126 (100.0)
American Indian or Alaskan Native	14 (0.2)	1 (0.4)	0 (0.0)	0 (0.0)
Native Hawaiian or other Pacific Islander	3 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)



Data Set	1-Year Risk, AUC (95% CI)	2-Year Risk, AUC (95% CI)	3-Year Risk, AUC (95% CI)	4-Year Risk, AUC (95% CI)	5-Year Risk, AUC (95% CI)	6-Year Risk, AUC (95% CI)	C-Index (95% CI)	CI)
MGH	0.86 (0.82 to 0.90)	0.82 (0.77 to 0.86)	0.79 (0.75 to 0.84)	0.79 (0.74 to 0.83)	0.78 (0.73 to 0.83)	NA	0.81 (0.77 to 0.85)	.85)
CGMH	0.94 (0.91 to 1.00)	0.87 (0.81 to 0.95)	0.81 (0.75 to 0.88)	0.79 (0.73 to 0.87)	0.77 (0.71 to 0.83)	0.74 (0.66 to 0.81)	0.80 (0.75 to 0.86)	.86)

Understanding Sybil's Predictions

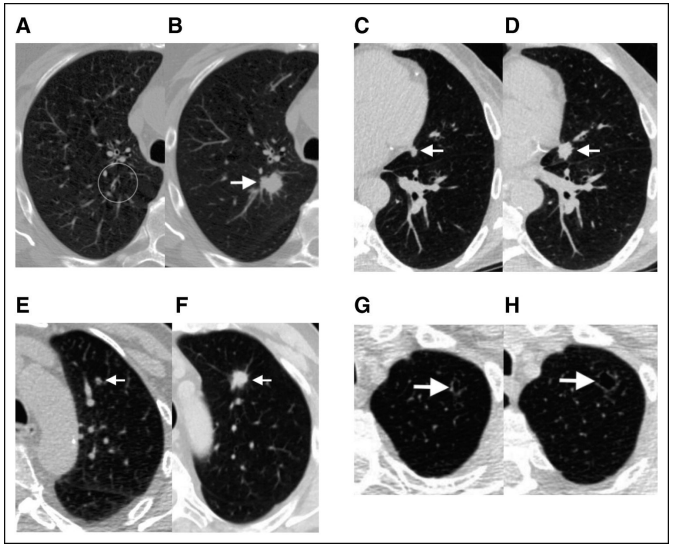
- The researchers analyzed Sybil and whether its prediction depended on the presence of radiographically visible cancerous lung nodules within the LCDTs
- By excluding cases where lung cancers had developed in the exact spot of lung nodules, the researchers thus analyzed the predictive capability of Sybil
- This decreased Sybil's performance to an extent, though it was nonetheless still capable even in the absence of cancerous nodules

Comparing it to Lungs-RAD

- Lungs-RAD is the clinical standard of care for standardizing lung cancer screening
- The researchers used an NLST set of 4201 LCDTs
- Lungs-RAD had an FPR (False Positive Rate of 0.10) while Sybil had an FPR of 0.08
- In considering only baseline LCDTs with no other information, Lungs-RAD had an FPR of 0.14 while Sybil had an FPR of 0.08

Clinical Application

- Identifying cases of missed cancers due to human error
- Decrease follow-up interval for patients
- Identify and give highest risk patients priority



Limitations

- Lack of comparator models to truly assess Sybil's performance
- Do not include sufficient amounts of data for Black, Hispanic race groups, is thus not broadly applicable
- Inconclusive about Sybil's ability of identifying lung cancer in nonsmokers due to external conditions

Discussion