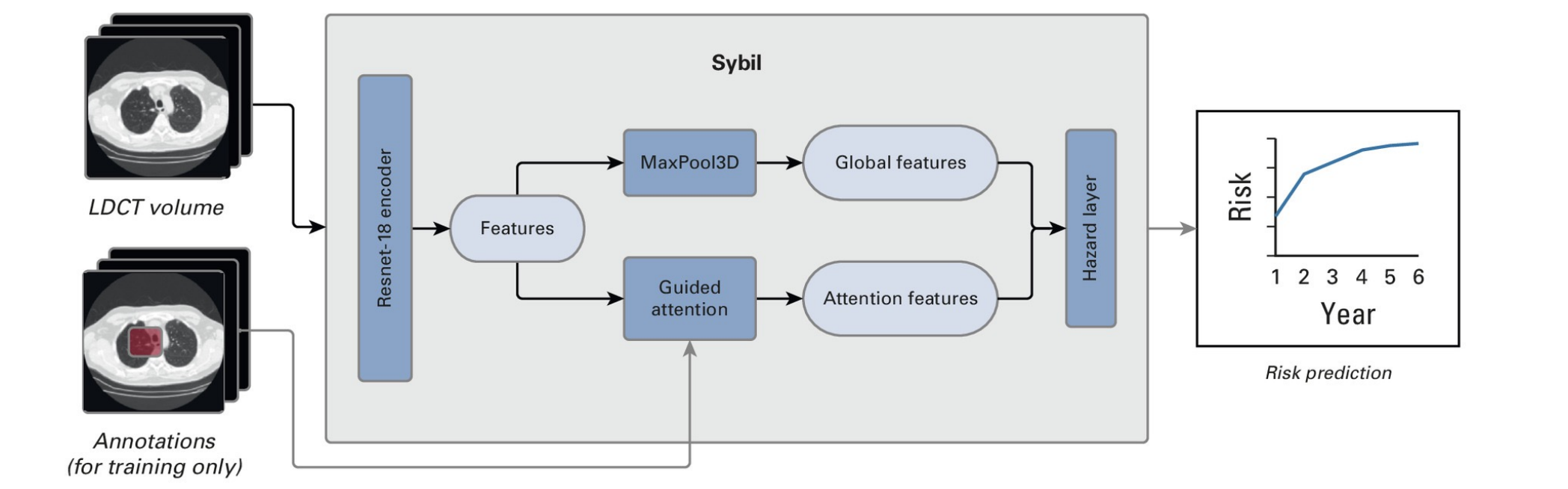


INTRODUCTION



Mikhael PG, Wohlwend J, Yala A, Karstens L, Xiang J, Takigami AK, Bourgoun PP, Chan P, Mirah S, Amayri W, Juan YH, Yang CT, Wan YL, Lin G, Sequist LV, Fintelmann FJ, Barzilay R. Sybil: A Validated Deep Learning Model to Predict Future Lung Cancer Risk From a Single Low-Dose Chest Computed Tomography. *J Clin Oncol*. 2023 Apr 20;41(12):2191-2200. doi: 10.1200/JCO.22.01345. Epub 2023 Jan 12. PMID: 36634294; PMCID: PMC10419602.

- Sybil is an AI tool to predict lung cancer risk from a single computed tomography (CT) scan
- We evaluated whether characteristics of nodules and patients among people with nodules can accurately predict the Sybil score
- Identified which characteristics are most strongly associated with the score

DATA & METHODS

Data sources: Baseline LDCT scans from the ILST Vancouver and Pan-Can lung cancer screening cohorts

Outcome: 6-year Sybil risk score

Features:

- **Patient:**
Age, sex, smoking status, smoking duration, pack-years, BMI, education, family history of cancer, number of nodules, and emphysema
- **Nodule features:**
Mean nodule diameter, nodule type, spiculation, location, segment site, segment status
- Nodules were ranked from highest to lowest risk using the Pan-Can malignancy probability; the highest-risk nodule was defined as the first-ranked nodule

DISCUSSION AND CONCLUSION

- Sybil scores are largely driven by the characteristics of the top-ranked (highest-risk) nodules. Patients' characteristics and lower-risk nodules contribute less
- Further studies should focus on improving the risk assessment tools for scans without nodules or with only low-risk nodules

MODEL & EVALUATION

Model Development

- A **Light-GBM regression model** was used to predict the 6-year Sybil score
- Data were randomly split into **70% training** and **30% testing** sets
- Hyperparameters were optimized using grid search and 5-fold cross-validation

Model Evaluation

- Model performance was evaluated using **MAE** and **RMSE**. The MAEs were 0.044 in the training set and 0.048 in the testing set (with corresponding RMSEs of 0.074 and 0.083), which suggest limited overfitting
- Feature importance was evaluated using gain-based **variable importance** and **SHapley Additive exPlanations (SHAP) values**. SHAP values were used to assess the vectorized impact of each feature regarding Sybil score predictions

REFERENCES

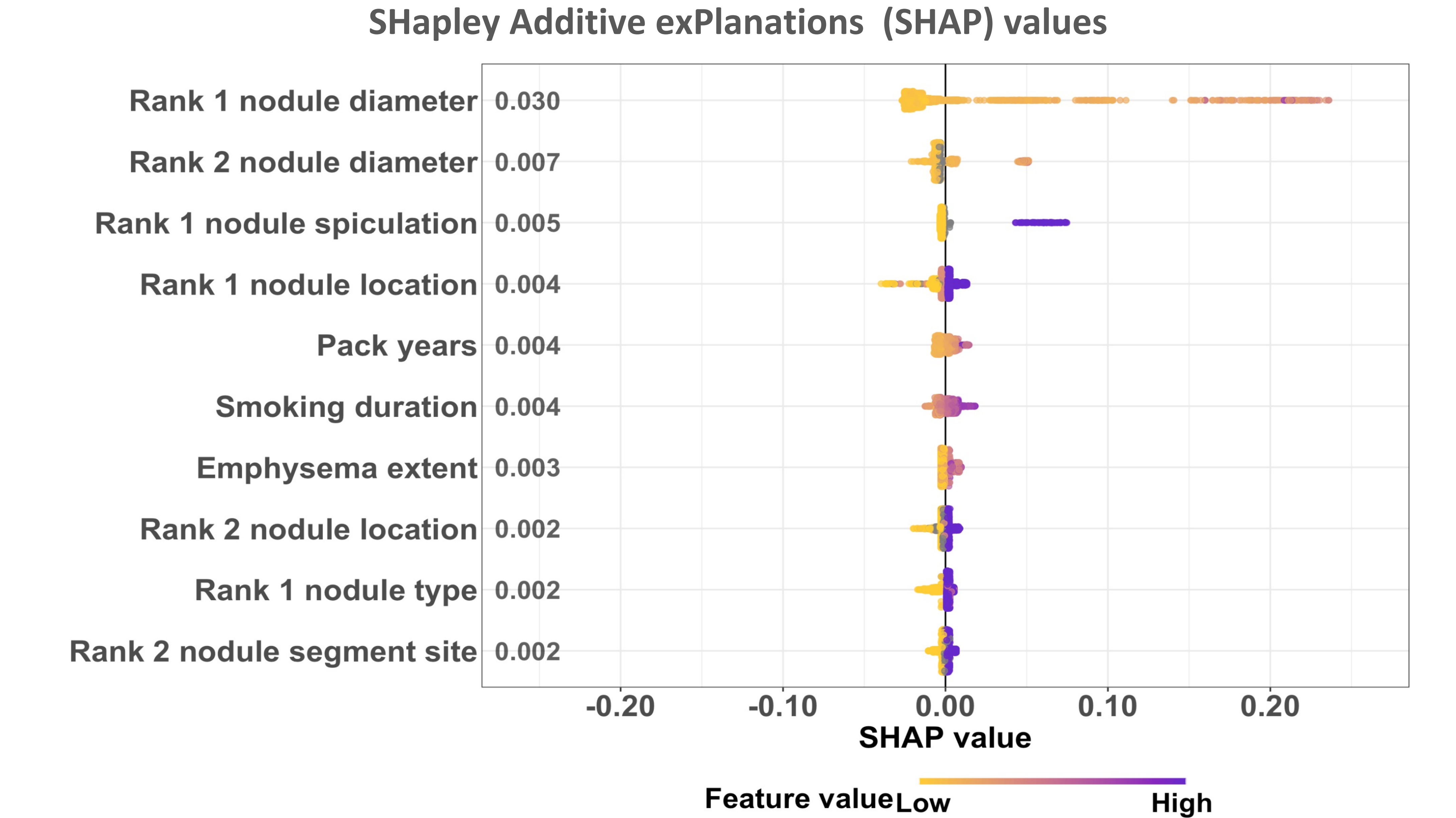
[1] Annette McWilliams, Martin C. Tammemagi, John R. Mayo, Heidi Roberts, Geoffrey Liu, Kam Soghrati, Kazuhiro Yasufuku, Simon Martel, Francis Laberge, Michel Gingras, Sukhinder Atkar-Khattra, Christine D. Berg, Ken Evans, Richard Finley, John Yee, John English, Paola Nasute, John Goffin, Serge Puksa, Lori Stewart, Scott Tsai, Michael R. Johnston, Daria Manos, Garth Nicholas, Glenwood D. Goss, Jean M. Seely, Kayvan Amjadi, Alain Tremblay, Paul Burrows, Paul MacEachern, Rick Bhatia, Ming-Sound Tsao, and Stephen Lam. Probability of cancer in pulmonary nodules detected on first screening ct. *New England Journal of Medicine*, 369(10):910–919, 2013.

[2] Yang Liu and Allan Just. SHAPforxgboost: SHAP Plots for 'XGBoost', 2020. R package version 0.1.0.

Artificial Intelligence (AI) Use Disclaimer
No generative AI tools were used in the preparation of this poster.

RESULTS

- A total of 3,164 individuals were included in the analysis, with a mean of 4 nodules per participant. Most nodules were solid (81.96%), followed by ground-glass opacity nodules (14.31%) and semi-solid nodules (3.73%)
- The Sybil score was predicted mainly by nodule-level features. Mean axial diameter of the first-ranked nodule was the dominant predictor, followed by the diameter of the second-ranked nodule, and other first-ranked nodule characteristics, including location and spiculation
- Patient-level variables, including emphysema extent, smoking duration, pack-years, age, and education, contributed less than the top-ranked nodule features



- Larger diameter of the first-ranked nodule was positively associated with higher predicted Sybil scores. For first-ranked nodules, spiculated ones had more positive SHAP values than non-spiculated ones, suggesting that the spiculation of the first ranked nodule is positively associated with predicted Sybil scores
- Nodule location also contributed to model predictions; subpleural nodules were associated with lower predicted sybil scores, whereas parenchymal nodules were associated with higher predicted sybil scores
- For nodule types, ground-glass opacity nodules were associated with lower predicted Sybil scores, while solid nodules were associated with higher predicted Sybil scores