

P1.061: Evaluation of Lung Cancer Risk Prediction Models in Diverse Populations and Development of Machine Learning Models

Yu Chen¹, Jiahao Lin¹, Xuyi Chen², Wenkai Huang³, Lan Wu¹, Zirui Luo¹, Sumei Cao¹, **Mengmeng Li¹**

¹State Key Laboratory of Oncology in South China, Guangdong Provincial Clinical Research Center for Cancer, Sun Yat-sen University Cancer Center, Guangzhou 510060, China, ²School of Public Health, Guangzhou Medical University, Guangzhou 511436, China, ³Office of Cancer Screening, National Cancer Center/National Clinical Research Center for Cancer/ Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, 100021, China

Correspondence: Dr. Mengmeng Li, Email: limm@sysucc.org.cn. Prof. Sumei Cao, Email: caosm@sysucc.org.cn

INTRODUCTION

- Current lung cancer (LC) screening criteria are inadequate for identifying high-risk individuals in Asian never-smokers and women.
- Most published LC risk models were developed in Western populations, and their performance in Chinese populations across subgroups defined by sex and smoking status remains unclear.
- We aimed to develop the Sex-Smoking-specific machine learning Lung Cancer prediction model (SSLC) to enhance the equity of LC screening.

METHODS

- Thirteen published LC risk models were externally validated in the Guangzhou screening cohort (N=64,738) across four subgroups defined by sex and smoking status.
- SSLC for LC incidence and mortality were developed using the China Kadoorie Biobank (N=512,255) for development and internal validation, and the Guangzhou screening cohort for external validation.
- The algorithms included random survival forest (RSF), gradient boosting survival analysis (GBSA), CoxNet, extreme gradient boosting with debiased binary cross-entropy (XGBBCE), and Weibull accelerated failure time (Weibull AFT).
- Performance was assessed by area under the receiver operating characteristic curve (AUC), Brier scores, and decision curve analysis.

DISCUSSION AND CONCLUSION

Actionable Take Away Message

- Sex-Smoking-specific machine learning Lung Cancer prediction model developed in large Chinese cohorts showed improved discrimination and calibration, whereas existing models showed marked variation in performance across subgroups.

Future Direction for Research

- These findings support the adoption of subpopulation-specific eligibility criteria for LC screening in China.

REFERENCES & ACKNOWLEDGEMENTS

REFERENCES

- Feng X, Guida F, Guenoun A, et al. Performance of Lung Cancer Risk Prediction Models in Different Racial and Ethnic Groups in the United States: Results From the Lung Cancer Cohort Consortium. *Ann Intern Med*. Published online June 30, 2026. doi:10.7326/ANNALS-25-03816
- Chen Z, Chen J, Collins R, et al. China Kadoorie Biobank of 0.5 million people: survey methods, baseline characteristics and long-term follow-up. *Int J Epidemiol*. 2011;40(6):1652-1666. doi:10.1093/ije/dyr120
- Wang F, Tan F, Shen S, et al. Risk-stratified Approach for Never- and Ever-Smokers in Lung Cancer Screening: A Prospective Cohort Study in China. *Am J Respir Crit Care Med*. 2023;207(1):77-88. doi:10.1164/rccm.202204-0727OC
- Callender T, Imrie F, Cebere B, et al. Assessing eligibility for lung cancer screening using parsimonious ensemble machine learning models: A development and validation study. *PLOS Medicine*. 2023;20(10):e1004287. doi:10.1371/journal.pmed.1004287
- Pan Z, Zhang R, Shen S, et al. OWL: an optimized and independently validated machine learning prediction model for lung cancer screening based on the UK Biobank, PLCO, and NLST populations. *eBioMedicine*. 2023;88. doi:10.1016/j.ebiom.2023.104443

WEB PLATFORM

To further access to our model, please view: <https://sslc.vercel.app/>, we have implemented the SSLC model as an interactive web-based platform for practical use.

ACKNOWLEDGEMENTS

The most important acknowledgement is to the participants in the study and the members of the survey teams in each of the 10 regional centers in CKB, as well as to the project development and management teams based at Beijing, Oxford, and the 10 regional centers. We also gratefully acknowledge the participants, investigators, and coordinating staff of the Guangzhou screening cohort.

Artificial Intelligence (AI) Use Disclaimer

Artificial intelligence (AI) and AI-assisted tools were not utilized at any stage during the preparation, drafting, or proofreading of this abstract and the poster.

RESULTS

- Among the existing models, the best-performing models for LC incidence and mortality across the four subgroups (female non-smokers, female smokers, male non-smokers, and male smokers) were different (Figure 1).
- The machine learning algorithms with the largest AUCs for each subgroup were selected to construct SSLC. SSLC showed better or comparable discrimination as compared to the previously published models (Figure 2).
- In mortality prediction, SSLC achieved lower Brier scores than the published models in all strata; in incidence prediction, calibration was better in female smokers and broadly comparable in the remaining subgroups.
- Decision curve analysis further suggested that the SSLC provided clinical benefit across subgroups, particularly for mortality risk prediction.

Figure 1: Discrimination of Previously Published Lung Cancer Risk Prediction Models Across Different Subgroups

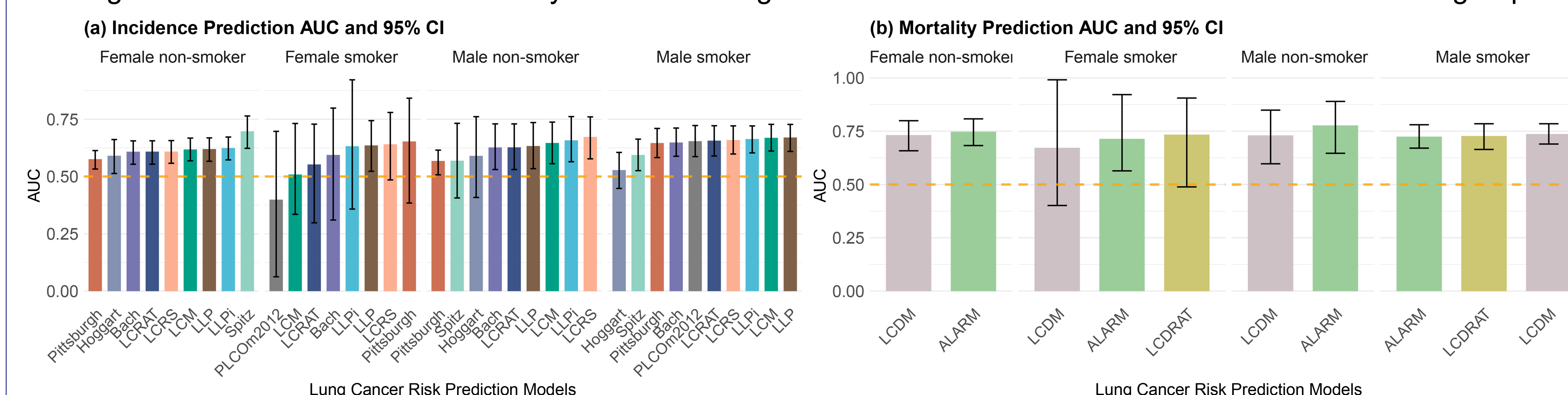


Figure 2: ROC Curves of SSLC and Previously Published Models Across Different Subgroups

