



Future lung cancer prediction using low-dose chest computed tomography

Imran Ashraf, Soojung Hur, Yongwan Park

Department of Information and Communication Technology, Yeungnam University, Gyeongsan-si, Republic of Korea

Correspondence to: Imran Ashraf, PhD. Department of Information and Communication Technology, Yeungnam University, 280 Daehak-ro, Gyeongsan-si 38541, Republic of Korea. Email: imranashraf@ynu.ac.kr.

Comment on: Mikhael PG, Wohlwend 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;41:2191-200.

Keywords: Lung cancer (LC); deep learning; computed tomography

Submitted Apr 06, 2023. Accepted for publication May 17, 2023. Published online May 23, 2023.

doi: 10.21037/tlcr-23-235

View this article at: <https://dx.doi.org/10.21037/tlcr-23-235>

Despite developments in surgery and chemotherapeutic approaches, lung cancer (LC) has been among the leading causes of death globally (1). Early diagnosis and timely treatment can be crucial to increase the probability of survival which becomes very difficult when clinical symptoms appear. Often diagnosed at advanced stages, LC becomes untreatable and leads to death.

For the most part, previous efforts focused on identifying those at a higher risk of LC. The emphasis has been on the clinical and demographic features, as well as, chest radiography. Controlled trials including the National Lung Screening Trial (NLST) and Dutch-Belgian Lung Cancer Screening Trial (NELSON) reported decreased mortality with LC screening (LCS) using low-dose computed tomography (LDCT). The efficacy of LCS using LDCT is established compared with radiography or no screening. Following the promising results of LDCT, the United States (US) recommends annual examinations for citizens aged 50 or higher (2). LCS faces a shortfall at the moment due to predominantly focusing on the screening of heavy smokers. With a large increase in cancer among never and lighter smokers, LCS needs to be broadened (3,4).

If the focus continues to be on smokers only for LCS, the gap can persist between the screen population and the disease population. Prioritizing individual assessments for predicting future LC risk is an attractive approach to overcoming this problem. However, the existing healthcare

system is facing many challenges that can hamper this solution. Lack of appropriate medical equipment and staff including radiologists and thoracic surgeons are a few of them. With the ongoing pandemic, the health system is overloaded, limiting its capability to keep up with LCS demand.

Artificial intelligence (AI) has made an impact in many fields including engineering, robotics, medicine, and healthcare. Further aided by the availability of big data and implantable sensors, AI has shown great promise for future healthcare applications. Coupled with LDCT images, AI approaches can provide potential solutions for predicting the future risk of LC. Contrary to traditional LCS approaches that rely on lung nodules, LDCT can hypothetically provide information beyond that to predict LC risk for 1 to 6 years in the future. Feature engineering is an important part of machine learning (ML) approaches and has a detrimental impact on models' predictive performance. Feature extraction transforms raw data into smaller groups of numerical values. Such values, though dimensionally reduced, represent the original dataset accurately. Feature extraction is very important, as different features from the same data have different significance to predict the label, and the choice of using a particular set of features can lead to different predictive performances of the same model. For several types of ML models, features are carefully crafted by human experts which limits their

wide application. On the contrary, ML models can also be used without such endeavor and can learn the underlying complex relationships provided a large dataset is given.

The study published in the *J Clin Oncol* by Mikhael *et al.* presents “*Sybil: a validated deep learning model to predict future lung cancer risk from a single low-dose chest computed tomography*”. The paper presents a deep learning approach based on LDCT images to predict the future risk of developing LC (5). The proposed approach is superior to several other similar approaches, as no other approach is available to predict future LC risk past 1 year using single LDCT scan.

The prime objective of this study is to test the hypothesis that LDCT images contain useful information, beyond lung nodules, concerning the prediction of future risk of LC. The impact of including lung nodule features is also analyzed. Two additional positive features of the proposed work are code availability and model validation. The use of two publicly available datasets is advantageous to replicate the results and advance further research in this direction.

LC prediction is performed on a single LDCT scan and multiple scans of an individual are treated as multiple subjects. LC-positive LDCT scans are those biopsy-confirmed within 6 years of LDCT scan. For model training, thoracic radiologists annotated the scans containing suspicious lesions using MD.AI software. Sybil, the proposed model, provides output as a set of six probability scores, corresponding to 1 to 6 years after LDCT.

The proposed model is evaluated using Uno’s concordance index and area under the receiver operating characteristic. The model proves to predict LC with high accuracy for the 1st year after the LDCT scan, and this accuracy decreases gradually for years 2 to 5. Moreover, the performance of Sybil against different subgroups like gender, age, smokers, etc. is retained. False positive rate (FPR) is also found to be less than Lung Imaging Reporting and Data System (Lung-RADS), considered a gold standard in the US.

Risk assessment of Sybil’s performance is also carried out using the addition and exclusion of radiological features. Additional experiments involve excluding scans containing visible lung nodules, those annotated by radiologists. A hampered performance is observed, thereby reducing the efficacy of the proposed approach.

The authors contrived the model based on LDCT scan only, ignoring clinical and radiological features, and show

a great potential to exceed existing models for predicting future risk of LC. Noteworthy, in addition to enhanced performance, is that radiologist annotation, demographic features, and clinical features are not needed. Additionally, the models predict LC by localizing the high-risk regions. The model does not spread the risk over the entire thorax and can locate specific regions at a high risk of cancer.

Existing approaches indicate challenges for LCS like inter-grader variability, high FPR, prediction for a limited number of years to incidence, and high false negative rate (FNR) (6,7). Mikhael *et al.*’s article overcomes the challenges of higher FPR and FNR and provides the prediction for 6 years after LDCT. Besides a predictive tool for the future risk of LC, the clinical application of the proposed approach is reducing the number of follow-up scans. Similarly, a biopsy is also not needed for patients with nodules who are predicted to be at low risk.

Sybil has shown its prospective role in LC risk prediction, yet, its generalizability has not been put to the test. Validation on two recent datasets indicates slightly different performances, though it shows predictive capabilities. The base dataset obtained in 2002, being old, requires a recent dataset to analyze how technological advancements in CT technology could affect the efficacy of the model. In addition, predominantly LDCT contains scan from white people, and the efficiency of the model using scans from other races need to be evaluated accordingly. LDCT scans of non-smokers are also needed to analyze the model’s efficacy, and so does population diversity for generalizability.

The study from Mikhael *et al.* and similar models from (6,7) show the potential of deep learning approaches for catering healthcare solutions, especially those involving medical images. With the lack of medical experts, the increased number of patients, and the burden of post-pandemic issues, the adoption of AI for automated diagnosis and prognosis of LC carries several benefits (8). However, embracing AI-based solutions not only requires legal frameworks but also needs a vote of confidence from the people. Designing and deploying AI solutions require datasets, both large and publicly available, which at the moment is not possible due to patient confidentiality and other legal issues.

LCS requires automated solutions, those involving no medical experts and physical visits to hospitals, however, such solutions require consensus from different stakeholders. With AI solutions embedded into health

care, faster and more accurate screening of LC patients is possible which can save countless lives.

Acknowledgments

Funding: None.

Footnote

Provenance and Peer Review: This article was commissioned by the editorial office, *Translational Lung Cancer Research*. The article has undergone external peer review.

Peer Review File: Available at <https://tocr.amegroups.com/article/view/10.21037/tocr-23-235/prf>

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at <https://tocr.amegroups.com/article/view/10.21037/tocr-23-235/coif>). The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Open Access Statement: This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license).

Cite this article as: Ashraf I, Hur S, Park Y. Future lung cancer prediction using low-dose chest computed tomography. *Transl Lung Cancer Res* 2023;12(7):1652-1654. doi: 10.21037/tocr-23-235

See: <https://creativecommons.org/licenses/by-nc-nd/4.0/>.

References

1. International Agency for Research on Cancer. 15-Lung-fact-sheet. 2020. Available online: <https://gco.iarc.fr/today/data/factsheets/cancers/15-Lung-fact-sheet.pdf>
2. US Preventive Services Task Force; Krist AH, Davidson KW, et al. Screening for Lung Cancer: US Preventive Services Task Force Recommendation Statement. *JAMA* 2021;325:962-70.
3. Tseng CH, Tsuang BJ, Chiang CJ, et al. The Relationship Between Air Pollution and Lung Cancer in Nonsmokers in Taiwan. *J Thorac Oncol* 2019;14:784-92.
4. Rivera GA, Wakelee H. Lung Cancer in Never Smokers. *Adv Exp Med Biol* 2016;893:43-57.
5. Mikhael PG, Wohlwend 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;41:2191-200.
6. Ardila D, Kiraly AP, Bharadwaj S, et al. End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. *Nat Med* 2019;25:954-61.
7. Huang P, Lin CT, Li Y, et al. Prediction of lung cancer risk at follow-up screening with low-dose CT: a training and validation study of a deep learning method. *Lancet Digit Health* 2019;1:e353-62.
8. American Cancer Society. Cancer Facts & Figures 2023. Special Section: Lung Cancer. Available online: <https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/annual-cancer-facts-and-figures/2023/2023-cancer-facts-and-figures.pdf> (accessed 05, April 2023).