Machine learning tool can predict malignancy in patients with multiple pulmonary nodules

24 February 2021

A machine learning-based tool was able to predict the risk of malignancy among patients presenting with multiple pulmonary nodules and outperformed human experts, previously validated mathematical models, and a previously established artificial intelligence tool, according to results published in *Clinical Cancer Research*, a journal of the American Association for Cancer Research.

Tools currently available can predict malignancy in patients with single nodules; predictive tools for patients presenting with multiple nodules are limited."With the adoption of widespread use of thoracic computed tomography (CT) for lung cancer screening, the detection of multiple pulmonary nodules has become increasingly common," said study author Kezhong Chen, MD, vice professor in the Department of Thoracic Surgery at Peking University People's Hospital in China. Among patients presenting with a pulmonary nodule on a CT scan in a previous lung cancer screening trial, roughly 50 percent presented with multiple nodules, Chen said. "Current guidelines recommend the use of clinical models that incorporate nodule and sociodemographic features to estimate the probability of cancer prior to surgical treatment, and while there are several tools for patients that present with a single nodule, no such tool currently exists for patients with multiple nodules, representing an urgent medical need," Chen added.

To address this unmet need, the researchers set out to develop a machine learning-based model to predict the probability of lung malignancy among patients presenting with multiple pulmonary nodules. First, the study authors used data from a training cohort of 520 patients (comprising 1,739 nodules) who were treated at Peking University People's Hospital between January 2007 and December 2018. Using both radiographical nodule characteristics and sociodemographic variables, the authors developed a model, termed PKU-M, to predict the probability of cancer. The performance of the model was evaluated by calculating the area under the curve (AUC), where a score of 1 corresponds to a perfect prediction. In the training cohort, the model achieved an AUC of 0.91. Some of the top predictive features of the model included nodule size, nodule count, nodule distribution, and age of the patient.

The model was then validated using data from a cohort of 220 patients (comprising 583 nodules) who underwent surgical treatment in six independent hospitals in China and Korea between January 2016 and December 2018. The performance of the PKU-M model in this cohort was similar to its performance in the training cohort, with an AUC of 0.89. The researchers also compared the performance of their model with four prior logistic regression-based models that were developed for the prediction of lung cancer. The PKU-M model outperformed all four of the prior models, whose AUC values ranged from 0.68 to 0.81.

Finally, the researchers conducted a prospective comparison between the PKU-M model, three thoracic surgeons, a radiologist, and a previously established artificial intelligence tool for the diagnosis of lung cancer called RX. This comparison was conducted on a separate cohort of 78 patients (comprising 200 nodules) who underwent surgical treatment at four independent hospitals in China between January 2019 and March 2019. Similar to the training and validation cohorts, the performance of the PKU-M model achieved an AUC of 0.87, which was higher than that from the surgeons (with AUCs ranging from 0.73 to 0.79), the radiologist (AUC of 0.75), and the

RX model (AUC of 0.76).

"The increasing detection rate of multiple pulmonary nodules has led to an emerging problem for lung cancer diagnosis," said study author Young Tae Kim, MD, Ph.D., professor in the Department of Thoracic and Cardiovascular Surgery at Seoul National University Hospital and the Seoul National University College of Medicine in the Republic of Korea. "Because many nodules are found to be benign either after long-term follow-up or surgery, it is important to carefully evaluate these nodules prior to invasive procedures. Our prediction model, which was exclusively established for patients with multiple nodules, can help not only mitigate unnecessary surgery but also facilitate the diagnosis and treatment of lung cancer."

"Models are developed to assist in clinical diagnosis, which means that they should be practical," said study author Jun Wang, MD, professor in the Department of Thoracic Surgery at Peking University People's Hospital. "We therefore designed a web-based version of the PKU-M model , where clinicians can input several clinical and radiological characteristics and the software will automatically calculate the risk of malignancy in a specific patient. This tool can quickly generate an objective diagnosis and can aid in clinical decision-making."

Because this study only used data from Asian patients, it may not be generalizable to a Western population or other populations, representing a limitation of this study.

More information: *Clinical Cancer Research* (2021). DOI: 10.1158/1078-0432.CCR-20-4007

Provided by American Association for Cancer Research