The recent development of artificial intelligence-based cancer occurrence risk prediction models

Shumin Ren

1Translation Informatics Center, Institutes for Systems Genetics, Frontiers Science Center for Disease-related Molecular Network, West China Hospital, Sichuan University, Chengdu, 610212, Sichuan, China;
2Department of Computer Science and Information Technology, University of A Coruña, 15071, A Coruña, Spain

Abstract. Artificial intelligence (AI) is playing an increasingly important role in developing cancer occurrence risk models. AI model can analyze vast amounts of data to identify patterns and correlations that may not be immediately apparent to clinicians, which can reduce overdiagnosis, timely identify risk factors, and lower incidence and mortality rates. This mini-review presented three specific articles that demonstrate the development process and application effectiveness of AI-based cancer occurrence risk models, providing inspiration and reference for future developments. These research allows for more accurate predictions of cancer risk based on a variety of factors such as imaging results, blood test result, etc. By identifying individuals at high risk for developing cancer, preventative measures can be taken to reduce their likelihood of developing the disease. Additionally, AI can help reduce overdiagnosis by distinguishing between benign and malignant conditions with greater accuracy. Overall, the use of AI in developing cancer risk models has the potential to greatly improve our ability to prevent and treat cancers.

Clinicians can use artificial intelligence (AI) algorithms to screen patients without symptoms who may be at risk for cancer, evaluate and prioritize patients with symptoms, and more accurately detect cancer occurrence (1). AI has the potential to enhance various aspects of cancer care including imaging, screening and diagnosis, treatment options and medication development (2). Compared to traditional models, AI can manage complex data sets and use higher fitting algorithms (3). Meanwhile, although there are currently some cancer risk prediction projects, such as the Tyrer-Cuzick (TC) model, which are widely embraced by the American Cancer Society and the National Comprehensive Cancer Network (NCCN) guidelines (4) (5), only providing life-time risk prediction, with restricted precision when applied to individuals and distinct periods. Therefore, it is necessary to establish a cancer risk occurrence model based personalized risk prediction of cancer occurrence using invasive markers. This mini-review introduced the latest AI-based cancer risk prediction models through summarizing three research articles and provides insights into related development directions.

Yala et al's study (6)assessed the effectiveness of Mirai, an AI-driven model for predicting breast cancer risk, in a variety of populations worldwide with the aim of enhancing early detection and minimizing unnecessary treatment. Mirai demonstrated higher sensitivity and specificity compared to traditional risk models in selecting high-risk cohorts, providing widespread and fair enhancements in treatment. The study collected 128,793 mammograms from a total of seven medical facilities located in five different nations and achieved similar or higher concordance indices compared to the original test set. The model performed better in hospitals with biennial screening, indicating the impact of screening patterns on performance. Mirai accurately predicted risk for individuals and distinct time frames, overcoming the limitations of traditional risk models that provide generate risk predictions for extensive patient populations. Mirai has the capability to supplant present risk assessment methods in protocols for magnetic resonance imaging (MRI) scans and surpassed current clinical standards based on the TC model at Massachusetts General Hospital (MGH). The study demonstrated that AI-based breast cancer models can offer significant advances over current models used in clinical practice, but prospective clinical studies must be conducted to verify the therapeutic benefit using Mirai and establish guidelines. Future work is needed to modify this technology to accommodate additional mammography providers and tomosynthesis imagery. The study had limitations, including the retrospective analysis and the need for prospective validation. However, The research constitutes the most extensive confirmation thus far of an AI-driven breast cancer model and indicates that the technology has the potential to provide widespread and fair enhancements in treatment. In conclusion, the research showed the usefulness of Mirai in precisely identifying high-risk groups and retaining its precision across varied test sets. Future clinical studies of this technology are justified to verify its clinical advantages.

Arai et al's study (7) was to create a machine learning model to forecast the occurrence of gastric cancer in individuals with chronic gastritis. The study included 1099 participants who underwent endoscopic examination and biopsy sampling of their gastric mucosa. The data was split into training and testing sets. The investigators evaluated the effectiveness of various machine learning models and discovered that the gradient-boosting decision tree (GBDT) model had the best predictive performance among all models examined. The GBDT model effectively divided the risk of gastric cancer into three groups and allowed for the generation of a personalized cumulative incidence prediction curve for each patient. The high-risk group is recommended to have annual endoscopic examination, while those in the low-risk group may not require yearly monitoring but still have some potential for cancer occurrence and should not be permanently removed from surveillance. This research emphasizes the potential of machine learning-based models for personalized risk forecasting and individualization of endoscopic surveillance intervals, leading to improved follow-up esophagogastroduodenoscopy (EGD) intervals. However, the study has several limitations, such as its retrospective single-center design and the absence of well-known risk factors for gastric cancer(8) (9) (10). The researchers recommend further validation of the model using larger datasets with more variables and external institutional data to confirm its general feasibility and improve its accuracy. Overall, the GBDT model developed in this study showed high predictive performance for gastric cancer incidence and may have clinical applications for risk stratification and prevention.

Soerensen et al.'s study(11) was to assess how well an artificial intelligence system can predict the likelihood of cancer in individuals who have been sent from their primary care provider for routine blood tests. The AI model's performance was evaluated against that of logistic regression. The study utilized a set of 25 standard laboratory blood tests. The primary outcome was whether or not a patient was diagnosed with cancer within a 90-day period. The findings indicated that the system using standard laboratory blood tests can generate a convenient risk score to forecast the likelihood of a cancer diagnosis within a 90-day period. The AI system's performance was found to be similar to that of the logistic regression model. The study's strengths include well-defined groups of participants and the confirmation of the model's accuracy in a separate validation test using a different group of participants from a different time period. This helps to assess how well the model performs over time. The fact that the model uses standard laboratory tests that are readily available to primary care providers enhances its practicality in a clinical setting, and the study's limitations include a relatively small study population, retrospective design, and lack of assessment of single blood tests' significance in detecting cancer. Prior to general clinical implementation, further testing is needed to confirm the usefulness of the risk score in different groups of people. In conclusion, the risk score generated by the AI system using standard laboratory blood tests could be a useful tool in helping doctors decide whether a patient needs further testing or if a wait-and-see approach is more appropriate. Future enhancements to the model could include tailoring it to specific genders and incorporating demographic information along with laboratory test results.

To sum up, AI has shown great potential in improving cancer screening and risk assessment through the development of personalized cancer risk prediction models. These models utilize large datasets, including patients' imaging results, blood samples and other test results, to accurately predict an individual's risk of developing cancer. By incorporating AI into cancer screening, early detection rates can be improved, reducing morbidity and mortality associated with cancer. The personalized cancer risk prediction models listed in this article have demonstrated significant improvements in identifying high-risk cohorts and maintaining accuracy across diverse test populations. These models also enable the generation of individualized cumulative incidence prediction curves, contributing to better follow-up intervals and potentially reducing overtreatment. However, prospective clinical studies must be conducted to verify the clinical advantage of these models and establish guidelines for their use. As AI-based cancer risk prediction models become more widely adopted, it is essential to ensure that they are ethically designed and implemented, prioritizing patient privacy and equity. In conclusion, AI-based cancer risk prediction models offer a promising future for cancer screening and risk assessment. These models have the potential to improve early detection, reduce overtreatment, and enhance personalized patient care. It is crucial to continue to advance and validate these models to ensure their safe and effective integration into clinical practice.

REFERENCE

1. Hunter B, Hindocha S, Lee RWJC. The role of artificial intelligence in early cancer diagnosis. 2022;14(6):1524.
2. Liang G, Fan W, Luo H, Zhu XJB, Pharmacotherapy. The emerging roles of artificial intelligence in cancer drug development and precision therapy. 2020;128:110255.

3. Fei Y, Li W-qJJocc. Improve artificial neural network for medical analysis, diagnosis and prediction. 2017;40:293.
4. Smith RA, Andrews KS, Brooks D, Fedewa SA, Manassaram-Baptiste D, Saslow D, et al. Cancer screening in the United States, 2019: A review of current American Cancer Society guidelines and current issues in cancer screening. CA: a cancer journal for clinicians. 2019;69(3):184-210.
5. Bevers TB, Ward JH, Arun BK, Colditz GA, Cowan KH, Daly MB, et al. Breast cancer risk reduction, version 2.2015. 2015;13(7):880-915.
6. Yala A, Mikhael PG, Strand F, Lin G, Satuluru S, Kim T, et al. Multi-Institutional Validation of a Mammography-Based Breast Cancer Risk Model. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2022;40(16):1732-40.
7. Arai J, Aoki T, Sato M, Niikura R, Suzuki N, Ishibashi R, et al. Machine learning–based personalized prediction of gastric cancer incidence using the endoscopic and histologic findings at the initial endoscopy. 2022;95(5):864-72.
8. Eom BW, Joo J, Kim S, Shin A, Yang HR, Park J, et al. Prediction Model for Gastric Cancer Incidence in Korean Population. PloS one. 2015;10(7):e0132613.
9. Cai Q, Zhu C, Yuan Y, Feng Q, Feng Y, Hao Y, et al. Development and validation of a prediction rule for estimating gastric cancer risk in the Chinese high-risk population: a nationwide multicentre study. Gut. 2019;68(9):1576-87.
10. Leung WK, Cheung KS, Li B, Law SYK, Lui TKL. Applications of machine learning models in the prediction of gastric cancer risk in patients after Helicobacter pylori eradication. Alimentary pharmacology & therapeutics. 2021;53(8):864-72.
11. Soerensen PD, Christensen H, Gray Worsoe Laursen S, Hardahl C, Brandslund I, Madsen JSJCC, et al. Using artificial intelligence in a primary care setting to identify patients at risk for cancer: a risk prediction model based on routine laboratory tests. 2022;60(12):2005-16.