Machine learning

https://doi.org/10.1038/s41591-023-02630-y

AI and imaging-based cancer screening: getting ready for prime time

Jörg Kleeff & Ulrich Ronellenfitsch

New data show that AI could enhance imaging-based screen for pancreatic cancer; however, its evaluation must be rigorous and adhere to the same standards used for conventional screening.

In recent years, artificial intelligence (AI) has become a pervasive element of our lives. Consciously or unconsciously, we interact with AI techniques when using search engines on the internet, posting on or reading social media, or using transportation. In clinical medicine, the uptake of AI has happened at a much slower pace, with diagnoses and treatment recommendations still almost exclusively based on human judgement. Only recently have AI techniques been evaluated for their applicability and potential benefit for several clinical scenarios, with video and imaging applications leading the way[1]. In this issue of *Nature Medicine*, Cao et al.[2] report the results of a study in which they have assessed AI techniques for detecting and classifying pancreatic lesions in non-contrast computerized tomography (CT) imaging. The approach tries to meet the clinical need for the early detection of pancreatic cancer, a challenging disease given its often unspecific symptoms, which result in late detection and poor prognosis[3].

Cao et al.[2] trained their AI algorithm on more than 3,200 image sets from a high-volume pancreatic cancer institution, of which about 70% stemmed from patients with a pancreatic lesion. The AI algorithm was validated in a multi-center cohort of about 6,200 patients comprising cases with confirmed pancreatic cancer and controls without pancreatic lesions, a multi-reader study in selected cases from the training cohort, and a second validation cohort comprising 20,530 patients from the same institution. The algorithm achieved a sensitivity of 92.9% and a specificity of 99.9% for the identification of pancreatic lesions, with an area under the curve (AUC) of 0.986–0.996. Human radiologists were outperformed by the algorithm for the detection of pancreatic adenocarcinoma – the most common and deleterious pancreatic tumor.

The results of Cao et al.[2] align well with recent evidence supporting the use of AI in a range of clinical settings. A scoping review of randomized controlled trials using AI for various purposes in patient management showed that 69 out of 84 trials (82%) reported positive results for their primary endpoint[1]. Although there may be publication bias, this finding is still compelling. A prominent example comes from a randomized controlled trial involving more than 80,000 women undergoing mammography screening for breast cancer, in which AI-supported mammography reading outperformed radiologist reading by a ratio of 1.2, enhancing rates of breast cancer detection from 5.1 to 6.1 per 1,000 participants while reducing reading workload by more than 40%[4]. AI-based imaging has even been evaluated as a screening tool for diseases not traditionally diagnosed by imaging alone.

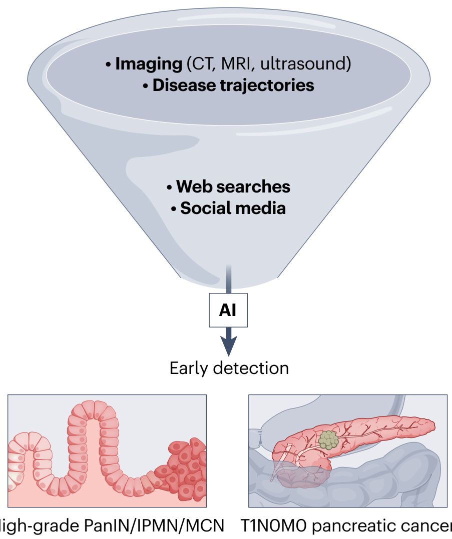

Fig. 1 | AI-assisted prediction and early detection of pancreatic cancer. In the future, information from imaging, individual disease trajectories, web searches and social media activities can possibly be combined and analyzed with AI techniques for the prediction and early detection of pancreatic cancer. PanIN, pancreatic intraepithelial neoplasms; IMPN, intraductal papillary mucinous neoplasms; MCN, mucinous cystic neoplasms. T1N0M0 indicates the diameter of the tumor cancer is 2 cm (or less) without evidence of lymph node or distant metastasis.

A deep-learning model applied to chest x-rays detected left ventricular ejection fraction below 40%, with a sensitivity and specificity of 82% and 86% (AUC 0.92), respectively – which is remarkable for a method traditionally believed not to provide sufficient information on cardiac functioning[5]. Even the analysis of retinal images for the detection of extraocular diseases such as stroke or Parkinson's disease – an approach which, to some, might sound like science-fiction – has shown promising results[6].

The accuracy metrics of the algorithm used by Cao et al.[2] are superior to those of several acknowledged screening methods such as Pap smears for cervical cancer or mammography for breast cancer[7,8]. This makes it tempting to call for integration of this specific method into large-scale screening efforts – by either offering non-contrast abdominal CT to specific population groups or applying the tool to non-contrast CT performed for other reasons (not related to cancer screening). However, the findings[2] also illustrate some of the challenges faced by AI-based screening studies, and these need to be considered before conclusive recommendations can be made.

The idea of applying AI-based screening to images acquired for other purposes, such as during visits to emergency departments or routine checkups, is appealing owing to the potential for synergy and cost reduction. However, the success of a screening method hinges on the detection of the disease in the early stages and the related amenability to curative treatment. Accuracy must always be specifically considered in this context, for each cancer type and population group under study. This is highly relevant for pancreatic cancer, owing to the poor prognosis in advanced stages of the disease. Any potential screening method for pancreatic cancer should thus detect early stages such as T1 lesions (<2 cm diameter). Cao et al.[2] report a sensitivity value in this subgroup of 85.7% in the test cohort and 92.2% in the validation cohort. Specificity and predictive values were not reported for this subgroup.

These metrics are particularly important if an examination such as a CT scan is primarily performed for other purposes and the screening is a by-product in a population with a low prevalence of the disease. False-positive findings startle patients and entail costly and often invasive further diagnostic measures. Accurate informed consent of patients is therefore paramount, and the possible consequences of a positive, but also a negative, finding on AI-based screening must be well explained. In addition, some patients might be reluctant to have their data assessed by an AI tool rather than a human, even in the presence of compelling data, owing to a general skepticism about AI[9].

Lastly, the value of any screening method for cancer lies in reducing all-cause mortality[10]. The earlier detection of tumors alone does not guarantee this. The study by Cao et al.[2] was of retrospective design and so could not assess the effect of screening on the mortality of included patients. Any definitive conclusion on the utility of the method would thus be premature, as AI-based screening should be evaluated with the same rigor as conventional screening — using randomized controlled trials to compare the approach with a valid comparator with regard to all-cause mortality.

Notwithstanding these challenges, it is possible to envision a future in which AI is used to combine information from routine imaging with information on clinical history[11], and even with individual online search trends and use of social media (as has been shown for disease outbreaks)[12] — for the prediction and early detection of challenging-to-treat diseases such as pancreatic cancer (Fig. 1). AI-based screening is a highly promising approach with potential for clinical impact in the near future; however, further thorough assessment is needed before it is ready to be introduced into widespread practice.

Jörg Kleeff ✉ & Ulrich Ronellenfitsch
Department of Visceral, Vascular and Endocrine Surgery, Martin-Luther-University Halle-Wittenberg, University Medical Center Halle (Saale), Halle (Saale), Germany.
✉e-mail: joerg.kleeff@uk-halle.de

Published online: 20 November 2023

References

1. Han, R. et al. Preprint at *medRxiv* https://doi.org/10.1101/2023.09.12.23295381 (2023).
2. Cao, K. et al. *Nat. Med.* https://doi.org/10.1038/s41591-023-02640-w (2023).
3. Kleeff, J. et al. *Nat. Rev. Dis. Primers* **2**, 16022 (2016).
4. Lång, K. et al. *Lancet Oncol.* **24**, 936–944 (2023).
5. Ueda, D. et al. *Lancet Digital Health* **5**, e525–e533 (2023).
6. Zhou, Y. et al. *Nature* **622**, 156–163 (2023).
7. Tadesse, G. F., Tegaw, E. M. & Abdisa, E. K. *J. Ultrasound* **26**, 355–367 (2023).
8. Hon, H. J., Chong, P. P., Choo, H. L. & Khine, P. P. *Asian Pac. J. Cancer Prev.* **24**, 2207–2215 (2023).
9. Zhou, Y., Shi, Y., Lu, W. & Wan, F. *Front Psychol.* **13**, 866124 (2022).
10. Bretthauer, M. et al. *JAMA Intern. Med.* https://doi.org/10.1001/jamainternmed.2023.3798 (2023).
11. Placido, D. et al. *Nat. Med.* **29**, 1113–1122 (2023).
12. Kurian, S. J. et al. *Mayo Clinic Proc.* **95**, 2370–2381 (2020).

Competing interests

The authors declare that they have no known competing financial or non-financial interests.