BJD
British Journal of Dermatology

Can clinical decision making be enhanced by artificial intelligence?

DOI: 10.1111/bjd.17110

Linked Article: Fujisawa *et al.* Br J Dermatol 2019; 373–381.

In this issue of the BJD, Fujisawa and colleagues[1] report on the development of a 'classifiers' algorithm first to distinguish between benign or malignant skin tumours, and then to determine the diagnostic category of the malignancies. The authors trained their deep convolutional neural network on a relatively small number of images (< 5000), in contrast to previous work that has used > 120 000 images for training.[2] Furthermore, many of the images in this study were obtained from rather uncommon skin tumours, in areas that are often difficult to photograph. For example, 52% of the melanomas and 38% of the benign lesions were located at acral sites, likely reflecting the ethnicity of the patients. The study also included 121 poromas, which are benign adnexal neoplasms with eccrine differentiation commonly located on acral sites. In contrast, the study did not include photographs of other common lesions such as angiomas and dermatofibromas, the latter often posing a differential diagnostic problem in clinical practice. Given the case mix, we can also assume that few images of atypical naevi on the trunk or extremities were included, although this was not mentioned in detail.

Despite these limitations, the analysis shows that deep neural networks can perform their intended purpose on the first 81% of this dataset, and give reasonable accuracy when tested against the remaining 19%. The authors report 96·3% sensitivity and 89·5% specificity to group the images correctly into malignant or benign diagnoses, and this result compares favourably against 13 board-certified dermatologists and nine dermatology trainees. However, given that any bias that exists in a particular health system would be perpetuated in future medical decisions, it is important to study the performance of the algorithm further in a fresh dataset that includes a different mix of lesion types.

The present analyses follow in the footsteps of other recent reports that used a deep convolutional neural network to derive algorithms and compared their performance against that of dermatologists. So far, studies based on image analyses seem to provide a good estimation of whether or not a photographed lesion is likely to be a skin cancer, with some even classifying the images into finer categories.[3] Recent examples include the work by Haennsle *et al.*,[4] where the algorithm was compared with the performance of 58 dermatologists, and results of a prominent paper by Esteva *et al.*,[2] where the algorithm outperformed 21 Stanford dermatologists.

So where to from here? Fujisawa and colleagues[1] correctly state that while the results are exciting, prospective evaluations

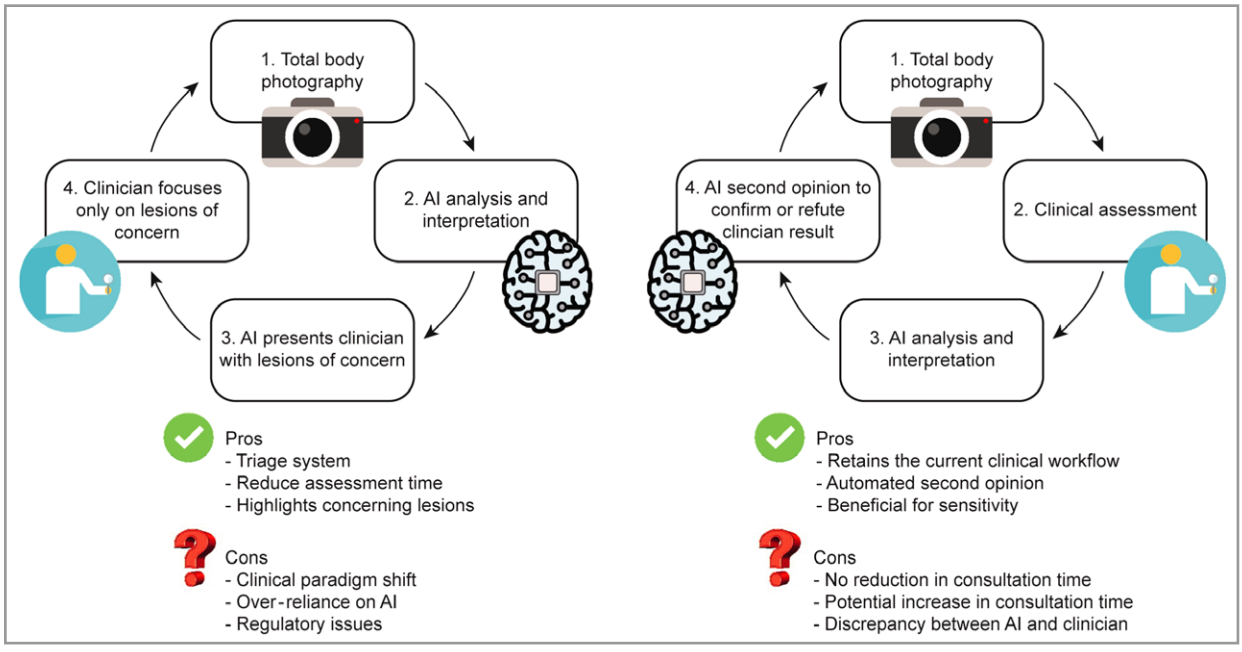

Fig 1. Placing artificial intelligence before vs. after the clinician: each has several pros and cons.

must still be done. As highlighted by Beam and Kohane,[5] data-derived algorithms have long supported clinicians in decision making, and they commonly represent just one small snippet of the many facets of a patient's presentation that clinicians must consider. What, then, would be an appropriate study design for a prospective evaluation in the clinic? Individual randomization may be prone to contamination, but it has been used successfully in the MAVARIC trial[6] in cervical cytology. MAVARIC compared manual reading only vs. paired automation-assisted plus manual reading, finding limited benefit from the addition of automation. A cluster-randomized trial whereby some clinics continue with usual care and others are assigned to usual care plus artificial intelligence (AI) may therefore be a design that could avoid contamination. However, it could have difficulty reaching its recruitment target if patients start to visit clinics that specifically offer AI reading. In both scenarios, the addition of AI would need to show improved sensitivity and specificity, as well as cost-effectiveness, before becoming standard care.[7]

In any assessment of benefit, the ideal positioning of AI in relation to the clinician also needs to be considered (Fig. 1).[8] Placing AI before the clinician would be similar to the model employed in cervical cytology screening, where automation allows the cytologist to focus on likely relevant locations above a threshold for suspicious cells.[9] Using AI in this way as a triage tool could greatly reduce the time a dermatologist needs to spend in the assessment of each patient, and therefore allow them to see more patients, or perform more telediagnoses.

Similarly to the current clinical approach, placing the clinician first, AI could provide an automated second opinion, either confirming or refuting the clinician's primary judgement. This would not save time, but it would help increase the sensitivity, and would be particularly useful for patients with many potentially suspicious lesions. Indeed, decision support systems are already being embedded in the software of many of the advanced dermatology imaging products now on the market.

If AI is proven to be effective in prospective studies, clinicians will likely welcome the support system, as it should decrease their risk of overlooking a melanoma in practice. However, the sensitivity will still not be 100% as there are unusual presentations that even AI would not have recognized.[10] However, overall, use of AI should free up clinicians' time to serve more patients, to explain treatment plans and options better, and to counsel patients distraught by a diagnosis of malignancy.

Conflicts of interest

H.P.S. is a shareholder and consultant for e-derm consult GmbH and MoleMap by Dermatologists Pty Ltd. He provides consultancy services for Canfield Scientific Inc.

M. Janda[1]
H.P. Soyer[2,3]

[1]*Centre for Health Services Research, Faculty of Medicine, The University of Queensland, Princess Alexandra Hospital, Brisbane, QLD, Australia*
[2]*Dermatology Research Centre, The University of Queensland, The University of Queensland Diamantina Institute, Brisbane, QLD, Australia* 4102
[3]*Dermatology Department, Princess Alexandra Hospital, Brisbane, QLD, Australia*
Correspondence: H. Peter Soyer.
E-mail: p.soyer@uq.edu.au

References

1 Fujisawa Y, Otomo Y, Ogata Y *et al.* Deep-learning-based, computer-aided classifier developed with a small dataset of clinical images surpasses board-certified dermatologists in skin tumour diagnosis. *Br J Dermatol* 2019; **180**: 373–381.
2 Esteva A, Kuprel B, Novoa RA *et al.* Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 2017; **542**:115–8.
3 Mar VJ, Soyer HP. Artificial intelligence for melanoma diagnosis: how can we deliver on the promise? *Ann Oncol* 2018; **29**:1625–8.
4 Haenssle HA, Fink C, Schneiderbauer R *et al.* Man against machine: diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists. *Ann Oncol* 2018; **29**:1836–42.
5 Beam AL, Kohane IS. Big data and machine learning in health care. *JAMA* 2018; **319**:1317–8.
6 Kitchener HC, Blanks R, Dunn G *et al.* Automation-assisted versus manual reading of cervical cytology (MAVARIC): a randomised controlled trial. *Lancet Oncol* 2011; **12**:56–64.
7 Janda M, Soyer HP. Automated diagnosis of melanoma. *Med J Aust* 2017; **207**:361–2.
8 Raphael AP, Soyer HP. Automated diagnosis: shedding the light on skin cancer. *Br J Dermatol* 2018; **178**:331–3.
9 Smith JHF. Cervical screening: history, current algorithms, and future directions. In: *Pathology of the Cervix. Essentials of Diagnostic Gynecological Pathology* (Herrington C, ed.), vol. 3. Cham: Springer, 2017; 45–65.
10 Mar VJ, Chamberlain AJ, Kelly JW *et al.* Clinical practice guidelines for the diagnosis and management of melanoma: melanomas that lack classical clinical features. *Med J Aust* 2017; **207**:348–50.