Y-LW reports personal fees from AstraZeneca, Eli Lilly, Roche, Pfizer, Bristol-Myers Squibb, Merck, and Boehringer Ingelheim; consulting or advisory role with AstraZeneca, Roche, Merck, and Boehringer Ingelheim; and research funding to Guangdong Lung Cancer Institute from Boehringer Ingelheim, AstraZeneca, and Roche. YL declares no competing interests.

1 Liu SY, Wu YL. An immunological storm for cancer therapy: 2018 Nobel Prize in Physiology or Medicine. *Sci Bull* 2018; **63:** 1608–70.
2 Wu YL, Planchard D, Lu S, et al. Pan-Asian adapted Clinical Practice Guidelines for the management of patients with metastatic non-small-cell lung cancer: a CSCO–ESMO initiative endorsed by JSMO, KSMO, MOS, SSO and TOS. *Ann Oncol* 2019; **30:** 171–210.
3 Reck M, Rodríguez-Abreu D, Robinson AG, et al. Pembrolizumab versus chemotherapy for PD-L1-positive non-small-cell lung cancer. *N Engl J Med* 2016; **375:** 1823–33.
4 Gandhi L, Rodríguez-Abreu D, Gadgeel S, et al. Pembrolizumab plus chemotherapy in metastatic non-small-cell lung cancer. *N Engl J Med* 2018; **378:** 2078–92.
5 West H, McCleod M, Hussein M, et al. Atezolizumab in combination with carboplatin plus nab-paclitaxel chemotherapy compared with chemotherapy alone as first-line treatment for metastatic non-squamous non-small-cell lung cancer (IMpower130): a multicentre, randomised, open-label, phase 3 trial. *Lancet Oncol* 2019; published online May 20. http://dx.doi.org/10.1016/S1470-2045(19)30167-6.
6 Arbour KC, Mezquita L, Long N, et al. Impact of baseline steroids on efficacy of programmed cell death 1 and programmed death ligand 1 blockade in patients with non-small-cell lung cancer. *J Clin Oncol* 2018; **36:** 2872–78.
7 Liu D, Wang S, Bindeman W. Clinical applications of PD-L1 bioassays for cancer immunotherapy. *J Hematol Oncol* 2017; **10:** 110.
8 Doroshow DB, Sanmamed MF, Hastings K, et al. Immunotherapy in non-small cell lung cancer: facts and hopes. *Clin Cancer Res* 2019; published online Mar 1. DOI:10.1158/1078-0432.CCR-18-1538.
9 Wang Z, Duan J, Cai S, et al. Assessment of blood tumor mutational burden as a potential biomarker for immunotherapy in patients with non-small cell lung cancer with use of a next-generation sequencing cancer gene panel. *JAMA Oncol* 2019; published online Feb 28. DOI:10.1001/jamaoncol.2018.7098.
10 Wang Z, Huang C, Yang JJ, et al. A randomised phase II clinical trial of nab-paclitaxel and carboplatin compared with gemcitabine and carboplatin as first-line therapy in advanced squamous cell lung carcinoma (C-TONG1002). *Eur J Cancer* 2019; **109:** 183–91.

Machine versus man in skin cancer diagnosis

Skin examinations for diagnosis of skin cancer are facilitated by the use of dermatoscopy, a non-invasive skin imaging technique that amplifies features of pigmented skin lesions that are not easily discernable when examined by the naked eye. Since its introduction in the late 1980s, dermatoscopy has substantially increased diagnostic accuracy in the identification of melanoma. However, interpretation of dermatoscopic criteria remains a subjective procedure that requires specific and dedicated training. Several strategies based on artificial intelligence and machine learning, including deep learning networks, have been developed as aids for less experienced dermatologists or other physicians to classify dermatoscopic images into benign and malignant skin tumours, as well as more specific diagnostic categories,[1–4] which in the long-term could revolutionise dermatology practice.

In *The Lancet Oncology*, Philipp Tschandl and colleagues[5] compared the accuracy of diagnoses produced by human readers with that of machine-learning algorithms for dermatoscopic images of pigmented skin lesions. Most outcome measures were outperformed by the machine-learning algorithms with respect to human readers. The best machine-learning algorithms achieved a mean of 7·94 (95% CI 7·76–8·12; p<0·0001) more correct diagnoses than the average human reader, and a mean of 6·65 (6·06–7·25; p<0·0001) more correct diagnoses than expert human readers with more than 10 years of experience.[5] This robustly powered study provides a timely contribution to an increasingly exciting debate in the field of dermatology and focuses on the more general issue of artificial intelligence use in diagnostic medicine. Tschandl and colleagues have aggregated a hugely balanced amount of information across patient populations from different geographical areas and human readers with varying experience from a large variety of geographically, ethnically, and culturally diverse countries. Sets of images representative of various types of disease (benign and malignant) were analysed and coupled with a robust statistical model, resulting in sound acceptability. The comparison of a large series of different machine-learning algorithms, which were developed by 77 laboratories that participated in the International Skin Imaging Collaboration Forum 2018 challenge,[6] is also novel.

Despite a marked effort to control for the intrinsic drawbacks of overfitting in machine learning, the authors underline that training too hard causes the predictive models to be too specific to the training set used.[5] Overfitting emerges as a problem with models that fit to the training data too closely, thus limiting the generalisability of the model and leading to poor success for data that might seem similar, but are actually different. The training set in an overfitted model might generate a very high performance, which


skvalval/Shutterstock.com

Published Online June 11, 2019 http://dx.doi.org/10.1016/S1470-2045(19)30391-2

See **Articles** page 938

For more on the **ISIC 2018 challenge** see https://challenge2018.isic-archive.com is unfortunately associated with poor to mediocre performance on unseen data. Tschandl and colleagues[5] also tested the machine-learning algorithms on images from sources different from those of the training set to properly quantify the overfitting issue.

Creating a gold standard of medical diagnosis is not a straightforward process. In a previous study by Tschandl and colleagues,[7] more than 50% of lesions, including all malignant tumours, were analysed histopathologically. Identification of benign unexcised lesions included more than 1·5 years of follow-up of sequential dermatoscopic imaging without changes, and judgment on the basis of expert consensus in cases of benign, banal, non-melanocytic lesions. For conclusive classification, seven simplified classes of diseases were generated, with specific attention paid to avoiding ambiguous classifications, and cases with uncertain histopathological diagnoses were excluded.

Where and when to appropriately use artificial intelligence is a matter of initial, but already intense, debate. Although appreciated as a powerful tool, artificial intelligence does not seem ready to replace the refined cognitive process of integrating morphological observations within a clinical context. In a real-world situation, in addition to dermatoscopic images clinicians consider several other features, including anatomical site, age, sex, and clinical history and evolution. Machine learning and humans should also be compared in terms of turnaround time and cost-effectiveness before becoming a standard of care.

In certain high-need settings, such as low-income and middle-income countries or health-care systems with few human resources, artificial intelligence might bring additional value, particularly as a screening tool. Furthermore, artificial intelligence could also be useful for the development of clinical support tools for inexperienced physicians and general practitioners, in view of an increasing demand for points of care for dermatology screening.[8] However, in future, automated classifiers and artificial intelligence algorithms could be integrated into clinical dermatology practice for a more accurate and effective triage of lesions in the context of human–machine collaboration. Although there is little doubt that artificial intelligence and machine-learning algorithms for skin cancer diagnosis are already gaining a central role in dermatology research, it is possible to anticipate that their application to clinical practice will require serious and robust validation in large, prospective studies.

*Daniela Massi, Marco Laurino

Section of Pathological Anatomy, Department of Health Sciences, University of Florence, Florence 50139, Italy (DM); and Institute of Clinical Physiology, National Research Council, Pisa, Italy (DM, ML)

daniela.massi@unifi.it

We declare no competing interests.

1 Codella N, Nguyen Q-B, Pankanti S, et al. Deep learning ensembles for melanoma recognition in dermoscopy images. *IBM J Res Dev* 2017; **61:** 1–15.
2 Esteva A, Kuprel B, Novoa RA, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 2017; **542:** 115–18.
3 Han SS, Kim MS, Lim W, Park GH, Park I, Chang SE. Classification of the clinical images for benign and malignant cutaneous tumors using a deep learning algorithm. *J Invest Dermatol* 2018; **138:** 1529–38.
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 Tschandl P, Codella N, Akay BN, et al. Comparison of the accuracy of human readers versus machine learning algorithms for skin lesion classification: an open, web-based, international, diagnostic study. *Lancet Oncol* 2019; published online June 11. http://dx.doi.org/10.1016/S1470-2045(19)30333-X.
6 Stoyanov D, Taylor Z, Sarikaya D, et al. OR 2.0 context-aware operating theaters, computer assisted robotic endoscopy, clinical image-based procedures, and skin image analysis. First International Workshop, OR 2.0 2018, 5th International Workshop, CARE 2018, 7th International Workshop, CLIP 2018, Third International Workshop, ISIC 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, Sept 16 and 20, 2018, Proceedings. New York: Springer, 2018.
7 Tschandl P, Rosendahl C, Kittler H. The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. *Sci Data* 2018; **5:** 180161.
8 Rogers HW, Weinstock MA, Feldman SR, Coldiron BM. Incidence estimate of nonmelanoma skin cancer (keratinocyte carcinomas) in the US population, 2012. *JAMA Dermatol* 2015; **151:** 1081–86.

Neoadjuvant therapy for melanoma: is it ready for prime time?

Published Online June 3, 2019 http://dx.doi.org/10.1016/S1470-2045(19)30377-8

See Articles pages 948 and 961

In the past, advanced melanoma with nodal or distant metastases was almost universally fatal, with a median overall survival of less than 1 year with few effective therapies.[1] However, with the advent of treatment with immune checkpoint blockade and *BRAF*-targeted therapy, among other strategies, a dramatic improvement has been seen in the survival of patients with stage IV disease, substantiating approval of such