Digital medicine

Can skin cancer diagnosis be transformed by AI?

For more on **Digital medicine** see **Comment** Lancet 2016; **388:** 740 and **Perspectives** Lancet 2019; **394:** 907

Skin conditions, especially different types of cancer, are common. Yet the number of dermatologists is fairly low. Dermatology is a specialty suited for artificial intelligence (AI) research and potential incorporation in clinical practice. AI has the potential to decrease dermatologist workloads, eliminate repetitive and routine tasks, and improve access to dermatological care.

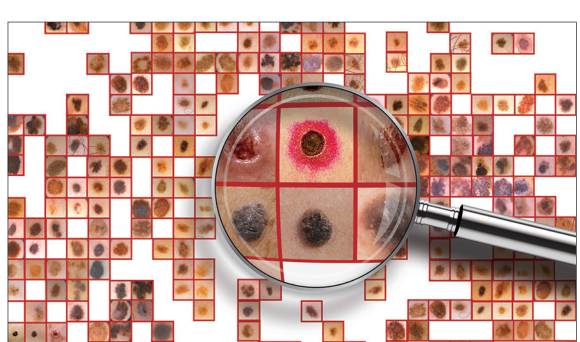
Today, it's straightforward to train and deploy deep learning algorithms for the classification, detection, and tracking of objects in images. In addition to visual inspection with the naked eye, dermatologists routinely use dermatoscopes for the analysis of cases. Using photographic and dermoscopic images, studies have shown physician-level diagnostic performance at single-lesion analysis for algorithms trained on these images. Image datasets comprise tens to hundreds of thousands of lesion images combined with millions of images of everyday objects. Typically, algorithms are trained using supervised learning, in which datasets are gathered that are composed of images and image labels, specifying the disease in the image. Datasets are then repeatedly shown to the algorithm until it learns to recognise disease states. These data are often labelled by many physicians, resulting in algorithms that learn from the collective intelligence of many experts.

Trained computer vision algorithms can be used on any medical device that collects image or video data. For instance, classification algorithms can be run on smartphones with or without dermatoscope attachments. These algorithms, however, are not yet ready for clinical use. There have only been a few retrospective studies that have compared the performance of algorithms with physicians at the task of image classification, and these were predominantly for melanoma cases. For deep learning, this can pose an issue since retrospective data used for training algorithms can differ from data processed prospectively in the real-world clinical environment. Additionally, deep learning algorithms are highly susceptible to overfitting, and their performance can be affected by miscorrelations in datasets. For instance, if malignant lesion training images contain the markers or rulers used by dermatologists to measure lesion diameters but benign lesion training images never do, then algorithms may learn that the presence of a marker implies the presence of cancer, leading to a drop in true performance. Similarly, if the images used to test an algorithm contain artifacts it has never encountered, then the algorithm is likely to misinterpret them as relevant. To clarify these concerns, algorithms must be validated prospectively by testing a functioning system (eg, a mobile app) against the clinical performance of physicians. Even slight differences in systems, such as the use of different mobile phones and dermatoscope attachments, must each be validated. Algorithms also need to be trained and validated for use across demographics. Studies have focused primarily on data from white people, and algorithms trained on such data perform inadequately in other populations. Regulatory guidelines need to include requirements that algorithms be validated on data from any demographic population for which they are certified.

To date, studies have largely focused on algorithms that can identify melanomas and dermal carcinomas from their benign counterparts. Preliminary research is beginning to expand these algorithms to other skin conditions. Here, the challenge lies not only in identifying the type of condition (eg, psoriasis *vs* eczema) but also in usefully quantifying change over time so that algorithms can help patients and providers understand the effectiveness of treatments. To serve as fully functioning clinical aids, deep learning algorithms must be rigorously validated for all diseases within a particular class. That is, if an algorithm is certified for the classification of melanoma, it must be able to exclude anything that could be mistaken for one.

If these challenges are overcome and deep learning systems are proven to be effective at general clinical use, the workflow of physicians could change for the better. Primary care doctors would be able to use a deep learning-enhanced smartphone app to screen patients for skin cancer during routine visits. Similarly, dermatologists could use deep learning for second opinions or for the rapid triage of lesions in difficult cases, such as inspecting patients with hundreds of moles. Eventually, patients might even benefit if at-home self-analysis is developed that enables them to avoid unnecessary clinic visits and pick up important signs early that might have been missed.

*Andre Esteva, Eric Topol
Stanford University, Stanford, CA, USA (AE); Medical AI, Salesforce Research, Palo Alto, CA 94301, USA (AE); and Scripps Research Translational Institute, La Jolla, CA, USA (ET)
andre.esteva@gmail.com



Faith Hark/Scripps Research Translational Institute

Further reading

Esteva A, Kuprel B, Novoa RA, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 2017; **542:** 115

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

Brinker TJ, Hekler A, Enk AH, et al. Deep learning outperformed 136 of 157 dermatologists in a head-to-head dermoscopic melanoma image classification task. *Euro J Cancer* 2019; **113:** 47–54

LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature* 2015; **521:** 436

Russakovsky O, Deng J, Su H, et al. Imagenet large scale visual recognition challenge. *Int J Comput Vis* 2015; **115:** 211–52

Winkler JK, Fink C, Toberer F, et al. Association between surgical skin markings in dermoscopic images and diagnostic performance of a deep learning convolutional neural network for melanoma recognition. *JAMA Dermatol* 2019; published online Aug 14. DOI:10.1001/jamadermatol.2019.1735

Adamson AS, Smith A. Machine learning and health care disparities in dermatology. *JAMA Dermatol* 2018; **154:** 1247–48