Artificial intelligence and melanoma detection: friend or foe of dermatologists?

The significance of early diagnosis for melanoma prognosis and survival cannot be understated. The public health benefits of melanoma prevention and detection have driven advances in diagnostics for skin cancer, particularly in the field of artificial intelligence. Evaluating the benefits and limitations of artificial intelligence in dermatology is paramount to its future development and clinical application.

Maria Charalambides[1]

Sonal Singh[2]

Author details can be found at the end of this article

Correspondence to:
Maria Charalambides;
mariacharalambides@hotmail.co.uk

What is artificial intelligence?

Artificial intelligence aims to mimic human cognitive functions by using machine learning algorithms. Following the input of training data and various data analytic methods it identifies patterns in data to make associations and, in the context of medicine, reach a diagnosis (Jiang et al, 2017) (Figure 1). Artificial intelligence could lead to a paradigm shift of diagnostics in dermatology; it is powered by the growing availability of data and rapid progress of analytic methods (Mar and Soyer, 2018).

Guidelines in dermatology

According to the National Institute for Health and Care Excellence (2015), current practice for skin cancer diagnosis involves a thorough history and examination, dermoscopic analysis and histopathology. Artificial intelligence can be integrated into the diagnostic pathway to supplement clinical examination, possibly through digital dermoscopic analysis.

Artificial intelligence in dermatology: the evidence

Research into the role of artificial intelligence in dermatology has been conducted largely in the context of melanoma detection. Studies suggest that convolutional neural networks may achieve comparable rates of diagnostic accuracy to dermatologists (Table 1). Thus, the integration of artificial intelligence into care has the potential to augment clinical decision making, especially in primary care, and offer a more standardised and accessible level of diagnostic accuracy.

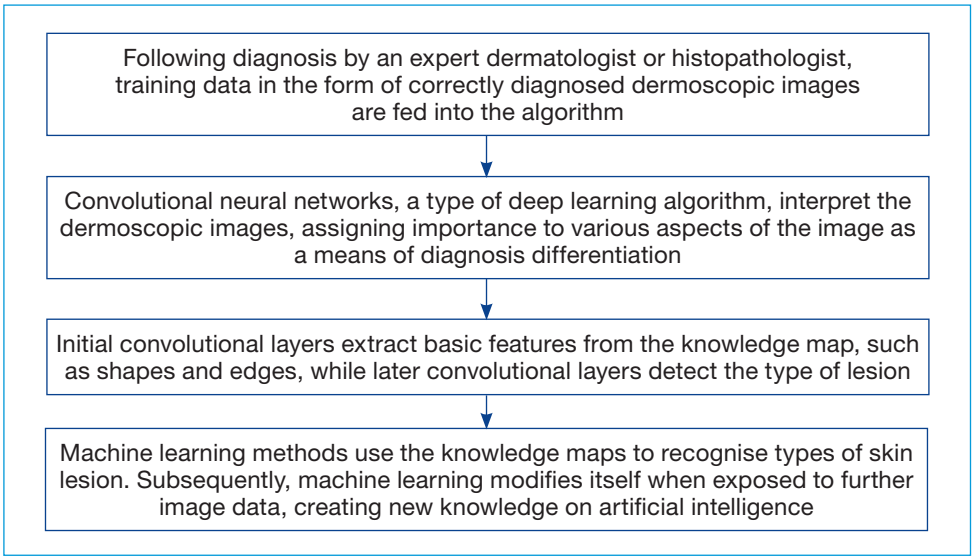

Figure 1. Flowchart demonstrating the process of dermoscopic image interpretation using artificial intelligence.

How to cite this article:
Charalambides M, Singh S. Artificial intelligence and melanoma detection: friend or foe of dermatologists? Br J Hosp Med. 2020;81(1):1–5. https://doi.org/10.12968/hmed.2019.0322

Table 1. Studies comparing CNN with dermatologists for melanoma detection

Reference	Dataset used	Comparison	Key findings
Esteva et al (2017)	129 450 clinical images used to train the CNN	21 dermatologists	ROC area under the curve graphs showed that CNN matched the performance of 21 dermatologists tested across three critical diagnostic tasks: keratinocyte carcinoma classification, melanoma classification and melanoma classification using dermoscopy
Haenssle et al (2018)	100 000 images of malignant melanoma and benign naevi	58 dermatologists (30 experts)	CNN ROC curve revealed a higher specificity of 82.5% compared to dermatologists with (75.7%, $P<0.01$) and without provision of clinical information (71.3%, $P<0.01$)
Yang et al (2018)	SD-198 dataset with 6584 clinical images and 198 diseases	General doctors ($n=2$), junior doctors ($n=2$), expert doctors ($n=2$)	Deep features derived from the fine-tuned Resnet resulted in the highest classification accuracy (53.35%): general doctor (49%), junior doctors (52%), expert (83.29%)
Maron et al (2019)	11 444 dermoscopic images used to train the CNN. A test set of 300 biopsy-verified images used to compare the classifier's performance with that of dermatologists	112 dermatologists from 13 German hospitals	Sensitivity and specificity of dermatologists for the primary end-point (correct classification of lesions into benign and malignant) were 74.4% and 59.8% respectively. At equal sensitivity, the algorithm achieved a specificity of 91.3%. For the secondary end-point (correct classification of the images into one of five diagnostic categories), the mean sensitivity and specificity of the dermatologists were at 56.5% and 89.2% respectively. At equal sensitivity, the algorithm achieved a specificity of 98.8%
Hekler et al (2019)	11 444 dermoscopic images were used to train a single CNN. The dermatologists and the CNN independently classified a set of 300 biopsy-verified skin lesions into five categories	112 dermatologists from 13 German university hospitals	By combining the CNN with dermatologist review, sensitivity increased from 86.1% (exclusively decided by CNN) to 89% by incorporating the human decision ($P<0.05$). At an equal sensitivity of 89%, the CNN achieved a specificity of only 81.5%
Fujisawa et al (2019)	4867 clinical images (including 14 diagnoses) obtained from 1842 patients diagnosed with skin tumours	13 consultant dermatologists; 9 dermatology trainees	CNN achieved 96.3% sensitivity for identifying malignant lesions and 89.5% specificity. CNN achieved accuracy (of identifying benign *vs* malignant) of 92.4±2.1% ($P<0.001$), dermatology consultants had an accuracy of 85.3±3.7% ($P<0.01$), while that of trainees was 74.4±6.8% ($P<0.01$)
Brinker et al (2019a)	4204 biopsy-proven melanoma images used to train the CNN. 804 images were evaluated, each by a mean of 21 dermatologists	74 dermatologists; 26 consultants; 56 trainees from 9 German hospitals	The respective sensitivity and specificity of lesion classification by the dermatologists were 67.2% (95% CI 62.6–71.7%) and 62.2% (95% CI 57.6–66.9%). The CNN achieved a higher sensitivity of 82.3% (95% CI 78.3–85.7%) and higher specificity of 77.9% (95% CI 73.8–81.8%)
Brinker et al (2019b)	12 378 images used to train the CNN. 100 clinical images were used to compare the CNN to the dermatologists	145 dermatologists from 12 German hospitals	The mean sensitivity and specificity achieved by the dermatologists was 89.4% (range 55.0–100%) and 64.4% (range 22.5–92.5%), with ROC of 0.769 (range = 0.613–0.9). At the same sensitivity, the CNN had a mean specificity of 68.2% (range 47.5–86.25%). The chief physicians showed the highest mean sensitivity of 92.8% at a mean specificity of 57.7%. With the same sensitivity of 92.8%, the CNN had a mean specificity of 61.1%
Brinker et al (2019c)	12 378 images used to train the CNN. 100 clinical images were used to compare the CNN to the dermatologists	157 dermatologists from 12 German hospitals	The dermatologists' mean sensitivity and specificity was 74.1% (range 40.0–100%) and 60% (range 21.3–91.3%) respectively. At a mean sensitivity of 74.1%, the CNN exhibited a mean specificity of 86.5% (range 70.8–91.3%). At a mean specificity of 60%, a mean sensitivity of 87.5% (range 80–95%) was achieved by the CNN. The CNN outperformed 136 of the 157 dermatologists. The chief physicians showed the highest mean specificity of 69.2% and sensitivity of 73.3%. With the same specificity of 69.2%, the CNN had a mean sensitivity of 84.5%

CI=confidence interval; CNN=convolutional neural network; ROC= receiver operating characteristics.

However, studies thus far are not representative of, or generalisable to, the real-world integration of artificial intelligence in clinical practise. The main limitations include dermatologists' diagnostic performance being evaluated in artificial settings without clinical information, and the test-sets usually did not including a diverse range of lesions or skin types, or if they did, artificial intelligence algorithms did not score as highly on identifying lesions like pigmented basal cell carcinomas, actinic keratoses, seborrheic keratoses or spitz naevi (Fujisawa et al, 2019).

Benefits of artificial intelligence in dermatology

In the clinical setting, artificial intelligence could be integrated into routine consultation to help the physician reach a diagnosis (Mar and Soyer, 2018). Artificial intelligence has the potential to save a substantial amount of money for health-care systems, although no dermatology-specific cost–benefit analysis data for artificial intelligence have yet been published (Jiang et al, 2017). From a global health perspective, artificial intelligence may prove transformative in resource-poor settings (Jiang et al, 2017). Furthermore, artificial intelligence can aid in the management of high-risk individuals who require total body photography and life-long surveillance for melanoma detection by being combined with two- and three-dimensional imaging technologies (Mar and Soyer, 2018). This would permit automated filtering of benign lesions, allowing more efficient management of suspicious lesions by specialists.

Beyond the clinical setting, artificial intelligence could be incorporated into smartphone applications to improve patient-led surveillance methods – fractal analysis algorithms aim to risk stratify lesions and could serve as a triaging tool to hasten diagnosis (Mar and Soyer, 2018). An application development by First Derm uses artificial intelligence to propose a differential diagnosis and provide user-friendly information on each one. Currently, the platform correctly identifies a condition with 40% accuracy and identifies the top five potential diagnoses with 80% accuracy; the accuracy is expected to increase over time as the database expands (Börve, 2018).

Limitations of artificial intelligence in dermatology

There are concerns surrounding the safety and regulatory control of publicly available applications claiming to diagnose skin cancer; most have not been validated and rigorous trials are lacking (Mar and Soyer, 2018). Non-specialist users may focus on naevi in exposed areas rather than more discrete lesions, which would be identified during full body dermatological examination, putting patient health at risk. Artificial intelligence may identify more false positives than clinicians, creating unnecessary anxiety, while amelanotic melanomas may be misdiagnosed, severely compromising patient health (Jiang et al, 2017).

Since studies have focused on developing convolutional neural networks for melanoma detection, artificial intelligence's usefulness in differentiating between non-melanoma skin cancers and pre-cancerous conditions such as actinic keratoses and Bowen's disease, where texture is an important feature, has not been well investigated. Furthermore, total body photography captures ~90–95% of the skin surface, limiting the utility of artificial intelligence in evaluating difficult to photograph areas such as the fingers, toes and scalp (Mar and Soyer, 2018).

Dermatological conditions can elicit psychosocial comorbidities, especially where a diagnosis of cancer is concerned, which cannot be addressed by artificial intelligence. Holistic patient assessment, psychological support, management and prevention advice can only be offered by physicians. Data comparing diagnoses from large groups of dermatologists to those from convolutional neural networks are lacking. Additionally, there is little experience of patients with non-Caucasian skin because fewer validated images have been inputted into convolutional neural network algorithms.

The future of artificial intelligence in dermatology

It is paramount that artificial intelligence systems are subjected to strict criteria and testing so that evidence-based medicine procedures are adhered to. Applications should

Key points

- The rise in melanoma incidence and high mortality rates has fuelled the development of advanced diagnostic techniques.
- Studies have reported comparable performance between convolutional neural networks and dermatologists.
- Convolutional neural networks have inadequate experience in diagnosing atypical melanomas, non-melanoma skin cancers and pre-cancerous lesions.
- Artificial intelligence cannot substitute holistic patient care offered by dermatologists in the clinical setting.
- Concerns regarding the accountability for inaccurate diagnoses made by artificial intelligence need to be addressed.
- Prospective studies are required to collect real-world generalisable data for the performance and acceptability of artificial intelligence in dermatology diagnostics.
- Artificial intelligence systems may serve as support tools to supplement the diagnosis of challenging lesions.

be regularly reviewed to ensure that they do not deceptively claim accurate diagnosis or calculation of melanoma risk, and that scientific evaluation of applications is publicised. Furthermore, concerns regarding regulation, data security, ownership of images and accountability for inaccurate diagnoses made by artificial intelligence need to be addressed (Jiang et al, 2017).

Prospective studies assessing the real-world impact and usefulness of convolutional neural networks for physicians and patients are required before artificial intelligence can be implemented into clinical practice. Economic evaluation of artificial intelligence in the context of the NHS in the UK is also warranted (Jiang et al, 2017). The success and development of artificial intelligence is dependent on large, high-quality, diverse data sets and continuous development using training data; artificial intelligence algorithms are only as effective as the data used to train them (Jiang et al, 2017). To address this, the International Skin Imaging Collaboration: Melanoma Project produced an archive of dermoscopic skin lesions for education and research. Early engagement of dermatologists in the potential implementation of artificial intelligence is essential to ensure that it is suitable for clinical use and that it does indeed benefit clinicians. Similarly, both the patient and physician's acceptability of artificial intelligence as a screening or diagnostic tool for skin cancer should be addressed and explored.

Conclusions

Use of artificial intelligence remains largely experimental in dermatology. Maintaining the doctor–patient relationship is central to the provision of holistic medical care and the dermatologist's flexibility, adaptability and experience cannot be substituted. A likely future scenario may involve artificial intelligence systems with rapid analytic skills serving as support tools to physicians of all levels of training and experience, supplementing and augmenting the diagnosis of challenging lesions (Mar and Soyer, 2018; Hekler et al, 2019).

The words of Jonathan Chen, at Stanford's Human Intelligence & Artificial Intelligence in Medicine Symposium in 2018, remain pertinent: 'Is the computer smarter than the physician? It's irrelevant. Together they can provide something better than either could alone.'

Author details

[1]College of Medical & Dental Sciences, University of Birmingham, Birmingham, UK

[2]Department of Dermatology, Wye Valley NHS Trust, Hereford, UK

References

Börve A. 2018. Dermatology (skin diseases) is well suited using artificial intelligence (AI). (accessed 4 October 2019) https://www.firstderm.com/artificial-intelligence-ai-dermatology-skin-diseases-first-derm/

Brinker TJ, Hekler A, Enk AH et al. Deep neural networks are superior to dermatologists in melanoma image classification. Eur J Cancer. 2019a;119:11–17. https://doi.org/10.1016/j.ejca.2019.05.023

Brinker TJ, Hekler A, Enk AH et al; Collaborators. A convolutional neural network trained with dermoscopic images performed on par with 145 dermatologists in a clinical melanoma image classification task. Eur J Cancer. 2019b;111:148–154. https://doi.org/10.1016/j.ejca.2019.02.005

Brinker TJ, Hekler A, Enk AH et al; Collaborators. Deep learning outperformed 136 of 157 dermatologists in a head-to-head dermoscopic melanoma image classification task. Eur J Cancer. 2019c;113:47–54. https://doi.org/10.1016/j.ejca.2019.04.001

Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, Thrun S. Dermatologist-level classification of skin cancer with deep neural networks. Nature. 2017;542(7639):115–118. https://doi.org/10.1038/nature21056

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(2):373–381. https://doi.org/10.1111/bjd.16924

Haenssle HA, Fink C, Schneiderbauer R et al; Reader study level-I and level-II Groups. 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(8):1836–1842. https://doi.org/10.1093/annonc/mdy166

Hekler A, Utikal JS, Enk AH et al; Collaborators. Superior skin cancer classification by the combination of human and artificial intelligence. Eur J Cancer. 2019;120:114–121. https://doi.org/10.1016/j.ejca.2019.07.019

Jiang F, Jiang Y, Zhi H et al. Artificial intelligence in healthcare: past, present and future. Stroke Vasc Neurol. 2017;2(4):230–243. https://doi.org/10.1136/svn-2017-000101

Mar VJ, Soyer HP. Artificial intelligence for melanoma diagnosis: how can we deliver on the promise? Ann Oncol. 2018;29(8):1625–1628. https://doi.org/10.1093/annonc/mdy193

Maron RC, Weichenthal M, Utikal JS et al; Collabrators. Systematic outperformance of 112 dermatologists in multiclass skin cancer image classification by convolutional neural networks. Eur J Cancer. 2019;119:57–65. https://doi.org/10.1016/j.ejca.2019.06.013

National Institute for Health and Care Excellence. 2015. Melanoma: assessment and management. (accessed 14 October 2019) www.nice.org.uk/guidance/ng14

Yang J, Sun X, Lang J, Rosin PL. 2018. Clinical Skin Lesion Diagnosis Using Representations Inspired by Dermatologist Criteria. The IEEE Conference on Computer Vision and Pattern Recognition (CVPR). https://doi.org/10.1109/CVPR.2018.00137