# Man versus machine: will machines replace dermatologists in the diagnosis of skin disease?

**Nikita Quinn**
Christchurch School of Medicine
Otago Medical School
University of Otago

**Nikita is a 5th year medical student based at Christchurch Hospital. She has particularly enjoyed her surgical attachments. Prior to medicine she completed a degree in Economics. This essay was the University of Otago winner of the Wilson-Allison Memorial Prize in Dermatology.**

Research into the role of artificial intelligence in medicine is rapidly growing. In 2016, healthcare-related artificial intelligence projects attracted more investment than artificial intelligence projects in any other sector of the global economy.[1] Artificial intelligence is a general term that refers to the use of a computer to model intelligent behaviour with minimal human intervention.[2] Recent advances in this field are numerous and include taking steps towards the automatic detection of diabetic retinopathy, better interpretation of radiography and more efficient diagnosis of skin cancer. In particular, the use of artificial intelligence to distinguish between malignant melanoma and benign lesions has garnered a lot of attention.

Skin cancer remains a major public health issue in New Zealand, with recent data revealing New Zealand has the second highest rate of melanoma in the world.[3] The 2018 skin cancer index published by German medical analyst group derma.plus stated almost 2,500 new melanoma cases are diagnosed in New Zealand every year.[4] Early detection of melanoma is critical to patient prognosis and survival. The five-year survival rate of early stage melanoma is 99%, falling to only 20% for melanoma that has spread to distant sites in the body.[5]

Currently the process of diagnosing a malignant lesion begins with visual examination by the general practitioner or dermatologist. Many physicians will also use a dermatoscope, a hand held microscope that provides low level magnification of the lesion. If these methods are inconclusive or lead the physician to suspect the lesion may be cancerous, a biopsy and subsequent histopathological examination are the next steps.[6] However, accurately distinguishing which lesions require a biopsy and which do not is often poorly achieved by medical professionals. Dermatologists and other medical practitioners formally trained in this field have been shown to have an average sensitivity for detecting melanoma of less than 80%.[7] This can have damming consequences for the patients affected, given the imperativeness of diagnosing melanoma at the earliest possible stage.

In recent years a lot of work has been carried out to develop automated computer image analysis of skin lesions, with the hope this may help physicians to more accurately identify potentially dangerous lesions. Traditional methods have focused on teaching computers to identify suspicious lesions on the basis of certain 'manmade criteria' such as lesions with an asymmetrical appearance, irregular border or multiple colours.[7] In 2017, a landmark paper from researchers at Stanford university proposed that the recognition of malignant lesions via machine learning was a feasible alternative.[8] The basis of machine learning is that the computer is programmed to 'figure out' the answers itself, rather than having answers pre-programmed into it. Not being restricted to certain man-made criteria allows a much broader range of malignant lesions to be identified, which is useful given the large variation that is seen in the appearance of melanoma.[7]

In 2018, leading cancer journal Annals of Oncology published a study that showed that a form of machine learning known as a deep learning convolutional neural network (CNN) was in fact better than most dermatologists at detecting skin cancer.[7] A CNN is an artificial neural network inspired by the biological processes used when neurons in the brain make connections with each other and respond to what is seen with our eyes.[9]

In the study, researchers from Europe and the United States of America trained a CNN to identify melanoma by showing it more than 100,000 dermoscopic images of the disease, as well as benign naevi, and attaching to each image what the correct diagnosis was. The network was able to learn rapidly from example, by deconstructing each image down to the pixel level, and creating its own diagnostic clues for classifying the images. After training the computer, the researchers created a set of 100 test images which again comprised both melanomas and benign naevi (these images had not been used for training and therefore had never been seen by the CNN before). The images were used to test the CNN and compare its performance to dermatologists around the world. 58 dermatologists agreed to participate in the study. In the first instance (level I), the dermatologists were shown each image on its own and asked to make a diagnosis of melanoma or benign naevi, and to indicate how they would manage the lesion (either surgical excision, short term monitoring of the lesion, or no further action required). In the second phase of the study (level two), the dermatologists were again shown each image and asked for a diagnosis and management decision, however this time they were also supplied with some additional clinical context (including the age and sex of the patient, and the location of the lesion).

In level one, the dermatologists on average correctly diagnosed 86.6% of melanomas, and 71.3% of benign naevi. When the CNN was tuned to have the same specificity as the dermatologists (i.e. to correctly identify 71.3% of benign naevi), the CNN was able to identify 95% of melanomas. The clinical context provided in level two of the study significantly improved the dermatologists' performance such that they accurately identified 88.9% of melanomas and 75.7% of benign naevi. However, while the performance of dermatologists improved when provided with more clinical information, the CNN continued to outperform them even at this level. These findings suggest the increased sensitivity and specificity provided by the CNN could result in fewer missed melanomas as well as less unnecessary biopsies if implemented into clinical practise.[7]

Lead researcher, Professor Holger Haenssle, from the University of Heidelberg, stated he does not envisage the CNN will replace dermatologists in diagnosing skin cancer, but that it could be used as an additional aid. 'Most dermatologists already use digital dermoscopy systems to image and store lesions for documentation and follow up. The CNN can then easily evaluate the stored image for an "expert opinion" on the probability of melanoma.[9]

As discussed above, at present the decision to investigate a skin lesion is dependent on the opinion of the treating clinician. Research has suggested the accuracy of this can vary widely depending on the training and experience of the doctor in question. It is hoped the use of automated computer image analysis may help to standardise the level of diagnostic accuracy seen across the world, such that all patients, regardless of where they live or which doctor they see, will be able to access the same level of care.[9]

While the technology currently exists on computers, there is a possibility it could become available as a smartphone app in the future, allowing almost ubiquitous access to skin lesion analysis right at our fingertips. There is also the potential for this technology to be used in combination with 2-D or 3-D total body skin imaging systems. These imaging systems are currently able to image close to 90–95% of the skin surface. This would mean the majority of a patient's benign lesions could be filtered by the machine, allowing dermatologists to focus more of their time on the more suspicious or concerning lesions. In addition, one of the major issues pertaining to the implementation of a melanoma screening programme is the lack of a suitable test – in that a whole body inspection by a physician lacks both sensitivity and specificity. The CNN may fill this gap by acting as a more precise screening tool.[6]

While this is an exciting development in the diagnosis of skin cancer, the concept is not without limitations. Firstly, in regards to the study, the dermatologists knew they were in an artificial setting and therefore were not making 'life or death' decisions. Difficulty in accessing validated images meant there was a lack of images from non-Caucasian ethnicities, raising concerns about the accuracy of the CNN when applied to a broader range of real-world settings. In addition, as this study shows, clinical context is crucial. Clinicians were not able to examine the rest of the patients' skin and look at their other moles and they could not ask questions such as what sun exposure the patient had experienced throughout their lifetime, if they had ever had a previous skin cancer, or if there was any relevant family history.[7] These are things that can be ascertained very quickly in a real-life clinical setting and would likely have a significant impact on a doctor's clinical decision making.

Further refining of the technology is also needed. Areas of the body that are difficult to image such as the scalp, fingers, and toes are problematic for this type of technology.[7] In addition, researchers have discovered the CNN can be tricked in unexpected ways. For example, previous studies have shown lesions with a ruler in the image are much more likely to be deemed malignant by the machine. This is because dermatologists are more likely to measure lesions they are concerned about and thus within the portfolio of validated images available for training, malignant lesions are more likely to have been photographed with a ruler.[8] This bias occurs due to the technology analysing the image in its entirety, rather than just the lesion alone. Other situations that could fool the technology could be unusual combinations of lesions such as a benign naevus in close proximity to a seborrheic keratoses, which could closely mimic a melanoma.[7] This also highlights another downfall of CNN technology – it is a black box system. This means that we do not know exactly what diagnostic clues the machine is using to formulate its diagnosis and thus its implementation is opaque.[10] If no clinician is involved in the diagnostic process, this could also lead to issues of accountability when the machine gets it wrong.[7]

It is also important to consider the impact this technology may have on the health-care system. Widespread adoption of a skin analysis app by consumers poses the potential for a flood of real and potential skin cancers to pour into the health-care system – rather than being replaced by machines, dermatologists may end up busier than ever. Dr Allan Halpern, chief of dermatology at the Memorial Sloan Kettering cancer centre in New York, stated 'what's not clear is what percentage of cancer cases can be left alone. Assuming there are a lot of cases that right now go undiagnosed, if all of a sudden artificial intelligence can bring all those cases into the healthcare sphere, it'll be enormous'.[11] This also raises the possibility of increased harm from overdiagnosis. It is possible increased analysis of skin lesions may result in skin cancers being diagnosed that would never have caused the patient any harm in the first place, resulting in unnecessary treatment.[11]

All in all, the use of artificial intelligence in the diagnosis of skin disease is likely to become a useful aid for dermatologists, however it is unlikely to ever replace them. The above research only relates to the diagnosis of melanoma, however, dermatologists are instrumental in diagnosing hundreds of different skin conditions. Furthermore, making a diagnosis is only the tip of the iceberg – dermatologists must then educate patients about their diagnosis, support them through the appropriate treatment, and guide them on how to best prevent future disease. In addition, many technological issues still need to be resolved, such as how to avoid the machine being tricked and how to image difficult areas such as the fingers, toes, and scalps.[7] More real-world research is also needed before the use of this technology can become widespread, including research on how acceptable using artificial intelligence to make a diagnosis would be to patients and clinicians. There is no guarantee clinicians would follow the recommendations of the machine, particularly if they do not entirely trust it.[7]

Overall, the use of artificial intelligence in the diagnosis of skin disease is a promising area of research that may well become an integral part of a dermatologist's tool kit in the future. This is also an exciting development for current medical students who are likely to see artificial intelligence become integrated into, not only the diagnosis of skin cancer, but across more and more areas of health care throughout their future careers. In summary, while artificial intelligence is likely to be a valuable resource, it is unlikely to ever become a full substitute for seeing a clinician and therefore, dermatologists should be encouraged to view artificial intelligence as an exciting opportunity rather than a threat.

## References

1. Buch V, Ahmed I, Maruthappu M. Artificial intelligence in medicine: current trends and future possibilities. Br J Gen Pract. 2018;68(668):143–4.

2. Harnet P, Tremblay J. Artificial intelligence in medicine. Metabolism. 2017;69:36–40.

3. American Institute for Cancer Research. Skin cancer statistics [Internet]. 2018. Available from: https://www.wcrf.org/dietandcancer/cancer-trends/skin-cancer-statistics

4. derma.plus. Skin cancer-index [Internet]. 2018 [cited 2018 Oct 25]. Available from: https://derma.plus/en/global-skin-cancer-index/

5. American Cancer Society. Survival rates for melanoma skin cancer by stage.Stanford News Service; 2017.

6. Kubota T. Deep learning algorithm does as well as dermatologists in identifying skin cancer [Internet]. Available from: https://news.stanford.edu/press-releases/2017/01/25/artificial-inteltify-skin-cancer/

7. Haenssle H, Fink C, Schneiderbauer, Tober F, Buhi T, Blum A, 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(8):1836–42.

8. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, et al. Dermatologist-level classification of skin cancer with deep neural networks. Nature. 2017;542(7639):115–8.

9. European Society for Medical Oncology. Man against machine: artificial intelligence is better than dermatologists at diagnosing skin cancer [Internet]. 2018. Available from: https://www.esmo.org/Press-Office/Press-Releases/Artificial-Intelligence-Skin-Cancer-Diagnosis

10. Swetter S. Will artificial intelligence replace dermatologists? Conference presentation presented at American Academy of Dermatology Annual Meeting. 2018;San Diego, California.

11. Bowes J. Should dermatologists fear machine learning? J Am Acad Dermatol. 2018;28(2)42–6.

**Correspondence:**

Nikita Quinn: quini093@student.otago.ac.nz