

Artificial intelligence and melanoma diagnosis: ignoring human nature may lead to false predictions

Aimilios Lallas¹, Giuseppe Argenziano²

¹ First Department of Dermatology and Venereology, Aristotle University, Thessaloniki, Greece

² Dermatology Unit, University of Campania, Naples, Italy

Citation: Lallas A, Argenziano G. Artificial intelligence and melanoma diagnosis: ignoring human nature may lead to false predictions. *Dermatol Pract Concept*. 2018;8(4):249-251. DOI: <https://doi.org/10.5826/dpc.0804a01>

Published: October 31, 2018

Copyright: ©2018 Lallas et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Corresponding Author: Aimilios Lallas, MD, First Department of Dermatology and Venereology, Aristotle University, 124 Delfon Street, 54623, Thessaloniki, Greece. Email: emlallas@gmail.com

Having spent most of our waking hours of the last decades working on, thinking and writing about, teaching, and discussing skin cancer diagnosis, this is the first time that we feel compelled to address a question that is so often asked: Are we (doctors) going to be left out of the game by artificial intelligence (AI)?

This question is usually asked in combination with an expression strongly suggestive of an underlying feeling: fear. This is absolutely understandable, to be honest. What kinds of feelings should we expect of a doctor who has spent a lifetime trying to improve his/her capacity to diagnose melanoma and who then reads that “machines perform at expert level or above” in recognizing melanoma? Joy? Happiness? Hope? Let’s be realistic—they are confusion, fear, and anger.

The resistance of professionals to technological developments that threaten to replace them might sound similar to what happened at the time of the Industrial Revolution. One could argue that now, like then, if machines can indeed perform better than humans, the reluctance of doctors to accept this will not be reason enough to delay the advance of this new development.

Going back to the key question about whether doctors will be replaced by automated algorithms for skin cancer diagnosis, our honest answer is, we don’t know. We don’t

think so, but we are not sure. In contrast, many other scientists are much more confident: Melanoma in the future will be diagnosed by AI, the only question is how soon it will happen. But with all the tremendous efforts of scientists and the investments of huge companies, it should be quite soon.

In fact, never before have we seen so many researchers in our field focusing their efforts on the same topic. Never before have we seen so many studies published in such a short period of time with the same aim: to develop algorithms that diagnose melanoma equally as or better than doctors. Most of them succeed in demonstrating this is so.

Without aiming to shake the confidence of those who foresee algorithms replacing doctors, we think that they might benefit by taking into account the following considerations:

- 1. What has been shown to date:** All studies comparing melanoma diagnosis by AI with diagnosis by humans have been conducted in an experimental setting. To be clear, these studies were conducted in front of a computer, tablet, or smartphone and were based on evaluation of images. None of these algorithms has been tested in a real clinical setting (though the reason is not known). Comparing a game with images played in front of a computer with a diagnosis made in a clinical setting is like comparing a car racing video game with a live Formula 1 race. The clinical

examination of patients is a much more multifactorial, unpredictable, and complicated process, as compared to evaluation of a clinical or dermoscopic photograph in front of a computer. This is quite clear for any doctor who works in clinical practice even for a short period.

2. **The accuracy of diagnosis:** Most of the algorithms are fed by images of histopathologically diagnosed lesions. Therefore, their great performance always has histopathology as a reference point and presupposes that the histopathological diagnosis was correct. However, it has been well documented that the interpretation of histopathology, especially for “borderline” melanocytic lesions, is far from what we would expect from a “gold standard” method. Last year, in *The BMJ*, Elmore et al reported a disagreement rate among different pathologists up to 75% (!) when trying to differentiate between nevi with moderate to severe dysplasia and early melanoma [1]. Therefore, AI algorithms are supplied with images of melanomas that might have been diagnosed as nevi by another pathologist and nevi that might have been diagnosed as melanomas by another pathologist. Can we imagine the potential effects of this? Clinicians are able to deal with this problem because they are aware of this limitation. They also know that morphology (clinical, dermoscopic, histopathological) does not always accurately predict biology. Melanomas that look like nevi do exist and the inverse is true as well. For this reason, clinical management decisions are not made only on the basis of morphology. For AI, all this is not comprehensible. AI requires clear endpoints (benign-malignant). “Don’t know” does not exist for AI, and this creates a huge risk.
3. **The usefulness of mistakes:** Throughout history, scientific knowledge has improved principally by learning from human error. Humans learn from their mistakes, and this has proved to be the foundation for progress. As soon as a doctor identifies a mistake, he/she tries to explain why it happened and what should be done to avoid repeating it. AI systems train themselves, without human guidance. They learn fast—much faster than we do—and we are unable to fully understand the way they become so accurate so quickly. As much as we fail to understand how an algorithm is able to accurately classify a lesion when humans do not, we also fail to adequately explain why AI is wrong when it is wrong. Therefore, we will never really know why a mistake happened, and it is quite likely that AI will repeat the same mistakes because we cannot train it to avoid them.
4. **The value of research:** Innovation is great, but the value of research is not measured by how innovative it is or by the impact factor of the journal in which it is published. The value of research is measured by its impact on humans. The discovery of penicillin was a great innovation that saved millions of people. The discovery of atomic bombs

and nuclear weapons was a great innovation that killed millions of people. It goes without saying that these 2 discoveries did not produce results with the same value. In fact, the results were polar opposites: extremely beneficial and extremely detrimental. If we can agree that the principal value of our ethical system is human life, then the value of research is assessed by the impact on it. In our field, for example, the research on epidemiology of melanoma has been of high value because it identified high-risk groups, positively affecting humanity. The research on histopathology of melanoma has been of high value because it identified prognostic factors, positively affecting humanity. The research on dermoscopy of melanoma has been of high value because it allowed for earlier detection, positively affecting humanity. The research on genetics of melanoma has been of high value because it identified mutations that might be potential targets of drugs. The research on drugs for metastatic melanoma has been of high value because it prolonged the survival of patients. The research on melanoma diagnosis with AI has not, as of this date, had a positive effect on humans. The only system that attempted to enter the clinical practice (that one with 98.5% sensitivity but 10% specificity) was never adopted by clinicians because it would have resulted in the unnecessary excision of millions of nevi [2]. What is good and what is bad is decided by the ethics of a society and not by mathematical models. AI recognizes no ethics—only mathematical models. Are we so sure that we are (or will be) able to protect our ethics if AI takes the lead?

5. **The nature of humans:** Humans like to interact with humans, in general, and even more in medicine. The moment that a physician examines a patient is a unique interaction during which a human being uses all available knowledge and mental effort to help another human being. There is a lot of interchange of energy in this procedure. This is very much superior to a simple judgment on the benign or malignant nature of a lesion. The result of a medical consultation is not measured only by the absolute improvement of the patient’s physical health. Think for a moment about patients with end-stage metastatic cancer, those not responsive to treatment, or patients with diseases with no available treatments. They build a strong relationship with their doctor, which is not measurable or explainable by the absolute improvement of their physical health. To think that medical care can be simply conducted by mathematical models is tantamount to ignoring human nature.

Even if someday it will be possible to satisfactorily address points 1 through 4 made above, ignoring human nature and investing against it is very likely a prelude to failure.

We are convinced that AI has the potential, in several ways, to become an additional precious tool in the hands of doctors struggling to reduce melanoma mortality. The main

obstacles of this goal are the misconceptions about our role as doctors. Maybe by rereading and rethinking what Hippocrates said 2,500 years ago we could find the way.

References

1. Elmore JG, Barnhill RL, Elder DE, et al. Pathologists' diagnosis of invasive melanoma and melanocytic proliferations: observer accuracy and reproducibility study. *BMJ*. 2017 Jun 28;357:j2813.
2. Monheit G, Cognetta AB, Ferris L, et al. The performance of MelaFind: a prospective multicenter study. *Arch Dermatol*. 2011;147(2):188-194.

AIMILIOS LALLAS, MD

Deputy Editor

GIUSEPPE ARGENZIANO, MD

Editor-in-Chief