ABSTRACT

Artificial intelligence (AI) based on machine learning and convolutional neuron networks (CNN) is rapidly becoming a realistic prospect in dermatology. Non-melanoma skin cancer is the most common cancer worldwide and melanoma is one of the deadliest forms of cancer. Dermoscopy has improved physicians’ diagnostic accuracy for skin cancer recognition but unfortunately it remains comparatively low. AI could provide invaluable aid in the early evaluation and diagnosis of skin cancer. In the last decade, there has been a breakthrough in new research and publications in the field of AI. Studies have shown that CNN algorithms can classify skin lesions from dermoscopic images with superior or at least equivalent performance compared to clinicians. Even though AI algorithms have shown very promising results for the diagnosis of skin cancer in reader studies, their generalizability and applicability in everyday clinical practice remain elusive.

Herein we attempted to summarize the potential pitfalls and challenges of AI that were underlined in reader studies and pinpoint strategies to overcome limitations in future studies. Finally, we tried to analyze the advantages and opportunities that lay ahead for a better future for dermatology and patients, with the potential use of AI in our practices.

PLAIN LANGUAGE SUMMARY

Artificial intelligence (AI) is the development of computer systems able to perform tasks that normally require human intelligence, such as visual perception, speech recognition, and translation between languages. The research on the use of AI in dermatology includes the ability of a machine to correctly classify a skin lesion. Skin cancer is the most common cancer worldwide and melanoma is the deadliest form of skin cancer. All skin cancers have a better prognosis when detected early in their development, hence their early detection is of paramount importance. Dermatologists use a dermatoscope—a specialized magnifying lens to improve their diagnostic capacity. However, even with the use of the dermatoscope, their ability to recognize skin cancer is far from perfect. AI has the ability to learn from dermoscopic images and subsequently provide an image-based diagnosis. Several studies
compared the performance of machines and humans in classifying skin lesions from these images and showed that machines can classify skin lesions as good (and sometimes better) than humans. However, the application of AI in everyday clinical practice remains a challenge. In this article, we attempt to summarize the limitations and challenges that researchers found in their studies, and we provide strategies to improve the design of future studies. Finally, we describe the advantages and opportunities that could lay ahead for a better future for dermatology and patients.

**Keywords:** Melanoma; Prevention; Skin cancer; Dermoscopy; Diagnosis; Artificial intelligence; Machine learning; Teledermatology

### Key Summary Points

Artificial intelligence (AI) is rapidly gaining ground in medicine, and dermatology in particular. Research has shown that AI algorithms can diagnose skin conditions in general and skin cancer specifically with a diagnostic accuracy comparable with that of skin cancer experts.

However, there are several challenges that need to be addressed; AI algorithms are often tricked by perturbations in image quality, magnification, image color, as well as rulers, skin markings, and pen markings. The generalizability of AI algorithms and their potential use in clinical practice remains to be eluded. Real-life clinical trials using AI algorithms are needed in order to amplify their potential use in everyday practice.

AI algorithms could be of aid to dermatologists and patients, particularly in the fields of teledermatology, 3D imaging, and sequential digital dermoscopy, while AI’s applications could potentially prove beneficial to the entire field of dermatology.

### INTRODUCTION

Artificial intelligence (AI) is the development of computer systems able to perform tasks normally requiring human intelligence, such as visual perception, speech recognition, decision-making, and translation between languages [1]. AI has become an indispensable part of our daily lives, while it constantly penetrates more and more human activities [2]. The evolution of AI includes classic AI, followed by machine learning leading to the era of deep learning in which we currently live [3]. In machine learning (ML), algorithms are trained to perform tasks by learning from data rather than by precise programming instructions [4]. We are now facing the evolution of deep learning—a subset of machine learning that uses an artificial neural network (ANN) structure, inspired by the biological neural network [5]. In this form of ML, there is the capacity to use an unlimited number of layers, where each layer within the neural network can be trained to recognize different features specific to the dataset [6]. Convolutional neural networks (CNN) are a special form of neural networks that have dominated in the field of image processing [7]. CNN consist of convolutional, pooling layers and fully connected layers. The primary purpose of a convolutional layer is to detect distinctive visual features, and it is vital for successful image processing tasks such as segmentation and classification [8]. In order for CNNs to recognize these visual features on their own capacity, they initially require an abundant amount of training data [9].

This deep learning flow has given medical society the potential to evolve through AI [10]. Currently, major advancements have been made mainly in the field of radiology and cardiology. FDA-approved medical devices first gained approval in 2016, giving healthcare professionals the ability to enhance medical practice through AI [11]. In radiology, ML-based image reading algorithms are used on brain images for hemorrhage and stroke detection, for image processing improvements, as well in acute care for the assessment of pneumothorax and injuries. These algorithms...
augment radiologists’ practices by enabling faster diagnosis and alert for emergency situations. Also, there are available algorithms for mammography analysis and lesion detection. Applications in cardiology include electrocardiogram readings for the detection of cardiac rhythm abnormalities [11]. In the field of diabetes, there are FDA-approved medical devices for managing blood glucose levels using monitoring systems with predictive alerts [12], as well as the detection of diabetic retinopathy in ophthalmology [13]. It is also worth mentioning the existence of algorithms in clinical practice for the detection of sleep disorders [11]. Moreover, there are health systems that are using simple ML models based on electronic health records (HER) to stratify hospitalized patients in need for admitting to intensive care units [14]. The above-mentioned medical devices are just the beginning of this era. Raw data from EHR can be used for prognostic models. Diagnosis enhancement by AI can minimize diagnostic errors by physicians. The most appropriate treatment could be chosen for each individual patient and automatic selection of patients eligible for new treatments in clinical trials from EHR will grant ultimate patient outcomes [10].

Dermatology, being an image-based field of medicine, retains a prevailing position in the AI evolution. The main aspect of dermatology where AI has shown very promising results is the recognition of skin cancer [15–19]. Skin cancer includes non-melanoma skin cancers (NMSC), i.e., basal cell carcinoma, squamous cell carcinoma, and melanoma (MM). NMSC is the most common cancer worldwide, while melanoma is the fifth leading cause of cancer death in the US [20]. The mainstay of treatment for all subtypes of MM remains early recognition and surgical excision of the tumor [21]. The 5-year relative survival rate of localized MM is 99.5%, and unfortunately drops to 31.9% for distant MM [20]. Consequently, early diagnosis of skin cancer is inevitably the cornerstone for improving both mortality and morbidity outcomes.

Dermoscopy is a non-invasive imaging technique that uses polarized and non-polarized light to improve sensitivity and specificity for skin cancer diagnosis [22–24]. Moreover, most commercially available dermatoscopes have a standardized 10× magnification, while at the same time preserving patient identifiers, making it an ideal test case for ML training. Accordingly, a big proportion of AI research has focused on dermoscopic images for early skin cancer detection [15–19]. The overarching goal behind these efforts is to improve early skin cancer diagnosis and accordingly the mortality and morbidity resulting from both melanoma and NMSC. The ground for that lies in the fact that our diagnostic accuracy, as physicians, even with the use of dermoscopy remains comparatively low [25]. Additionally, there are two main challenges we must overcome. First, in several countries, including the US, access of the general population to dermatologists is difficult, leading to less than 25% of the adult population having ever had a total body skin examination (TBSE) by an expert dermatologist [26]—a diagnostic practice that can identify otherwise undiagnosed cutaneous malignancies [27, 28]. Second, only 25% of MM are diagnosed by a healthcare provider [29]. On this ground, AI could provide invaluable aid in the early evaluation and diagnosis of skin cancer.

The aim of this commentary is to provide a critical appraisal of current AI achievements and limitations in dermatology, focusing on relevancy in clinical practice. In addition, possible strategies to overcome these limitations and future perspectives are explored.

This article is based on previously conducted studies and does not contain any new studies with human participants or animals performed by any of the authors.

CURRENT ACHIEVEMENTS

In the last decade, there has been a breakthrough in new research and publications in the field of AI. This AI evolution has led to a compelling discussion in the scientific community regarding the potential role that AI could play [30]. Although some may perceive the advent of automated diagnosis as a threat, an effective AI system has the potential to improve the accuracy, accessibility, and efficiency of patient care.
The existence of a large public dataset (International Skin Image Collaboration Archive—ISIC) paved the road to remarkable research and became the reference standard for research in the field [32]. The results of landmark studies overemphasize AI’s usefulness, as they demonstrated superior or at least equivalent performance of CNN-based classifiers compared to clinicians [17]. Despite the fact that AI algorithms have shown very promising results for the diagnosis of skin cancer in reader studies, their generalizability and applicability in everyday clinical practice remains elusive. While there are already FDA-approved AI algorithms that have been embedded into clinical practice, mainly in radiology [11], there is no public clinical use of AI algorithm devices in dermatology. MelaFind, a device approved by the US Food and Drug Administration that used multispectral digital skin lesion analysis, has been shown to have high melanoma sensitivity and to improve both the sensitivity and specificity of dermatologists after clinical and dermoscopic examination of suspicious skin lesions via reader studies. Despite these apparent strengths, the device was discontinued in 2017 due to inadequacy of tangible benefit [33, 34].

PITFALLS IN AI

At this point, we need to identify the potential pitfalls of AI and pinpoint the advantages and opportunities that lay ahead. Many studies have shown that AI systems are susceptible to the presence of confounding factors, negatively impacting their classification performance [31, 35–38]. Those factors are mainly associated with variables regarding image quality and standardization. Perturbations in image magnification, adversarial “noise” (intended perturbations such as ink spots that aim to “confuse” MLA), image rotation, brightness/contrast manipulation [31], rulers, ink markings, blurry photos and dark corners of the tubular lens [35] are all variables that depend on the quality of the image that a clinician provides (Fig. 1). Specifically, a study showed that the AI algorithm appeared more likely to interpret images with rulers as malignant. The algorithm inadvertently was trained to recognize such findings as malignant as images presenting a MM had rulers more often than benign lesions [36]. Another study found that skin markings significantly interfered with CNN’s correct diagnosis of nevi by increasing the melanoma probability scores and consequently the false-positive rate, most likely because of the same reason [37]. Finally, a study found that the diagnostic accuracy of AI algorithms is heavily dependent on whether the image is in focus and well centered [38]. These biases in AI models are inherent unless specific attention is paid to addressing inputs with variability or incorporating stringent standards [36].

On the contrary, air bubbles, hairs, background skin diseases, sun-damaged skin, and peculiar anatomic sites [36] are confounding factors impacting the CNNs’ performance that cannot be eliminated by humans. A study showed that entities that do not generally present with crust (such as vascular lesions, dermatofibromas, and nevi) were frequently miscategorized in the presence of crusts [19]. The same study showed that the presence of hair affected the misclassification of actinic keratosis (36 vs. 56% without hair) [19]. Finally, another study showed that the anatomic site of a lesion plays a critical role in the performance of CNN [38]. Studies [18, 39–41] have shown the potential of CNN-based classification for special anatomic sites—such as face, palms, and soles that have different normal dermoscopic signs—but more extensive and diverse datasets as well as further research are needed to extend the application of AI in rare anatomic sites (e.g., genital area) and rare skin cancer subtypes (e.g., mucosal or desmoplastic MM) [17]. On the other side, banal-looking, benign lesions such as angiomas, dermatofibromas, or nevi are most often underrepresented or absent from studies’ training sets, leading to underperformance of the algorithms [18]. Inclusion of typical benign lesions avoids verification bias and thus eliminates such limitations as shown by a study from Tschandl et al. [42].

Beyond standardization pitfalls, AI technology must additionally overcome generalizability limitations. A frequent critique to both
artificial intelligence researchers and the ISIC Archive highlights that the training datasets for AI algorithms mainly consist of Caucasian patients, thus limiting the representation of possible variation in disease presentation [43].

Moreover, a study showed that rarer distributions of specific skin lesions, such as non-pigmented nevi and non-pigmented MMs decreased the accuracy of the algorithms’ classifications compared to common distributions. These results highlight that algorithms should be tested on both usual and unusual types of lesions and imaging attributes [19].

Finally, more research is needed on clinical close-up images combined with dermoscopic images as defined by combined convolutional neuron networks (cCNN) in order to provide a more accurate and realistic presentation of the lesion examined. These close-up images can provide additional datasets for future AI applications in preclinical evaluations [10]. Overall, deep learning is an intensely data-demanding technology, requiring an abundant number of labeled examples to achieve accurate classification [44]. ISIC is the genesis of a publicly available image dataset that needs to expand with the synergist approach of frontline physicians in order to develop and train classifiers with the best outcomes, as ML can only be as good as the quality of data it gets.

**CLINICAL LIMITATIONS**

Clinical evaluation, including patient history and patient examination, is the groundwork for every physician. In the artificial settings of challenges and studies to evaluate CNN’s performance, the clinician’s ability is often underestimated [17]. Many studies acknowledge the
lack of inclusion of clinical-related factors such as age, sex, degree of sun damage, anatomic site, and personal and family history [42, 43]. In a clinical environment, dermatologists would consider total body examination for comparison of variabilities such as the macroscopic Ugly Duckling sign (a nevus that stands out from the rest in a given individual) [45], the Little Red Riding Hood sign (a nevus that looks benign but differs from the rest in a given individual) [46], as well as dermatoscopic predominant nevus patterns (defined as the pattern seen in more than 30% of all nevi) [46, 47]. These approaches increase sensitivity and specificity but require a more holistic approach [42]. Additionally, in an experimental design, in vivo dermoscopy has been shown to be intrinsically better than in the artificial setting solely based on digital images [48].

Consumer trust in AI is an additional barrier to the application of AI in clinical practice. A recent study tried to shed some light on understanding citizens’ trust and expectations concerning the use of AI across multiple countries. Researchers highlighted that concerns about the adequacy of current regulations and laws for the safe use of AI is the strongest driver that influences their trust. Also, 63% of them reported being unwilling or ambivalent about trusting AI in healthcare [49]. The black box nature of ML lies in the fact that there is no explanation and consistent process on how the algorithm reached a specific diagnosis. This approach can potentially lead to trust issues from patients, particularly since the model gives a diagnostic output while being unable to explain the results. Physician interpretation is necessary to explain why a diagnosis or treatment should be chosen [50, 51].

As aforementioned, TBSE is not yet widely available to patients [26]. Studies showed that primary care physicians (PCP) as well as dermatologists do not usually perform TBSE as part of their standard practice examination [52, 53]. Even though the importance of TBSE has been highlighted in many studies [27, 28], there are still no consistent recommendations among professionals regarding skin cancer screening to seal the use of this diagnostic practice [54]. Moreover, inadequate time [53] as well as insufficient training in medical students and residents prevents TBSE from being established in clinical practice [55]. The use of AI in TBSE could be a helpful assistant to PCP and other non-dermatology specialists, as it will provide the patients a skin cancer screening that they would not otherwise have, without being time-consuming for physicians. Even the greatest algorithm or physician will fail to diagnose an unexamined or unimaged melanoma, so our efforts should undoubtedly be focusing in this direction.

Lastly, we are still not aware if all these efforts for early detection of skin cancer could be more helpful than harmful. As a vast range of disorders—some of unknown biology—are labeled as cancers, overdiagnosis is a term that we often hear [56, 57]. Within that spectrum, we are not aware of whether the excessive use of technology can mislead far from the desired results. The dilemma has two main concerns regarding the detection of an extremely early stage of melanocytic tumors that are of uncertain malignant potential [58] and the over-detection of NMSC in elderly patients. This can lead to a negative psychological impact on patients with early stages of MM [59] and redundant excisions regarding NMSC in patients with short life expectancy, making the benefit of such excisions doubtful [60].

**STRATEGIES TO OVERCOME LIMITATIONS**

The acknowledgment of the weaknesses of these emerging technologies is the cornerstone for the improvement and evolvement of the existing algorithms. There are possible strategies to overcome these weaknesses before these algorithms become a part of daily clinical practice. Firstly, we need to expand CNN training sets to reflect the variety of the general population. The immigration waves underly this need since physicians are required to examine patient populations with whom they are less familiar and less to assess. A trained algorithm will hence be an advantage. Most algorithms are trained on either Caucasian or Asian patients [9, 38, 40, 61], but early screening of patients
with skin of color could be more beneficial, as more advanced disease and lower survival rates have been reported due to delays in diagnosis in this population [62]. Algorithms tend to underperform when they are given data from different populations that are not included in the dataset. This highlights the necessity to train the same algorithm on a broader range of images from different ethnicities [61].

The inclusion of metadata for the patients under examination should also become an inseparable part of the provided data for the algorithm, such as age, gender, skin type, and anatomic location [31]. This will establish a more realistic environment in studies for clinicians who always include a patient’s history in their diagnostic approach. This metadata should be provided to both clinicians and machines, as the algorithm could also analyze the information. Some studies have already integrated clinical metadata and their results are encouraging for a more accurate classifier [63, 64]. Future studies will determine if this approach gives clinicians a better accuracy compared to CNNs [65].

Clinical close-up images can also be harnessed for the artificial classification of skin lesions [64]. Macroscopic examination of a lesion is the first approach for a clinician to decide whether to proceed to dermoscopy or not. Clinical images of the lesion can provide additional data that are not visible under dermoscopy, such as the pearly appearance and shiny surface of BCC [66] and “stuck-on” appearance of seborrheic keratosis [67]. The combination of clinical and dermoscopic image analysis, as aforementioned, is called combined CNN (cCNN) and will probably become the predominant reference point in future studies. Studies to date have already used cCNN classifiers for improving the algorithm’s performance [18, 63, 64]. Those datasets of clinical images could potentially be used to train algorithms for smartphone applications. However, even the best of these algorithms still have a long way to go, as shown on Figs. 1 and 2.

Special attention should be given to images with confounding factors. Part of these confounding factors, such as lesion-adjacent artifacts could be overcome with the use of image segmentation—a technique to separate the lesion from the background image. A variety of techniques have been proposed for lesion segmentation and thus could be used in future studies [35]. A study showed that the overall performance of lesion classifiers trained in segmented images was comparable to that of unsegmented images. However, researchers pointed out that segmentation quality must be controlled, as this approach might introduce new pitfalls that require further investigation [68].

Haggenmuller et al., in their systematic review on skin cancer classification via CNN, pointed out that the vast majority of the reader studies used holdout data exclusively. Holdout data refers to data from the same source of training and validating data [17]. Navarrete et al. used Han et al.’s [69] publicly available algorithm to explore its generalizability on external testing. Their results demonstrated inferior sensitivity of the algorithm when applied to a different data set [31]. The data used to validate an algorithm that do not follow the training distribution are called out-of-distribution (OOD) data. Future research should consider the use of OOD images (e.g., from a
different source) for the evaluation of classifiers as the gold standard [17].

Finally, the scientific community should favor studies that propel the collaboration of human and artificial intelligence, instead of considering them as opponents in reader studies. In the study of Hekler et al., dermatologists and a single trained CNN classifier independently classified a set of biopsy-verified skin lesions. The researchers combined their results to a new ensemble classifier that demonstrated superior sensitivity to the best individual classifier [70]. Tschandl et al. also found that when physician’s diagnosis-making is supported by an AI algorithm, the diagnostic accuracy improves over that of either AI or physicians alone [15]. These findings have also been confirmed by other researchers [71, 72]. The impact of the previous use of the aforementioned MelaFind system in dermatologists’ decisions to biopsy atypical lesions was evaluated in a study by Hauschild et al. The study showed that dermatologists did not follow MelaFind’s results systematically but rather used the information as complementary in their decision to biopsy, resulting in increased sensitivity [73]. Future research should aim at human–machine collaboration studies, as these approaches will probably suit best clinical practices and we are currently in the process of running such studies.

The use of a standardized protocol for quality use of imaging is essential. Digital Imaging and Communications in Medicine (DICOM) is the international standard for medical imaging. It defines the formats for medical images that can be exchanged with the data and quality necessary for clinical use [74]. Even though the use of DICOM has become the standard method of image processing in other specialties such as radiology and cardiology, in dermatology there is still room for improvement. Images are being collected by non-standardized methods via smartphones and cameras without the capacity to include inclusion supplementary material. Caffery et al. [75] gave a detailed explanation of the role of DICOM in AI in dermatology. They highlighted that objects such as resized or down-sampled images, segmentation images, and the algorithm’s lesion classification output, as well as metadata can be attached to a DICOM file [75]. The existence of such datasets can eliminate the aforementioned pitfalls in AI. These datasets can be used for external validation of ML algorithms (OOD), contributing to generalizability enhancement. Additionally, the use of metadata-based retrieval could facilitate retrieving images from ML datasets with specific characteristics for future studies. Finally, researchers declared that a patient’s identity can be preserved for privacy concerns in clinical trials with DICOM’s de-identification profiles [76].

**FUTURE PERSPECTIVES**

**Dermoscopy and Body Scanning**

We now live in an era where these emerging technologies are integrating into different medical fields. At this point, we need to identify under which conditions AI algorithms would be useful in the clinical setting of dermatology. Human–machine collaboration has revealed promising results for future applications [15, 70, 72]. To further extend this approach, machines could assist clinicians in time-consuming practices that occasionally are not being performed [26]. A study showed that automated mapping of pigmented skin lesions from an automated total body scanning system as well as the detection of change in aligned images can be successfully applied under specific circumstances [77]. Furthermore, naevus count is another time-consuming task that dermatologists need to perform in order to stratify patients’ risk. Betz-Stablein et al. showed that CNN algorithms can successfully perform naevus count from 3D total body photography (TBP) [78].

However, even the most advanced of these systems still display intrinsic biases and need larger datasets to optimize their performance, as shown in Fig. 3.

Additionally, we need to acknowledge the need for life-long surveillance in high-risk patients. Sequential digital dermoscopy (SDD) was shown to be effective for early melanoma detection in high-risk patients [79]. Moreover, short-term follow-up of suspicious pigmented
lesions is commonly used to avoid unnecessary excisions and assure the correct diagnosis [80]. Side-by-side comparison of sequential dermoscopic images is necessary for the detection of substantial changes in dermatoscopic parameters that are not specific to MM and hence guarantee an earlier diagnosis [81]. The detection of dynamic changes in sequential dermoscopic images could be time-consuming and is susceptible to subjectivity [82]. The unique nature of the above patients is highlighted by legal and safety concerns regarding the inadequacy of detecting changes in SDD from less-experienced dermatoscopists, as well as non-compliance of patients for follow-up [83]. Currently, the research in the performance of AI in comparing SSD is in its infancy. The results of a promising study showed that automated algorithms could detect computing image differences among consecutive dermoscopic images and provide earlier detection of MM on the first follow-up compared to clinicians [84]. We believe that this is an essential field that demands the attention of AI applications, as there is room for improvements that will further improve early MM detection.

**Precision Prevention**

Precision medicine is the practice of personalized and individualized medicine for patients. It is the stratification of individual patients based on genetic, biomarker, phenotypic, and psychosocial characteristics that aims to distinguish every patient to provide a targeted and unique treatment for the best clinical outcomes [85]. As ML has the capability of unlimited storage, the above data can be stored and analyzed in ML algorithms to provide a holistic predictive result. Lee et al. gave a remarkable approach to precision prediction of MM [51]. They proposed a holistic risk stratification that can be produced from AI computer-aided diagnostics in order to provide the appropriate level
of surveillance to patients depending on their risk factors. Clinical phenotype and deep image-based phenotype using CNN and 3D total body photography can be assessed for naevus count and photo-numeric scales for sun damage. Genotype, digital, and molecular markers from lesion assessment can all be combined to produce a personalized skin score. This approach aims to conserve early melanoma detection and improve surveillance for high-risk patients, while minimizing overdiagnosis [51, 86].

Teledermatology/Smartphone Apps

Finally, we need to address the fact that a high percentage of patients will not have access to specialized care, which can lead to unfavorable disease outcomes [26]. Many studies have highlighted the usefulness of both store and forward teledermatology and automated smartphone apps [87]. Teledermatology and potential teledermoscopy—with the use of a dermatoscope that is attachable to the smartphone—can provide convenience and accessibility, reduce time travel, assist general practitioners, and improve triage and management [88]. In a recent study, patients were invited to use a phone application to send pictures of their suspicious skin lesions. About 70% of participants stated that they would not have seen a dermatologist without the program, indicating that teledermatology could help promote patient awareness [89]. Incorporating CNN algorithms into smartphone applications will provide an automated and trustful system that will classify the patient’s lesion and provide instructions accordingly. Consequently, teledermatology and teledermoscopy are likely to aid in improving patients’ screening and awareness, and also reduce dermatologists’ workloads and give time for the attention needed for high-risk patients. Rajkomar’s et al. work on the use of ML in medicine might be a guide for the future direction of dermatology [10].

CONCLUSIONS

Machine learning plays a tremendous role in dermatology and skin cancer detection. The opportunities that lie ahead are unlimited, starting from the automated classification of skin cancer through CNN, automated total body photography and sequential digital dermoscopy to AI precision prevention and automated teledermoscopy. However, the application of these algorithms in clinical practice is premature. Limitations concerning lack of generalizability and standardization, consumer trust, and potential overdiagnosis must continue to be addressed in order to bring these new technologies safely into the real world. Datasets with diverse images and different populations, the inclusion of metadata and close-up images, segmentation tools and use of OOD and DICOM standards will help eliminate the current limitations for future studies. Finally, a clinician’s role, especially in oncological patients remains fundamental, as no machine can ever replace a human-to-human relationship. Still, we have in our hands a tool that if we use rationally under the correct guidance and supervision, we can enhance medical care. A new era is spreading ahead us, whether we embrace it or not. We have a duty to our patients to assess the opportunities and create a better future for medicine and patients.

ACKNOWLEDGEMENTS

Funding. No funding or sponsorship was received for this study or publication of this article.

Authorship. All named authors meet the International Committee of Medical Journal Editors (ICMJE) criteria for authorship for this article, take responsibility for the integrity of the work as a whole, and have given their approval for this version to be published.

Author Contributions. All authors contributed to the commentary conception and design. Material preparation, data collection, and analysis were performed by Konstantinos Liopyris, Stamatios Gregoriou, Julia Dias, and Alexandros J. Stratigos. The first draft of the manuscript was written by Julia Dias and edited...
Disclosures. Konstantinos Liopyris, Statamos Gregoriou, Julia Dias, and Alexandros J. Stratigos have nothing to disclose.

Compliance with Ethics Guidelines. This paper reflects the authors’ personal perspectives on the current and future trajectories of AI in dermatology. The authors acknowledge that ethical issues might arise in future applications of AI in dermatology and these will pose an ethical challenge for society; however, the paper does not focus on these issues. This article is based on previously conducted studies and does not contain any new studies with human participants or animals performed by any of the authors.

Open Access. This article is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License, which permits any non-commercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc/4.0/.

REFERENCES

1. Artificial Intelligence. Oxford Dictionary. https://doi.org/10.1093/oi/authority.20110803095426960.
2. Vermeulen O, Eisenhauer M, Sundmaeker H, et al. Internet of things cognitive transformation technology research trends and applications.
3. Bini SA. Artificial intelligence, machine learning, deep learning, and cognitive computing: what do these terms mean and how will they impact health care? J Arthroplast. 2018;33(8):2358–61. https://doi.org/10.1016/j.arth.2018.02.067.
4. Erickson BJ, Korfiatis P, Akkus Z, Kline TL. Machine learning for medical imaging. Radiographics. 2017;37(2):505–15. https://doi.org/10.1148/rg.2017160130.
5. Hogarty DT, Mackey DA, Hewitt AW. Current state and future prospects of artificial intelligence in ophthalmology: a review. Clin Exp Ophthalmol. 2019;47(1):128–39. https://doi.org/10.1111/ceo.13381.
6. Hogarty DT, Su JC, Phan K, et al. Artificial intelligence in dermatology—where we are and the way to the future: a review. Am J Clin Dermatol. 2020;21(1):41–7. https://doi.org/10.1007/s40257-019-00462-6.
7. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. http://code.google.com/p/cuda-convnet/.
8. Lee JG, Jun S, Cho YW, et al. Deep learning in medical imaging: General overview. Korean J Radiol. 2017;18(4):570–84. https://doi.org/10.3348/kjr.2017.18.4.570.
9. Tschandl P. Artificial intelligence for melanoma diagnosis. Ital J Dermatol Venereol. 2021;156(3):289–99. https://doi.org/10.23736/S2784-105X.2021.1563003.
10. Rajkomar A, Dean J, Kohane I. Machine learning in medicine. N Engl J Med. 2019;380(14):1347–58. https://doi.org/10.1056/nejmra1814259.
11. Benjamens S, Dhunnoo P, Meskó B. The state of artificial intelligence-based FDA-approved medical devices and algorithms: an online database. NPJ Digit Med. 2020. https://doi.org/10.1038/s41746-020-00324-0.
12. Abraham SB, Arunachalam S, Zhong A, Agrawal P, Cohen O, McMahon CM. Improved real-world glycemic control with continuous glucose monitoring system predictive alerts. J Diabetes Sci Technol. 2019;15(1):91–7. https://doi.org/10.1177/1932296819859334.
13. van der Heijden AA, Abramoff MD, Verbraak F, van Hecke MV, Liem M, Nijpels G. Validation of automated screening for referable diabetic retinopathy with the IDx-DR device in the Hoorn Diabetes Care System. Acta Ophthalmol. 2018;96(1):63–8. https://doi.org/10.1111/aos.13613.
14. Escobar GJ, Turk BJ, Ragins A, et al. Piloting electronic medical record-based early detection of
inpatient deterioration in community hospitals. J Hosp Med. 2016;11:S18–24. https://doi.org/10.1002/jhm.2652.

15. Tschandl P, Rinner C, Apalla Z, et al. Human–computer collaboration for skin cancer recognition. Nat Med. 2020;26(8):1229–34. https://doi.org/10.1038/s41591-020-0942-0.

16. Esteva A, Kuprel B, Novoa RA, et al. Dermatologist-level classification of skin cancer with deep neural networks. Nature. 2017;542(7639):115–8. https://doi.org/10.1038/nature21056.

17. Haggenmüller S, Maron RC, Hecker A, et al. Skin cancer classification via convolutional neural networks: systematic review of studies involving human experts. Eur J Cancer. 2021;156:202–16. https://doi.org/10.1016/j.ejca.2021.06.049.

18. Tschandl P, Rosendahl C, Akay BN, et al. Expert-level diagnosis of nonpigmented skin cancer by combined convolutional neural networks. JAMA Dermatol. 2019;155(1):58–65. https://doi.org/10.1001/jamadermatol.2018.4378.

19. Combalia M, Codella N, Rotemberg V, et al. Validation of artificial intelligence prediction models for skin cancer diagnosis using dermoscopy images: the 2019 International Skin Imaging Collaboration Grand Challenge. Lancet Digit Health. 2022;4(5):e330–9. https://doi.org/10.1016/S2589-7500(22)00021-8.

20. National Cancer Institute Melanoma of the Skin-Cancer Stat Facts. Published May 2022. Accessed 18 May 2022. https://seer.cancer.gov/statfacts/html/melan.html

21. Swetter SM, Tsao H, Bichakjian CK, et al. Guidelines of care for the management of primary cutaneous melanoma. J Am Acad Dermatol. 2019;80(1):208–50. https://doi.org/10.1016/j.jaad.2018.08.055.

22. Pan Y, Gareau DS, Scope A, Rajadhyaksha M, Mullan NA, Marghoob AA. Polarized and nonpolarized dermoscopy the explanation for the observed differences. http://archderm.jamanetwork.com/

23. Pehamberger H, Steiner A, Wolff K. In vivo epiluminescence microscopy of pigmented skin lesions. I. Pattern analysis of pigmented skin lesions. J Am Acad Dermatol. 1987;17(4):571–83. https://doi.org/10.1016/S0190-9622(87)70239-4.

24. Kittler H, Pehamberger H, Wolff K, Binder M. Diagnostic accuracy of dermoscopy. Lancet Oncol. 2002;3(3):159–65. https://doi.org/10.1016/S1470-2045(02)00679-4.

25. Argenziano G, Cerroni L, Zalaudek I, et al. Accuracy in melanoma detection: a 10-year multicenter survey. J Am Acad Dermatol. 2012. https://doi.org/10.1016/j.jaad.2011.07.019.

26. Lakhani NA, Saraiya M, Thompson TD, King SC, Guy GP. Total body skin examination for skin cancer screening among U.S. adults from 2000 to 2010. Prev Med (Baltim). 2014;61:75–80. https://doi.org/10.1016/j.ypmed.2014.01.003.

27. Jiang A, Jefferson IS, Robinson SK, et al. Skin cancer discovery during total body skin examinations. Int J Womens Dermatol. 2021;7(4):411–4. https://doi.org/10.1016/j.ijwdd.2021.05.005.

28. Argenziano G, Zalaudek I, Hofmann-Wellenhof R, et al. Total body skin examination for skin cancer screening in patients with focused symptoms. J Am Acad Dermatol. 2012;66(2):212–9. https://doi.org/10.1016/j.jaad.2010.12.039.

29. Avilés-Izquierdo JA, Molina-López I, Rodríguez-Lomba E, Marquez-Rodas I, Suarez-Fernandez R, Lazaro-Ochaita P. Who detects melanoma? Impact of detection patterns on characteristics and prognosis of patients with melanoma. J Am Acad Dermatol. 2016;75(5):967–74. https://doi.org/10.1016/j.jaad.2016.07.009.

30. Scheetz J, Rothschild P, McGuinness M, et al. A survey of clinicians on the use of artificial intelligence in ophthalmology, dermatology, radiology and radiation oncology. Sci Rep. 2021. https://doi.org/10.1038/s41598-021-84698-5.

31. Navarrete-Dechent C, Dusza SW, Liopyris K, Marghoob AA, Halpern AC, Marchetti MA. Automated dermatological diagnosis: hype or reality? J Investig Dermatol. 2018;138(10):2277–9. https://doi.org/10.1016/j.jid.2018.04.040.

32. International Skin Imaging Collaboration (ISIC) Archive. www.isic-archive.com

33. Monheit G, Cognetta AB, Ferris L, et al. The performance of MelaFind: a prospective multicenter study. Arch Dermatol. 2011;147(2):188–94. https://doi.org/10.1001/archdermatol.2010.302.

34. Strata Skin Sciences, Inc. and subsidiary notes to unaudited condensed consolidated financial statements. 2017.

35. Okur E, Turkan M. A survey on automated melanoma detection. Eng Appl Artif Intell. 2018;73:50–67. https://doi.org/10.1016/j.engappai.2018.04.028.

36. Narla A, Kuprel B, Sarin K, Novoa R, Ko J. Automated classification of skin lesions: from pixels to...
practice. J Investig Dermatol. 2018;138(10):2108–10. https://doi.org/10.1016/j.jid.2018.06.175.

37. 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;155(10):1135–41. https://doi.org/10.1001/jamadermatol.2019.1735.

38. Navarrete-Dechent C, Liopyris K, Marchetti MA. Multiclass artificial intelligence in dermatology: progress but still room for improvement. J Invest Dermatol. 2021;141(5):1325–8. https://doi.org/10.1016/j.jid.2020.06.040.

39. Haenssle HA, Winkler JK, Fink C, et al. Skin lesions of face and scalp—classification by a market-approved convolutional neural network in comparison with 64 dermatologists. Eur J Cancer. 2021;144:192–9. https://doi.org/10.1016/j.ejca.2020.11.034.

40. Yu C, Yang S, Kim W, et al. Acral melanoma detection using a convolutional neural network for dermoscopy images. PLoS One. 2018. https://doi.org/10.1371/journal.pone.0193321.

41. Abbas Q, Ramzan F, Ghani MU. Acral melanoma detection using dermoscopic images and convolutional neural networks. Vis Comput Ind Biomed Art. 2021. https://doi.org/10.1016/j.jid.2020.06.040.

42. Tschandl P, Codella N, Akay BN, et al. Comparison of the accuracy of human readers versus machine-learning algorithms for pigmented skin lesion classification: an open, web-based, international, diagnostic study. Lancet Oncol. 2019;20(7):938–47. https://doi.org/10.1016/S1470-2045(19)30333-X.

43. Adamson AS, Smith A. Machine learning and health care disparities in dermatology. JAMA Dermatol. 2018;154(11):1247–8. https://doi.org/10.1001/jamadermatol.2018.2348.

44. Charttrand G, Cheng PM, Vorontsov E, et al. Deep learning: a primer for radiologists. Radiographics. 2017;37(7):2113–31. https://doi.org/10.1148/rg.2017170077.

45. Gaudy-Marqueste C, Wazaefi Y, Bruneu Y, et al. Ugly duckling sign as a major factor of efficiency in melanoma detection. JAMA Dermatol. 2017;153(4):279–84. https://doi.org/10.1001/jamadermatol.2016.5500.

46. Mascaro JM. The dermatologist’s position concerning nevi: a vision ranging from “the ugly duckling” to “little red riding hood.” Arch Dermatol. 1998;134(11):1484–5. https://doi.org/10.1001/archderm.134.11.1484.

47. Zalaudek I, Docimo G, Argenziano G. Using dermoscopic criteria and patient-related factors for the management of pigmented melanocytic nevi. Arch Dermatol. 2009;145(7):816–26. https://doi.org/10.1001/archdermatol.2009.115.

48. Carli P, de Giorgi V, Argenziano G, Palli D, Gianotti B. Pre-operative diagnosis of pigmented skin lesions: in vivo dermoscopy performs better than dermoscopy on photographic images. 2002.

49. Gillespie N, Lockey S, Curtis C. Trust in artificial intelligence: a five country study. 2021. https://doi.org/10.14264/e34bfa3.

50. Chan S, Reddy V, Myers B, Thibodeaux Q, Brownstone N, Liao W. Machine learning in dermatology: current applications, opportunities, and limitations. Dermatol Ther (Heidelb). 2020;10:365–86. https://doi.org/10.6084/m9.figshare.12006789.

51. Lee KJ, Betz-Stablein B, Stark MS, et al. The future of precision prevention for advanced melanoma. Front Med (Lausanne). 2022. https://doi.org/10.3389/fmed.2021.818096.

52. LeBlanc WG, Vidal L, Kirshner RS, et al. Reported skin cancer screening of US adult workers. J Am Acad Dermatol. 2008;59(1):55–63. https://doi.org/10.1016/j.jaad.2008.03.013.

53. Federman DG, Kravetz JD, Kirshner RS. Skin cancer screening by dermatologists: prevalence and barriers. J Am Acad Dermatol. 2002;46(5):710–4. https://doi.org/10.1067/mjd.2002.120531.

54. Johnson MM, Leachman SA, Aspinwall LG, et al. Skin cancer screening: recommendations for data-driven screening guidelines and a review of the US Preventive Services Task Force controversy. Melanoma Manag. 2017;4(1):13–37. https://doi.org/10.2217/mmt-2016-0022.

55. Wise E, Singh D, Moore M, et al. Rates of skin cancer screening and prevention counseling by US medical residents. Arch Dermatol. 2009. https://doi.org/10.1001/archdermatol.2009.242.

56. Welch HG, Black WC. Overdiagnosis in cancer. J Nati Cancer Inst. 2010;102(9):605–13. https://doi.org/10.1093/jnci/djq099.

57. Esserman LJ, Thompson IM, Reid B, et al. Addressing overdiagnosis and overtreatment in cancer: a prescription for change. Lancet Oncol. 2014. https://doi.org/10.1016/S1470-2045(13)70598-9.

58. Kutzner H, Jutzi TB, Krahl D, et al. Overdiagnosis of melanoma—causes, consequences and solutions.
59. Bell KJL, Mehta Y, Turner RM, et al. Fear of new or recurrent melanoma after treatment for localised melanoma. Psychooncology. 2017;26(11):1784–91. https://doi.org/10.1002/pon.4366.

60. Linos E, Parvataneni R, Stuart SE, Boscardin WJ, Landefeld CS, Chren MM. Treatment of nonfatal conditions at the end of life: nonmelanoma skin cancer. JAMA Intern Med. 2013;173(11):1006–12. https://doi.org/10.1001/jama.2013.639.

61. 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 Investig Dermatol. 2018;138(7):1529–38. https://doi.org/10.1016/j.jid.2018.01.028.

62. Ward-Peterson M, Acuña JM, Alkhallalh MK, et al. Association between race/ethnicity and survival of melanoma patients in the United States over 3 decades. Medicine (United States). 2016. https://doi.org/10.1097/MD.0000000000003315.

63. Kharazmi P, Kalia S, Lui H, Wang ZJ, Lee TK. A feature fusion system for basal cell carcinoma detection through data-driven feature learning and patient profile. Skin Res Technol. 2018;24(2):256–64. https://doi.org/10.1111/srt.12422.

64. Yap J, Yolland W, Tschandl P. Multimodal skin lesion classification using deep learning. Exp Dermatol. 2018;27(11):1261–7. https://doi.org/10.1111/exd.13777.

65. Haenssle HA, Fink C, Toberer F, et al. Man against machine reloaded: performance of a market-approved convolutional neural network in classifying a broad spectrum of skin lesions in comparison with 96 dermatologists working under less artificial conditions. Ann Oncol. 2020;31(1):137–43. https://doi.org/10.1016/j.annonc.2019.10.013.

66. McDaniel B, Badri T, Steele RB. Basal cell carcinoma. 2022.

67. Greco MJ, Bhutta BS. Seborrheic keratosis. 2022.

68. Maron RC, Hekler A, Krieghoff-Henning E, et al. Reducing the impact of confounding factors on skin cancer classification via image segmentation: technical model study. J Med Internet Res. 2021. https://doi.org/10.2196/21695.

69. 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(8):1836–42. https://doi.org/10.1093/annonc/mdy166.

70. Hekler A, Utikal JS, Enk AH, et al. Superior skin cancer classification by the combination of human and artificial intelligence. Eur J Cancer. 2019;120:114–21. https://doi.org/10.1016/j.ejca.2019.07.019.

71. Marchetti MA, Codella NCF, Dusza SW, et al. Results of the 2016 international skin imaging collaboration international symposium on biomedical imaging challenge: comparison of the accuracy of computer algorithms to dermatologists for the diagnosis of melanoma from dermoscopic images. J Am Acad Dermatol. 2018;78:270-277.e1. https://doi.org/10.1016/j.jaad.2017.08.016.

72. Han SS, Park I, Eun Chang S, et al. Augmented intelligence dermatology: deep neural networks empower medical professionals in diagnosing skin cancer and predicting treatment options for 134 skin disorders. J Investig Dermatol. 2020;140(9):1753–61. https://doi.org/10.1016/j.jid.2020.01.019.

73. Hauschild A, Chen SC, Weichenthal M, et al. To excise or not: Impact of MelaFind on German dermatologists’ decisions to biopsy atypical lesions. JDDG J German Soc Dermatol. 2014;12(7):606–14. https://doi.org/10.1111/ddg.12362.

74. Digital Imaging and Communications in Medicine. https://www.dicomstandard.org/about. Accessed 19 May 2022.

75. Caffery LJ, Rotemberg V, Weber J, Soyer HP, Malvehy J, Clunie D. The role of DICOM in artificial intelligence for skin disease. Front Med (Lausanne). 2021. https://doi.org/10.3389/fmed.2020.619787.

76. DICOM PS3.15 2022b—Security and System Management Profiles. https://dicom.nema.org/medical/dicom/current/output/html/part15.html#chapter_E. Accessed 20 May 2022.

77. Korotkov K, Quintana J, Puig S, Malvehy J, Garcia R. A new total body scanning system for automatic change detection in multiple pigmented skin lesions. IEEE Trans Med Imaging. 2015;34(1):317–38. https://doi.org/10.1109/TMI.2014.2357715.

78. Betz- Stablein B, D’Alessandro B, Koh U, et al. Reproducible naevus counts using 3D total body photography and convolutional neural networks. Dermatology. 2022;238(1):4–11. https://doi.org/10.1159/000517218.

79. Salerni G, Terán T, Puig S, et al. Meta-analysis of digital dermoscopy follow-up of melanocytic skin lesions: a study on behalf of the International Dermoscopy Society. J Eur Acad Dermatol Venereol. 2013;27(7):805–14. https://doi.org/10.1111/jdv.12032.
80. Menzies SW, Emery J, Staples M, et al. Impact of dermoscopy and short-term sequential digital dermoscopy imaging for the management of pigmented lesions in primary care: a sequential intervention trial. Br J Dermatol. 2009;161(6):1270–7. https://doi.org/10.1111/j.1365-2133.2009.09374.x.

81. Babino G, Lallas A, Agozzino M, et al. Melanoma diagnosed on digital dermoscopy monitoring: a side-by-side image comparison is needed to improve early detection. J Am Acad Dermatol. 2021;85(3):619–25. https://doi.org/10.1016/j.jaad.2020.07.013.

82. Tschandl P, Doz Philipp Tschandl P. Sequential digital dermatoscopic imaging of patients with multiple atypical nevi. Rev Dermatol Pract Concept. 2018;8(3):231–7. https://doi.org/10.5826/dpc.0803a16.

83. Winkler JK, Tschandl P, Toberer F, et al. Monitoring patients at risk for melanoma: may convolutional neural networks replace the strategy of sequential digital dermoscopy? Eur J Cancer. 2022;169:180–8. https://doi.org/10.1016/j.ejca.2021.10.030.

84. Yu Z, Nguyen J, Nguyen TD, et al. Early melanoma diagnosis with sequential dermoscopic images. IEEE Trans Med Imaging. 2022;41(3):633–46. https://doi.org/10.1109/TMI.2021.3120091.

85. Jameson JL, Longo DL. Precision medicine-personalized, problematic, and promising. 2015.

86. Smithers BM, Dunn J, Soyer HP. Whither melanoma in Australia? Med J Aust. 2017;207(8):330–1. https://doi.org/10.5694/mja17.00740.

87. Rat C, Hild S, Sérandour JR, et al. Use of smartphones for early detection of melanoma: systematic review. J Med Internet Res. 2018. https://doi.org/10.2196/jmir.9392.

88. Koh U, Horsham C, Soyer HP, et al. Consumer acceptance and expectations of a mobile health application to photograph skin lesions for early detection of melanoma. Dermatology. 2018;235(1):4–10. https://doi.org/10.1159/000493728.

89. Cazzaniga S, Castelli E, Di Landro A, et al. Mobile teledermatology for melanoma detection: assessment of the validity in the framework of a population-based skin cancer awareness campaign in northern Italy. J Am Acad Dermatol. 2019;81(1):255–7. https://doi.org/10.1016/j.jaad.2019.01.082.