



Artificial Intelligence in Skin Cancer

Ofer Reiter^{1,2} · Veronica Rotemberg¹ · Kivanc Kose¹ · Allan C. Halpern¹

Published online: 6 July 2019

© Springer Science+Business Media, LLC, part of Springer Nature 2019

Abstract

Purpose To review recent developments in artificial intelligence for skin cancer diagnosis.

Recent Findings Major breakthroughs in recent years are likely related to advancements in utilization of convolutional neural networks (CNNs) for dermatologic image analysis, especially dermoscopy. Recent studies have shown that CNN-based approaches perform as well as or even better than human raters in diagnosing close-up and dermoscopic images of skin lesions in a simulated static environment. Several limitations for the development of AI include the need for large data pipelines and ground truth diagnoses, lack of metadata, and lack of rigorous widely accepted standards.

Summary Despite recent breakthroughs, adoption of AI in clinical settings for dermatology is in early stages. Close collaboration between researchers and clinicians may provide the opportunity to investigate implementation of AI in clinical settings to provide real benefit for both clinicians and patients.

Keywords Machine learning · Artificial intelligence · Convolutional neural networks · Skin cancer · Melanoma · Non-melanoma skin cancer

Introduction

Skin Cancer Diagnosis

With over 5 million new cases of cutaneous malignancies reported every year [1], automated methods for diagnosing skin cancer are a huge area of clinical need and research effort.

In the past, the diagnosis of skin cancer relied on a clinical examination by a dermatologist. In recent years, dermoscopy has gained in popularity and is currently being used by 81% of US dermatologists. The rate is even higher among young dermatologists at 98% [2]. The added value of dermoscopy over the naked eye examination was demonstrated in several meta-analyses. For the diagnosis of melanoma, dermoscopy increased the diagnostic accuracy of naked eye examination with a relative diagnostic odds ratio of 4.7–5.6 [3]. For basal cell

carcinoma (BCC), sensitivity increased from 67 to 85% and specificity from 97 to 98% [4]. However, dermoscopy is an operator-dependent test and requires training and experience.

Newer technologies, such as reflectance confocal microscopy (RCM) and optical coherence tomography (OCT), are used as add-on tests to supplement dermoscopy in certain specialized centers and may further increase diagnostic accuracy [5, 6]. However, these tests are still not widely implemented, mainly due to the high costs and need for specific training.

Artificial Intelligence

Artificial intelligence (AI) is the notion of developing intelligent machines that can automatically carry on a task. This idea dates back to the 1930s. The first landmark article on machines that “think” was published in 1950 by Allan Turing, who also suggested that a machine that passes the “Turing test”, meaning it could hold a conversation in a way that is indistinguishable from a human, can be considered “thinking” [7].

Machine learning (ML) is a subfield of AI that studies how computers can learn tasks without being explicitly programmed to conduct them [8]. ML is mathematically modeling the relation between the data (e.g., dermoscopy images) and the task (e.g., diagnosis) while optimizing a given penalty

This article is part of the Topical Collection on *Skin Cancer*

✉ Ofer Reiter
Ofer.rtr@gmail.com

¹ Dermatology Service, Department of Medicine, Memorial Sloan Kettering Cancer Center, New York, NY 10022, USA

² Sackler Faculty of Medicine, Tel Aviv University, 6997801 Tel Aviv, Israel

(e.g., accuracy of diagnosis). Among several methods available in the literature, deep neural networks have gained popularity in recent years due to their high representation and classification power. Neural networks as a ML technique was first proposed in the 1940s, but it wasn't until the early 2010s when the idea began to be implemented regularly among the machine learning community, following the seminal work of Krizhevsky [9]. The main reasons were (i) neural network models with high recognition capabilities are hard to train due to their computational complexity and (ii) they do not perform well (overfit to the training examples by memorizing them) in the lack of a large set of training data [10]. Advances in computing hardware (e.g., graphical processing units (GPU), tensor processing units (TPU)), parallel computing techniques, and availability of vast amount of digital data enabled investigators to train neural networks that can model complex relations between given training data and tasks. In particular, diagnostic analysis of medical images provides an incredible opportunity for machine learning to impact clinical care. In recent years, there has been an incredible increase in research directed at automated analysis of clinical and dermoscopic images for the purpose of diagnosing skin cancers, particularly melanoma.

History of AI and Skin Cancer

Dermoscopy allows for imaging skin surface and subsurface morphological structures of lesions. It demonstrates a controlled environment for imaging of skin lesions with predefined features of lighting conditions, camera distance, angle, etc. Having such a controlled imaging environment, it provides reasonably “clean” datasets for AI development purposes. The consistency of dermoscopy combined with its common use in dermatology clinics has made it the optimal training ground for AI investigation of dermatology problems.

At the initial stages of ML, research for skin lesion diagnoses mostly focused on the classic workflow of machine learning: preprocessing, color- or texture-based segmentation (marking the borders of the lesion), feature extraction, and classification [11, 12]. In this setting, ML algorithms were trained to diagnose skin lesions using features that are modeling the human-generated criteria (such as the 7-point checklist).

The first report of the ML application in the diagnosis of dermoscopic images of nevi and melanoma was by Binder et al. in 1994. The authors used artificial neural networks and reported similar diagnostic accuracy compared with human investigators [13]. However, over the next two decades, progress in this field was slow. This was likely due to a lack of systematically collected large image datasets that represent the general cohort as well as limited computational and algorithmic capabilities to digest and analyze the available data.

Recent Breakthroughs in the Field of AI and Skin Cancer

In the past few years, the developments in the capability of training neural network models with larger sizes and higher representation power allowed for significant progress. Particularly, a specialized type of neural networks called convolutional neural networks (CNNs) has shown great success in image analysis and recognition applications, including dermoscopic image analysis. CNNs are a stack of filters operating on local regions of interest (e.g., filters calculate relationship among a group of pixels within a certain neighborhood) at a time. They are specialized in learning the relationship between the input (e.g., image) and the classification task (e.g., diagnosis) by analyzing the local relations between a certain neighborhood of samples (e.g., pixels). As mentioned earlier, feature extraction is an integral part of the machine learning process and CNNs no longer require a separate user-defined feature extraction phase. CNNs work through several layers, where the first layer is the input (raw pixel data) and the last layer is the output, which includes the classification/diagnosis of the image/lesion. The first and the last layers are the only two layers that are accessible by the user. However, there are multiple hidden layers between them that can be indirectly accessible through the input and output layers. These hidden layers are the part of the network that models the relation between the input and the output tasks. The network does this by mapping the input lesion image into a distinctive and potentially unique mathematical/numerical description of its interpixel relations called feature representation, which can then be reliably classified using the output layer of the network.

There are two major advantages of CNNs: Unlike previous image analysis pipelines, where the user needs to define the best feature representation to perform the aimed task, CNN-based algorithms are capable of learning the feature representation and solve the classification/recognition task directly from the data in a joint manner without the need for user guidance. Moreover, as the analysis is conducted with convolution operations, which can be computed in a very efficient way by parallel processing over graphical processing units (GPUs), large amounts of data can be easily utilized to achieve successful and highly robust models. CNNs have been shown to perform much better in image analysis compared with previous technologies [14].

While CNNs have dramatically improved the landscape of AI research for skin cancer diagnosis, they suffer from large data requirements and lack of interpretability. As these models have a large number of tunable parameters, they require large image datasets to train [15]. Moreover, in many cases, the analysis that is performed over the hidden layers, as well as the features that are used to generate the diagnosis, cannot be

fully interpreted by the users. Therefore, it can be challenging to understand how the algorithms reach their “conclusions”.

AI Diagnostic Accuracy for Skin Lesions

In 2017, a landmark research letter was published in *Nature* by Esteva et al. [16•]. The authors trained a CNN with 129,450 clinical images including 3000 dermoscopic images and compared its ability to differentiate between keratinocyte carcinomas and seborrheic keratosis and between melanomas and nevi to that of human experts. The CNN achieved performance on par with all tested experts. As opposed to previous work, in this study, the CNN was not restricted to man-made segmentation criteria but rather was only images with respective diagnosis and it created its own diagnostic rules (classification model).

Another landmark publication was that of Han et al. in 2018 [17•]. In this article, the authors trained a CNN with several private image datasets that included 12 different skin diseases (BCC, squamous cell carcinoma [SCC], Bowen’s disease, actinic keratosis, seborrheic keratosis, melanoma, nevus, lentigo, pyogenic granuloma, hemangioma, dermatofibroma, and warts). They reported that the CNN’s performance was similar to that of 16 dermatologists and had an area under the curve (AUC) of receiver operator characteristic (ROC) curve of 0.90–0.96 for BCC, 0.83–0.91 for SCC, and 0.82–0.88 for melanoma. As a novel method for testing, the group has allowed anyone to test their algorithms and has made results public, which can transform the field forward.

After several studies have shown that CNNs can be used to diagnose pigmented skin lesions, Tschandl et al. tested their performance on non-pigmented skin lesions [18]. They trained their model on almost 13,724 images of excised lesions. They then tested its performance on 2000 images and compared it with the performance of 95 human raters. The AUC ROC curve of the CNN was 0.742 compared with 0.695 for the human raters. When specificity was fixed to the mean level of human raters (51.3%), the CNN’s sensitivity (80.5%) was higher than that of the human raters (77.6%). The authors concluded that the CNN achieved a higher rate of correct specific diagnoses compared with the novice raters, but not compared with dermoscopy experts.

Several systematic reviews and meta-analyses have been published in the past 2 years summarizing the available data on the diagnostic accuracy of AI for skin lesions. A Cochrane review summarizing the data available up until August 2016 included a meta-analysis of 22 studies that used dermoscopy-based AI. They found that the sensitivity for the diagnosis of melanoma was 90.1% and specificity was 74.3%. The authors commented that the studies had high variability and high risk of bias and included only specific populations [19].

Marka et al. published a systematic review about AI and non-melanoma skin cancers (NMSC) that included studies

published up until 2018. They reported a diagnostic accuracy of 72–100% and AUC of 0.832 to 1.0. Again, the studies included had a high risk of bias and methodological limitations [20].

In two recent studies from Germany, the ability of CNNs to correctly diagnose skin lesions was further demonstrated. In the first study, CNN was trained and tested on dermoscopic images of pigmented nevi and melanomas and its performance was compared with 58 international dermatologists. Adjusted to the dermatologists’ sensitivity, the CNN had a higher specificity. Interestingly, this was true even when the investigators provided the dermatologists with additional non-dermoscopic close-up images and clinical data [21].

In the second study, the authors trained a CNN solely on dermoscopic images and then tested its performance on clinical images and compared it with 145 dermatologists. Adjusted to the mean sensitivity of the dermatologists (89.4%), the CNN showed a slightly higher specificity (68.2% vs. 64.4%), even though it was never trained on clinical images [22].

International Skin Imaging Collaboration

Sponsored by the International Society for Digital Imaging of the Skin, International Skin Imaging Collaboration (ISIC) is an academia and industry collaboration aimed at improving melanoma diagnoses and reducing its mortality rates through the use of digital imaging technologies. It provides a public database that is used to benchmark machine learning algorithms and host public challenges.

In general, the activity of ISIC can be divided to two major parts:

1. ISIC working groups—These groups of experts are working on developing standards for skin imaging in different aspects, including imaging technologies, imaging techniques, terminology, and metadata standards (standards for the technical and clinical data that should be stored with the image).
2. ISIC archive [23]—ISIC has developed and currently maintains the largest publicly available image database of skin lesions. The images are collected from leading centers around the world. Currently, the archive consists of more than 40,000 images (mostly dermoscopic but also clinical) of ~30 different skin lesions tagged with their diagnosis.

ISIC has been a major driver in the development of AI technologies in the field of skin cancer. First, the ISIC archive is an open access website and the images are available for everyone to download and use for training AI software. Second, ISIC has been hosting annual challenges to further engage the tech community. The challenges consist of a

training set and a test set, and participants are invited to submit their algorithms and to compete for the most accurate algorithm. The challenges include three steps: (1) segmentation of the lesion from the background of the images; (2) detection of different dermoscopic features; and (3) classification of the lesion in the image.

Each participant is provided with a training set (images + respective classification and segmentation information) to develop ML algorithms that carry on the tasks defined in each phase. The performance of the developed methods is assessed over an independent test set (only images are available to the participants). The participant can upload their processing results over the test set to the challenge web portal, at which the results are evaluated in almost real-time and published on the challenge score boards. In this way, the participants can compete the methods and assess their performance against other participants.

Each year, the challenges have become more complicated with more images and more diagnoses that are included. The 2016 challenge included 900 images in the training set and 350 images in the test set. Two diagnoses were included: melanomas and nevi. Two dermoscopic features were examined—streaks and globules. The performance of AI algorithms was compared with that of 8 dermatologists. The dermatologists' performance was similar to the top individual algorithm (sensitivity of 82% and specificity of 59%), but not as good as a fusion algorithm that combined 16 individual-automated predictions (specificity of 76% when sensitivity was set to the dermatologists' level at 82%) [24••].

The 2017 challenge included 2000 images in the training set and 600 images in the test set. Three diagnoses were included: melanomas, nevi, and seborrheic keratosis. Four dermoscopic features were examined: pigment network, negative network, streaks, and milia-like cysts. The top algorithm reached an AUC of 0.91 across all disease categories [25].

The most recent ISIC challenge of 2018 used the HAM10000 dataset [26], which included more than 10,000 dermoscopic images of 7 disease categories (melanoma, nevi, seborrheic keratosis, BCC, Bowen's disease and actinic keratosis, vascular lesions, and dermatofibromas). Five dermoscopic features were examined: pigment network, negative network, streaks, milia-like cysts, and globules (including dots). The best performing algorithms achieved an average sensitivity of 88.5% in diagnosing all disease categories. The algorithms' performance was compared with the performance of over 500 human participants, but these results have not been published yet.

ISIC challenges have helped dramatically to increase the amount of research and publications related to AI and skin lesion diagnosis. Dozens of papers describing the different algorithms that were used in the challenges were published following each challenge.

New Developments and Technologies

Most major breakthroughs in the implementation of AI in skin cancer diagnosis so far have involved creating predictions for the classification of dermoscopic images of skin lesions. However, there are other new and exciting fields of research that have been published in recent years.

1. Use of metadata—As mentioned before, metadata is text-based data that provides additional non-visual information to the rater/AI system. Classically, metadata consists of two components: (a) clinical metadata that includes patient demographics, medical history, and lesion evolution; and (b) technical metadata that includes information about the image acquisition, and technology. Yap et al. examined the use of a classifier that combines imaging modalities with patient metadata and compared it with the performance of a baseline classifier that only used a single macroscopic image. They found that the combined classifier performed better than the baseline classifier in detecting melanoma as well as other lesions such as BCC and SCC [27]. Roffman et al. trained a CNN to predict the risk of NMSC. They were able to reach a sensitivity of 88.5% and a specificity of 62.2% solely based on a questionnaire, which did not even include ultraviolet exposure [28].
2. AI in smartphone apps—A quick search on the different app stores leads to multiple smartphone apps that offer an “automated diagnosis” of skin lesions. However, a recent systematic review by the Cochrane library found very sparse evidence-based information for the efficacy of these technologies. They were able to identify only two studies, both with high risk of bias, that tested four automated diagnosis apps. Sensitivity for the diagnosis of melanoma or “high risk”/“problematic” lesions ranged from 7 to 73% and specificity from 37 to 94%. The authors concluded that smartphone apps have not yet demonstrated sufficient evidence for accuracy and the data that exists suggests they are at high risk of missing melanoma [29].
3. Diagnostic tests other than dermoscopy—Dermoscopy is the most prevalent method used by dermatologists for skin cancer screening. However, other methods exist. Several studies investigated the implementation of AI technologies in these methods. Examples include CNN-based classification on OCT images of BCC (95.4% sensitivity and specificity) [30], semantic segmentation of morphological patterns of melanocytic lesion in RCM mosaics collected at a dermal epidermal junction level (76% sensitivity and 94% specificity) [31], delineation of stratum corneum and dermal epidermal junction in RCM image stacks [32–34], and CNN-based classification on hyperspectral imaging of selected nevi and melanomas (100% sensitivity and 36% specificity) [35].

Limitations and Challenges

After reading this review and the diagnostic accuracy data, one might think that AI currently outperforms human dermatologists in the diagnosis of skin cancer. However, this is not the case. Most of the studies described in this review have been performed in a controlled experimental environment with a limited number of diagnoses and on close-up or dermoscopic images. On the other hand, the clinicians are not trained to diagnose lesions only from dermoscopy or clinical images. In real clinical settings, human raters use additional information to reach a diagnosis, including the patient and the lesion's clinical data and history, the combination of the naked eye and dermoscopic appearance of the lesion, the comparison of the lesion to other lesions on the patient's body, and the ability to palpate the lesion. Therefore, the use of metadata is critical and in most of the studies, no metadata was included. In this sense, this environment favors AI over human rates and does not represent true clinical settings. In the existence of all these metadata, the performance of the human raters would improve significantly. However, it is still an active area of research on how much AI can benefit from such side information.

Challenges in the Development of AI for Skin Cancer Screening

In addition, there are numerous challenges in the development and implementation of AI in the field of skin cancer screening. This section will review the main ones.

1. The need for large data pipelines

The development of clinical level AI requires large data pipelines to train the algorithms. Aside from the ISIC archive, most open access image databases of skin lesions currently include a relatively small number of images. Among other things, the development of open access databases is limited by various factors, such as image copyright issues and patient privacy. In addition, most image databases currently include primarily dermoscopic images and lack clinical images or full-body photography images, which limit the use of AI for screening dermoscopic images and not patients.

2. “Ground truth” diagnosis

For an ML algorithm to train and “learn” the relationship between pixel data and lesion classification, it requires “ground truth” diagnosis tagging for the images. In dermatology, “ground truth” diagnosis is traditionally considered to be the histological diagnosis. This poses two major challenges: first, histology is an operator-dependent test and some cases are read as different diagnoses by different pathologists [36]. Second, the inclusion of only biopsied lesions creates a bias in

the training sets in favor of malignant lesions and may hamper the algorithm's ability to accurately diagnose the most common benign skin lesions, such as angiomas or benign nevi, which are not routinely biopsied.

3. Lack of imaging standards

Many different factors may influence the appearance of a lesion in an image: lighting conditions, camera angle, camera distance, color calibration, etc., which may change due to variations in (i) the imaging conditions and/or (ii) the specifications of imaging device manufactured by different companies (or even models). For the reproducibility of AI algorithms using different datasets/clinical settings, it is best if all these factors are standardized. However, today, most dermatological image acquisition process is non-standardized. There are attempts at creating standards in dermatological photography, but they are still complex and not easy to implement [37].

4. Lack of metadata

Two similarly appearing lesions can have different clinical significance in different clinical settings (e.g., a new Spitzoid lesion on a young child vs an elderly individual). However, today, most imaging databases that are used to train AI do not include any metadata, and algorithms are trained based on pixel data alone. Including metadata along with images in the future could enhance AI's ability to obtain a more accurate diagnosis, as demonstrated already by Yap et al. [27].

5. Lack of generalization

While there are thousands of different disease entities in dermatology, most image databases include large numbers of images on a limited number of diagnoses. A ML algorithm that is trained on only a few types of lesions will not perform well in a clinical environment where there are dozens of different lesions. In addition, imaging archives have been criticized for not including the entire spectrum of skin types, ethnicities, and geographies, and have a disproportional representation of lighter skin types. Algorithms trained with these databases may not perform well in clinical settings that include all skin types [38].

6. Lack of prospective studies in a clinical setting.

As mentioned above, all previous studies of ML technologies in the diagnosis of skin cancer have been performed in a controlled experimental environment on dermoscopic and/or close-up images that does not accurately represent the clinical settings. To create ML algorithms that will be relevant for real clinical settings, there is a need for prospective studies that will be performed in the same settings.

Future Considerations

With the development of CNNs in recent years, AI no longer depends on predefined features and is capable of learning them from the raw pixel data to generate classifications. This process includes multiple “hidden” layers that are unknown to the operator and lack clinical meaning. As for the approaches of future AIs, we expect the literature to move towards more transparent and clinically relevant diagnostic methods, which would be more relevant to the physician in clinical settings.

In addition, we expect the future to bring larger and better organized image databases, which will make it possible to train more accurate and comprehensive ML algorithms. This can be attributed to several trends—first, more physicians and medical centers use photography [39], generating a large pool of skin disease images. Second, images are expected to be standardized and coupled with metadata as efforts in creating these standards are increasing. Third, resolving regulatory and legal issues will make more images available from diverse geographical, ethnic, and cultural backgrounds.

Finally, many efforts and resources are invested in the development of AI, but the best way to implement AI in a clinical setting is still unclear. Will the end users be patients or physicians? How will the predictions be delivered to them? And what will the AI’s role be in guiding diagnosis and management? For example, it is not clear how an unexperienced clinician or a patient should deal with the result of 2% probability for melanoma [40]. It is the authors’ opinion that AI will not replace physicians in the near future. Rather, it will be a tool in their hands. Even if all the challenges mentioned above are overcome, there is still the issue of human nature: humans prefer to interact with humans, especially in medicine and even more so in the case of cancer diagnosis [41]. A potential way to implement AI in a clinical setting can be found in a study by Tschandl et al. [42]. The authors reported a neural network that presents the clinician with visually similar images based on features from the image in question. This type of models can be used as a tool to help enhance the clinician’s diagnostic accuracy.

Conclusions

The past 2 years have seen a dramatic progress in the development of AI for the diagnosis of skin lesions, mainly of pigmented skin lesions through dermoscopic images. However, these breakthroughs have all been in a controlled experimental environment with the exclusion of very critical metadata. In this sense, the adoption of AI-based diagnostic in the real dermatology clinical setting still is in its early stages and limited. Overall, the foremost issue is to establish a synergistic research environment between the dermatologists and computer scientists, where each side understands the needs

and constraints of both fields. Dermatologists should lead the discussion as to where AI should be integrated into skin cancer screening in a clinical setting in order to provide a real benefit to both the clinician and patient and to avoid any confusion or unnecessary stress and biopsies. On the other hand, computer scientists should lead the discussions on the data wise needs to achieve these aims and on providing new ways of analyzing and presenting the data to make the clinical practice more efficient.

Compliance with Ethics Guidelines

Conflict of Interest Allan C. Halpern has done work for Canfield Scientific, Inc. and for the Scibase Advisory Board.

Ofer Reiter, Veronica Rotemberg, and Kivanc Kose declare no conflict of interest.

Human and Animal Rights and Informed Consent This article does not contain any studies with human or animal subjects performed by any of the authors.

References

Papers of particular interest, published recently, have been highlighted as:

•• Of major importance

1. American Cancer Society, Cancer facts & figures 2019 [database on the Internet] 2019. Available from: <https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/annual-cancer-facts-and-figures/2019/cancer-facts-and-figures-2019.pdf>. Accessed:
2. Murzaku EC, Hayan S, Rao BK. Methods and rates of dermoscopy usage: a cross-sectional survey of US dermatologists stratified by years in practice. *J Am Acad Dermatol*. 2014;71(2):393–5. <https://doi.org/10.1016/j.jaad.2014.03.048>.
3. Dinnes J, Deeks JJ, Chuchu N, Ferrante di Ruffano L, Matin RN, Thomson DR et al. Dermoscopy, with and without visual inspection, for diagnosing melanoma in adults. The Cochrane database of systematic reviews. 2018;12:Cd011902. doi:<https://doi.org/10.1002/14651858.CD011902.pub2>.
4. Reiter O, Mimouni I, Gdalevich M, Marghoob AA, Levi A, Hodak E, et al. The diagnostic accuracy of dermoscopy for basal cell carcinoma: a systematic review and meta-analysis. *J Am Acad Dermatol*. 2018;80:1380–8. <https://doi.org/10.1016/j.jaad.2018.12.026>.
5. Dinnes J, Deeks JJ, Saleh D, Chuchu N, Bayliss SE, Patel L et al. Reflectance confocal microscopy for diagnosing cutaneous melanoma in adults. The Cochrane database of systematic reviews. 2018;12:Cd013190. doi:<https://doi.org/10.1002/14651858.Cd013190>.
6. Ferrante di Ruffano L, Dinnes J, Deeks JJ, Chuchu N, Bayliss SE, Davenport C et al. Optical coherence tomography for diagnosing skin cancer in adults. The Cochrane database of systematic reviews. 2018;12:Cd013189. doi:<https://doi.org/10.1002/14651858.Cd013189>.
7. Turing am. i.—Computing machinery and Intelligence. *mind*. 1950;LIX(236):433–460. doi:<https://doi.org/10.1093/mind/LIX.236.433> %J Mind.
8. Samuel AL. Some studies in machine learning using the game of checkers %J IBM. *J Res Dev*. 1959;3(3):210–29. <https://doi.org/10.1147/rd.33.0210>.

9. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. Proceedings of the 25th international conference on neural information processing systems - volume 1; Lake Tahoe, Nevada 2999257: Curran Associates Inc; 2012. p. 1097–105.
10. Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: a simple way to prevent neural networks from overfitting %J J. Mach Learn Res. 2014;15(1):1929–58.
11. Brinker TJ, Hekler A, Utikal JS, Grabe N, Schadendorf D, Klode J, et al. Skin cancer classification using convolutional neural networks: systematic review. *J Med Internet Res*. 2018;20(10):e11936. <https://doi.org/10.2196/11936>.
12. Gabriella Fabbrocini GB, Di Leo G, Liguori C, Paolillo A, Pietrosanto A, Sommella P, et al. Epiluminescence image processing for melanocytic skin lesion diagnosis based on 7-point check-list: a preliminary discussion on three parameters. *Open Dermatol J*. 2010;4(1):6.
13. Binder M, Steiner A, Schwarz M, Knollmayer S, Wolff K, Pehamberger H. Application of an artificial neural network in epiluminescence microscopy pattern analysis of pigmented skin lesions: a pilot study. *Br J Dermatol*. 1994;130(4):460–5.
14. Olga Russakovsky JD, Hao S, Krause J, Satheesh S, Ma S, Huang Z, et al. ImageNet large scale visual recognition challenge. *Int J Comput Vis*. 2015;115(3):211–52.
15. Weber P, Tschandl P, Sinz C, Kittler H. Dermatoscopy of neoplastic skin lesions: recent advances, updates, and revisions. *Curr Treat Options in Oncol*. 2018;19(11):56. <https://doi.org/10.1007/s11864-018-0573-6>.
16. 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 doi:10.1038/nature21056. **This is a landmark study that started the current trend of using CNNs for dermatologic image classification.**
17. 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 Invest Dermatol*. 2018;138(7):1529–38. <https://doi.org/10.1016/j.jid.2018.01.028> **This is the second landmark study that showed the ability of CNNs to differentiate between multiple lesion types.**
18. Tschandl P, Rosendahl C, Akay BN, Argenziano G, Blum A, Braun RP, 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. Ferrante di Ruffano L, Takwoingi Y, Dinnes J, Chuchu N, Bayliss SE, Davenport C, et al. Computer-assisted diagnosis techniques (dermoscopy and spectroscopy-based) for diagnosing skin cancer in adults. *Cochrane Database Syst Rev*. 2018. <https://doi.org/10.1002/14651858.Cd013186>.
20. Marka A, Carter JB, Toto E, Hassanpour S. Automated detection of nonmelanoma skin cancer using digital images: a systematic review. *BMC Med Imaging*. 2019;19(1):21. <https://doi.org/10.1186/s12880-019-0307-7>.
21. Haenssle HA, Fink C, Schneiderbauer R, Toberer F, Buhl 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. <https://doi.org/10.1093/annonc/mdy166>.
22. Brinker TJ, Hekler A, Enk AH, Klode J, Hauschild A, Berking C, et al. A convolutional neural network trained with dermoscopic images performed on par with 145 dermatologists in a clinical melanoma image classification task. *Eur J Cancer*. 2019;111:148–54. <https://doi.org/10.1016/j.ejca.2019.02.005>.
23. ISIC ARCHIVE. <https://www.isic-archive.com>.
24. Marchetti MA, Codella NCF, Dusza SW, Gutman DA, Helba B, Kallou A, 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(2):270–7 e1. <https://doi.org/10.1016/j.jaad.2017.08.016> **This is the first ISIC publication, where for the first time, AI developers were invited to use a ML challenge. It was open for everyone to participate and images were provided for free.**
25. Gutman D, Codella, N.C., Celebi, M.E., Helba, B., Marchetti, M.A., Mishra, N.K., & Halpern, A. . Skin lesion analysis toward melanoma detection: a challenge at the 2017 International symposium on biomedical imaging (ISBI), hosted by the international skin imaging collaboration (ISIC). *IEEE 15th international symposium on biomedical imaging (ISBI 2018)*, 168–172. 2018.
26. Tschandl P, Rosendahl C, Kittler H. The HAM10000 dataset: a large collection of multi-source dermoscopic images of common pigmented skin lesions. 2018.
27. 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>.
28. Roffinan D, Hart G, Girardi M, Ko CJ, Deng J. Predicting non-melanoma skin cancer via a multi-parameterized artificial neural network. *Sci Rep*. 2018;8(1):1701. <https://doi.org/10.1038/s41598-018-19907-9>.
29. Chuchu N, Takwoingi Y, Dinnes J, Matin RN, Bassett O, Moreau JF, et al. Smartphone applications for triaging adults with skin lesions that are suspicious for melanoma. *Cochrane Database Syst Rev*. 2018. <https://doi.org/10.1002/14651858.Cd013192>.
30. Marvdashti T, Duan L, Aasi SZ, Tang JY, Ellerbee Bowden AK. Classification of basal cell carcinoma in human skin using machine learning and quantitative features captured by polarization sensitive optical coherence tomography. *Biomed Opt Express*. 2016;7(9):3721–35. <https://doi.org/10.1364/BOE.7.003721>.
31. Bozkurt A, Kose K, Alessi-Fox C, Gill M, Dy J, Brooks D et al. A multiresolution convolutional neural network with partial label training for annotating reflectance confocal microscopy images of skin: 21st international conference, Granada, Spain, September 16–20 2018, Proceedings, Part II. 2018. p. 292–299.
32. Bozkurt A, Kose K, Alessi-Fox C, Dy JG, Brooks DH, Rajadhyaksha M. Unsupervised delineation of stratum corneum using reflectance confocal microscopy and spectral clustering. *Skin Res Technol : Off J Int Soc Bioeng Skin (ISBS) [and] Int Soc Digital Imaging of Skin (ISDIS) [and] Int Soc Skin Imaging (ISSI)* 2017;23(2):176–185. doi:<https://doi.org/10.1111/srt.12316>.
33. Kurugol S, Kose K, Park B, Dy JG, Brooks DH, Rajadhyaksha M. Automated delineation of dermal-epidermal junction in reflectance confocal microscopy image stacks of human skin. *J Invest Dermatol*. 2015;135(3):710–7. <https://doi.org/10.1038/jid.2014.379>.
34. Bozkurt A, Kose K, Coll-Font J, Alessi-Fox C, Brooks DH, Dy JG et al. Delineation of skin strata in reflectance confocal microscopy images using recurrent convolutional networks with Toeplitz Attention arXiv e-prints 2017.
35. Hosking AM, Coakley BJ, Chang D, Talebi-Liasi F, Lish S, Lee SW, et al. Hyperspectral imaging in automated digital dermoscopy screening for melanoma. *Lasers Surg Med*. 2019;51(3):214–22. <https://doi.org/10.1002/lsm.23055>.
36. Elmore JG, Barnhill RL, Elder DE, Longton GM, Pepe MS, Reisch LM, et al. Pathologists' diagnosis of invasive melanoma and melanocytic proliferations: observer accuracy and reproducibility study. *BMJ (Clinical research ed)*. 2017;357:j2813. <https://doi.org/10.1136/bmj.j2813>.
37. Finnane A, Curiel-Lewandrowski C, Wimberley G, Caffery L, Katragadda C, Halpern A, et al. Proposed technical guidelines for the acquisition of clinical images of skin-related conditions. *JAMA Dermatol*. 2017;153(5):453–7. <https://doi.org/10.1001/jamadermatol.2016.6214>.

38. 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>.
39. Milam EC, Leger MC. Use of medical photography among dermatologists: a nationwide online survey study. *J Eur Acad Dermatol Venereol : JEADV*. 2018;32(10):1804–9. <https://doi.org/10.1111/jdv.14839>.
40. Zakhem GA, Motosko CC, Ho RS. How should artificial intelligence screen for skin cancer and deliver diagnostic predictions to patients? *JAMA Dermatol*. 2018;154(12):1383–4. <https://doi.org/10.1001/jamadermatol.2018.2714>.
41. Lallas A, Argenziano G. Artificial intelligence and melanoma diagnosis: ignoring human nature may lead to false predictions. *Dermatol Pract Concept*. 2018;8(4):249–51. <https://doi.org/10.5826/dpc.0804a01>.
42. Tschandl P, Argenziano G, Razmara M, Yap J. Diagnostic accuracy of content-based dermatoscopic image retrieval with deep classification features. *Br J Dermatol*. 2018. <https://doi.org/10.1111/bjd.17189>.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.