Published in final edited form as:

J Am Acad Dermatol. 2022 December ; 87(6): 1352–1360. doi:10.1016/j.jaad.2020.05.053.

Deep Learning for Dermatologists: Part II. Current Applications

Pranav Puri, B.A.[2,10], **Nneka Comfere, M.D.**[3,8,10], **Lisa A. Drage, M.D.**[3,10], **Huma Shamim, MBBS**[3,10], **Spencer A. Bezalel, M.D.**[3,10], **Mark R. Pittelkow, M.D.**[5,10], **Mark D.P. Davis, M.D.**[3,10], **Michael Wang, M.D.**[4], **Aaron R. Mangold, M.D.**[5,10], **Megha M. Tollefson, M.D.**[3,10], **Julia S. Lehman, M.D.**[3,8,10], **Alexander Meves, M.D.**[3,10], **James A. Yiannias, M.D.**[5], **Clark C. Otley, M.D.**[3,10], **Rickey E. Carter, Ph.D.**[6], **Olayemi Sokumbi, M.D.**[7,9,10], **Matthew R. Hall, M.D.**[7,10], **Alina G. Bridges, D.O.**[3,8,10], **Dennis H. Murphree, Ph.D.**[1,10]

[1]Department of Health Sciences Research, Division of Digital Health Sciences, Mayo Clinic, Rochester, Minnesota

[2]Mayo Clinic Alix School of Medicine, Scottsdale, Arizona

[3]Department of Dermatology, Mayo Clinic, Rochester, Minnesota

[4]Department of Dermatology, University of California San Francisco, San Francisco, California

[5]Department of Dermatology, Mayo Clinic, Scottsdale, Arizona

[6]Department of Health Sciences Research, Division of Biomedical Statistics and Informatics, Mayo Clinic, Jacksonville, Florida

[7]Department of Dermatology, Mayo Clinic, Jacksonville, Florida

[8]Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, Minnesota

[9]Department of Laboratory Medicine and Pathology, Mayo Clinic, Jacksonville, Florida

[10]Mayo Clinic Office of Artificial Intelligence in Dermatology (AIDE)

Abstract

Due to a convergence of the availability of large datasets, graphics-specific computer hardware, and important theoretical advancements, artificial intelligence (AI) has recently contributed to dramatic progress in medicine. One type of artificial intelligence known as deep learning (DL) has been particularly impactful for medical image analysis. Deep learning applications have shown promising results in dermatology and other specialties including radiology, cardiology and ophthalmology. The modern clinician will benefit from an understanding of the basic features of deep learning in order to effectively use new applications as well as to better gauge their utility and limitations. In this second article of a two part series, we review the existing and emerging clinical applications of deep learning in dermatology and discuss future opportunities and limitations. Part 1 of this series offered an introduction to the basic concepts of deep learning to facilitate effective communication between clinicians and technical experts.

Corresponding author: Nneka Comfere M.D., 200 First St SW, Rochester, MN 55905, **Telephone:** 507-284-2511 comfere.nneka@mayo.edu.

Keywords

deep learning; machine learning; dermatology; artificial intelligence

INTRODUCTION

Artificial intelligence (AI) was initially defined by John McCarthy in 1955 as the "science and engineering of creating intelligent machines that have the ability to achieve goals like humans via a constellation of technologies."[1,2] Since then, AI and its various subfields have played a transformative role in shaping human social and cultural institutions. The American Academy of Dermatology has recognized the potential impact of AI in dermatology and recently released an official position statement on the topic.[3] In this statement, augmented intelligence (AuI) is proposed as an alternative term that highlights the complementary and assistive role of AI in human decision making.[3] This terminology affirms the central role of clinicians in the physician-patient relationship and highlights machine-learning as an assistive tool to enhance patient care. Deep learning is a subset of AI that applies artificial neural networks to make predictions. Deep learning's potential to enhance the practice of dermatology has been widely discussed[3–6] with applications ranging from image classification to discovering novel risk factors for non-melanoma skin cancers.[7,8] At the same time, a recent JAAD systematic review found that dermatologists are underrepresented in publications on AI in dermatology, authoring only 41% of published papers on the subject.[9] Here we highlight the emerging clinical applications of deep learning in dermatology and discuss future opportunities and challenges.

BRIEF REVIEW

Most recent advances of AI in dermatology have leveraged a type of algorithm known as a neural network (NN), and more specifically a variety known as a deep neural network (DNN)[10–13]. Neural networks are small computer programs that take input data and process it to output predictions. Examples include diagnosing a skin disease from a clinical photograph or highlighting tumor regions in a pathology slide. NNs are trained for a particular task using examples with known outcomes. NNs then make predictions on these examples, and these predictions are evaluated for performance. Through the process of learning, internal parameters called weights are updated based on the performance, and new predictions are generated. This cycle is iterated until the network's predictive capability is deemed acceptable at which point it can be considered for more general use.

An important improvement to the traditional training of DNNs is transfer learning. In transfer learning, instead of starting from scratch, one begins with a network that is known to perform well on a similar problem. Transfer learning dramatically reduces the amount of training data required, which is particularly important in medicine, where examples with known outcomes can be challenging to acquire. Unsurprisingly, many of the published dermatology deep learning studies to date, including some of those discussed in this review, used transfer learning to train their DNNs.

Part I of this series offers a more detailed overview of both neural networks and transfer learning. Table 1 provides a summary of relevant literature on applications of deep learning in dermatology. Table 1 includes papers that are not discussed in detail in the body of this manuscript.

APPLICATIONS

Classification and Differentiation

The most frequently used form of deep neural network in image analysis is a variant known as a convolutional neural network (CNN). In dermatology, deep CNNs have primarily been employed to classify images of pigmented and non-pigmented skin lesions. For example, Esteva et al.[14] developed a CNN model to identify epidermal and melanocytic lesions, then compared its performance to 21 board-certified dermatologists on two specific tasks: distinguishing squamous cell carcinomas (SCC) from benign seborrheic keratoses (SK) and malignant melanomas from benign nevi. On a biopsy-proven test set of 135 epidermal, 130 melanocytic non-dermoscopy images and 111 melanocytic dermoscopy images, dermatologists were asked whether to biopsy, treat the lesion or reassure the patient without biopsy. In parallel, the CNN was tasked with classifying the same lesions. The network outperformed the average performance of the dermatologists in each case. The authors conclude by graphical inspection that the CNN's performance was similar to that of the board-certified dermatologists.[14] However, we note that no formal statistical test was applied..

Haenssle *et al.*[15] similarly sought to compare the performance of a CNN trained to recognize melanoma in dermoscopic images to 58 international dermatologists with varying levels of experience in dermoscopy (29% beginner, 19% skilled, and 52% expert by self-report). The dermatologists were asked to classify lesions in two experiments termed level-I and level-II. In level-I, dermatologists classified lesions based on dermoscopy only. After a 4-week washout period, level-II was conducted in which dermatologists were provided dermoscopy, clinical images and additional clinical information. The CNN was trained on images only. Performance was assessed by Receiver Operator Characteristic (ROC) Area Under Curve (AUC). For the level-I task, the CNN's performance (0.86) exceeded that of the dermatologists (expert: 0.82, skilled: 0.77, beginner: 0.75). For the level-II tasks, the dermatologists' accuracy significantly improved with additional clinical information; however, their AUC (expert: 0.84, skilled: 0.84, beginner: 0.79) did not surpass the CNN (0.86).[15] This study highlights the importance of including a large group of dermatologists with varying levels of experience, as well as using open source datasets and lesions from different anatomic sites and of different histologic types during CNN training. They also demonstrate the importance of integrating clinical information and clinical experience when comparing human performance to algorithmic performance.

While the above studies are highlighted for their CNN performance in comparison to human dermatologists, there are numerous studies that address lesion identification.[16–19] Most of these studies are focused on improved algorithmic performance, but others demonstrate noteworthy results. For example, Han et al demonstrated that CNNs trained on images from Asian patients performed poorly on Caucasian patients and vice-versa, highlighting

the importance of training CNNs with skin lesions from a wide range of age groups and ethnicities.[20]

Dermatopathology

Deep learning in dermatopathology is centered around whole slides that are digitized into images by scanners. Andres et al. developed a diagnostic support tool to identify mitotic cells within detected tumor regions for whole slide images (WSI). The authors report a diagnostic accuracy of 83% for their model trained on 59 WSIs.[21] This tool could augment a dermatopathologist's practice by identifying areas of the slide with the highest density of mitotic figures, and could also potentially reduce the need for the immunohistochemical stains for mitosis. Olsen et al similarly trained a CNN using 450 WSI to classify basal cell carcinomas, dermal nevi, and seborrheic keratoses. Their Visual Geometry Group(VGG) network achieved an AUC of 0.99 for basal cell carcinomas, 0.97 for dermal nevi, and 0.99 for seborrheic keratoses.[22]

Hart et al. developed a CNN to differentiate between Spitz and conventional melanocytic lesions on histopathology. They trained their model on 100 curated whole slide images and first evaluated their model on curated image sections. Their model demonstrated 99% accuracy in this experiment. They then conducted a second experiment evaluating the model's performance on noncurated image patches of the entire slide. In contrast to the curated experiment, the model achieved a significantly lower accuracy of 52.3% on the non-curated patches.[23] Hekler et al built a similar CNN trained on 695 whole slide images to classify images as melanoma or benign nevi. They compared the performance of their CNN to dermatopathologists. Performance was evaluated on randomly cropped 10x magnification sections. The CNN achieved a melanoma sensitivity/specificity/accuracy of 76%/60%/68% respectively, while the 11 dermatopathologists achieved a mean sensitivity/specificity/accuracy of 51.8%/66.5%/59.2% respectively. However, these results should be interpreted with caution. In a normal clinical setting, pathologists have the ability to evaluate the whole slide and are not restricted to randomly cropped segments.[24]

FUTURE DIRECTIONS

To date, deep learning applications have been limited by the lack of large, labeled dermatologic image datasets. Such databases have yet to be created due to patient privacy concerns and high costs of obtaining expert-labeled images. In a recent JAAD article, Park. *et. al* point to the promise of the AAD's DataDerm clinical registry as a potential resource for de-identified, clinically robust data for dermatologic images.[25] This type of aggregate database would enable the creation of more powerful and accurate CNNs for dermatologic applications.

Future efforts should ensure training datasets include images representative of diverse populations in terms of race, skin phototype, age, and anatomic body site. To the same end, data augmentation techniques such as rotating and translating images from left to right and vice versa can improve CNN performance in scenarios when additional data cannot be obtained. For example, the ISIC database under-represents patients of color.[26] However, data augmentation techniques could be applied to artificially increase sample sizes

of patients of color and improve diagnostic accuracy for patients from underrepresented patient populations in existing datasets.

Investigations into the applications of deep learning to identify novel associations and predictive factors for skin diseases are warranted. A study by Roffman et al. developed a novel approach of using electronic health record data rather than images to predict non-melanoma skin cancer risk. The study identified 13 health parameters that were predictive of NMSC including body mass index, diabetic status, emphysema, and exercise habits. The model achieved a sensitivity of 86.2% and a specificity of 62.7%, despite not knowing UVR exposure or family history.[8] However, it is important to note that this study was not externally validated; and therefore, it is unclear whether these results are broadly generalizable.

Decision support systems employing deep learning could ensure diagnostic consistency, efficiency, and accuracy, thereby enhancing the quality and safety of patient care. Innovative telemedicine and teledermatology care delivery models that incorporate AI supported diagnostic systems could alleviate the global shortage of access to expert dermatologic care, particularly in rural and underserved communities[44–46]. This can facilitate the rapid and efficient provision of AI-assisted diagnoses and differential diagnoses with their associated probabilities directly to primary care clinicians at the point of care. Diagnostic decision support could also facilitate appropriate care triage and referral. In fact, numerous smartphone apps have recently been developed to detect high risk skin lesions in digital images. However, a Cochrane systematic review noted wide variability and inconsistencies in the accuracy of these apps, with sensitivities ranging from 7–73% and specificities ranging from 37–94%.[27]Additionally, workflow efficiencies could be gained by relieving dermatologists and dermatopathologists from time consuming or repetitive tasks such as detailed clinical assessments of multiple atypical nevi with dermoscopy in patients with dysplastic nevus syndrome and performing dermal mitotic counts in pathology respectively, for example. Automation of these tasks could enable dermatologists to more efficiently integrate and apply clinical data towards higher-order medical decision making.

To summarize, the potential benefits of deep learning tools include: 1) higher diagnostic consistency and accuracy, 2) earlier diagnosis and treatment, 3) improved efficiency and work flow, and 4) improved access to dermtatologic care.

LIMITATIONS

Dermatologists should view deep learning as a set of tools to augment clinicians' human capabilities rather than replace them. Physician and patient relationships grounded in trust and empathy are the foundation of medical practice.[28,29] Thus, the history taking, physical examination, consultation, and emotional reassurance dermatologists provide their patients are not threatened by advances in deep learning. Rather, ideally, deep learning tools will be developed to liberate dermatologists from repetitive and manual tasks so that they can provide more patient-centered and humanistic care.[3]

One of the most substantial limitations of deep learning is that the inner workings of the algorithm are essentially an unknown process or “black-box.” Although the inputs and outputs of the algorithm are observable, the mechanisms through which the algorithm forms its predictions are invisible to the user. This raises important concerns because clinicians strive to have justifications for their medical decision making. For example, information gained from the physical exam description of a lesion, its distribution, clinical course, and histopathological features provide a reasonably clear rationale in support of the clinical diagnosis. However, it currently is impossible for a dermatologist to extract a similar rationale for a suggested diagnosis from a neural network. Some argue that an accurate yet opaque prediction is preferable to an inaccurate transparent prediction. There are multiple examples of scenarios in dermatologic practice whereby medications used to treat diseases do not have a clear or defined underlying mechanism of action such as naltrexone for the management of Hailey-Hailey or hydroxychloroquine for management of photodermatoses.[30]

Another important limitation of deep learning is its susceptibility to selection bias. For example, the neural network developed by Esteva *et al.* “learned” to identify metric rulers as predictors of malignancy.[14] Certainly, rulers are not inherent predictors of malignancy, but rather the placement of a ruler in a clinical image reflects a clinician’s increased suspicion of a malignancy. Similarly, Winker et. al noted that surgical skin markings may have influenced CNNs to incorrectly classify lesions as melanocytic. Surgical skin markings increased their CNN’s false-positive rate by approximately 40%.[31]

Most studies to date have been retrospective and have utilized *in silico*, meaning computer based, validation. Future studies, however, should be prospectively validated in real world, clinical settings. This is of paramount importance because applying algorithms on curated, existing datasets is significantly different from applying algorithms on uncontrolled, noncurated data in a high-stakes clinical environment. Similarly, well-designed prospective clinical trials are able to minimize the effects of selection bias. Wang *et al.* conducted one of few existing prospective, randomized-controlled trials to prove the efficacy of a CNN for detecting polyps during colonoscopy.[32] Dermatologists should also subject CNNs to similar rigorous evaluation before incorporating CNNs into clinical practice.

Similarly, clinical adoption of CNNs potentially poses risks to patients in terms of inappropriate diagnoses, privacy breaches of identifiable data, and other harms.[33] Therefore, clinical adoption of CNNs will require regulatory oversight that includes extensive postmarket surveillance mechanisms to ensure the effectiveness and safety of CNN applications. The law and policy governing AI systems in health care has yet to be fully defined. As such, it is essential that dermatologists remain involved in policy-making efforts to ensure AI systems are safe and ethical for patients.

When evaluating a deep learning application, dermatologists should consider whether the clinical question is within the scope of the application. Since deep learning applications are statistical in nature, their performance is generally reduced when applied to tasks outside of their intended scope, environment, or patient population. For example, even though the ISIC collaboration found deep learning classifiers to outperform human experts for the diagnosis

of pigmented skin lesions, machine performance dropped dramatically when evaluated on images from different institutions or sources.[34] Deep learning performance is measured by sensitivity and specificity; however, clinicians must consider other factors including the training and testing patient populations as well as data quality and quantity.

Deep learning offers the promise of improving diagnosis for rare dermatologic conditions. Data augmentation can be utilized to artificially increase sample sizes and train CNNs for these applications. Clinicians must keep in mind the basic principles of conditional probability and Bayes' theorem. That is, algorithms for conditions with low prevalence are subject to high rates of false positives. Even if algorithms are developed to be highly sensitive and specific, they may not have high positive and negative predictive values due to low prevalence. For example, consider a patient presenting with epidermolysis bullosa (EB), a rare condition with a prevalence of 20 per 1 million live births.[35] Even if a CNN were trained to be 99% sensitive and 99% specific for the detection of EB, the positive predictive value would only be 0.2%.

CONCLUSIONS

Deep learning offers numerous innovations, solutions and support to enhance the clinical practice of dermatology. Deep learning applications have already demonstrated dermatologist-level accuracy in the classification of numerous skin lesions. These algorithms should be further refined and rigorously validated in prospective randomized controlled trials to improve patient care and safety, enhance the productivity of dermatologists and dermatopathologists, and improve access to high quality dermatologic care.

REFERENCES

1. Topol E Deep Medicine. How Artificial Intelligence can make healthcare human again. New York, NY: Basic Books by Hachette Book Group; 2019.
2. McCarthy J What is AI? http://jmc.stanford.edu/articles/whatisai/whatisai.pdf. Accessed August 18, 2019.
3. Dermatology AAo. Position Statement on Augmented Intelligence. American Academy of Dermatology;2019.
4. Aractingi S, Pellacani G. Computational neural network in melanocytic lesions diagnosis: artificial intelligence to improve diagnosis in dermatology? Eur J Dermatol. 2019;29(S1):4–7. [PubMed: 31017580]
5. Schlessinger DI, Chhor G, Gevaert O, Swetter SM, Ko J, Novoa RA. Artificial intelligence and dermatology: opportunities, challenges, and future directions. Semin Cutan Med Surg. 2019;38(1):E31–37. [PubMed: 31051021]
6. Hulstaert E, Hulstaert L. Artificial intelligence in dermato-oncology: a joint clinical and data science perspective. Int J Dermatol. 2019.
7. Han SS, Park GH, Lim W, et al. Deep neural networks show an equivalent and often superior performance to dermatologists in onychomycosis diagnosis: Automatic construction of onychomycosis datasets by region-based convolutional deep neural network. PLoS One. 2018;13(1):e0191493. [PubMed: 29352285]
8. Roffman 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. [PubMed: 29374196]
9. Zakhem GA, Fakhoury JW, Motosko CC, Ho RS. Characterizing the role of dermatologists in developing artificial intelligence for assessment of skin cancer: A systematic review. J Am Acad Dermatol. 2020.

10. Litjens G, Ciompi F, Wolterink JM, et al. State-of-the-Art Deep Learning in Cardiovascular Image Analysis. JACC Cardiovasc Imaging. 2019;12(8 Pt 1):1549–1565. [PubMed: 31395244]
11. Ting DSW, Peng L, Varadarajan AV, et al. Deep learning in ophthalmology: The technical and clinical considerations. Prog Retin Eye Res. 2019;72:100759. [PubMed: 31048019]
12. Valliani AA, Ranti D, Oermann EK. Deep Learning and Neurology: A Systematic Review. Neurol Ther. 2019;8(2):351–365. [PubMed: 31435868]
13. Wang S, Yang DM, Rong R, Zhan X, Xiao G. Pathology Image Analysis Using Segmentation Deep Learning Algorithms. Am J Pathol. 2019;189(9):1686–1698. [PubMed: 31199919]
14. Esteva A, Kuprel B, Novoa RA, et al. Dermatologist-level classification of skin cancer with deep neural networks. Nature. 2017;542(7639):115–118. [PubMed: 28117445]
15. 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):1836–1842. [PubMed: 29846502]
16. Thomsen K, Iversen L, Titlestad TL, Winther O. Systematic review of machine learning for diagnosis and prognosis in dermatology. J Dermatolog Treat. 2019:1–15.
17. 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(2):270–277.e271. [PubMed: 28969863]
18. Brinker TJ, Hekler A, Hauschild A, et al. Comparing artificial intelligence algorithms to 157 German dermatologists: the melanoma classification benchmark. Eur J Cancer. 2019;111:30–37. [PubMed: 30802784]
19. Yap J, Yolland W, Tschandl P. Multimodal skin lesion classification using deep learning. Exp Dermatol. 2018;27(11):1261–1267. [PubMed: 30187575]
20. 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):1529–1538. [PubMed: 29428356]
21. Andres C, Andres-Belloni B, Hein R, et al. iDermatoPath - a novel software tool for mitosis detection in H&E-stained tissue sections of malignant melanoma. J Eur Acad Dermatol Venereol. 2017;31(7):1137–1147. [PubMed: 28107565]
22. Olsen TG, Jackson BH, Feeser TA, et al. Diagnostic Performance of Deep Learning Algorithms Applied to Three Common Diagnoses in Dermatopathology. J Pathol Inform. 2018;9:32. [PubMed: 30294501]
23. Hart SN, Flotte W, Norgan AP, et al. Classification of Melanocytic Lesions in Selected and Whole-Slide Images via Convolutional Neural Networks. J Pathol Inform. 2019;10:5. [PubMed: 30972224]
24. Hekler A, Utikal JS, Enk AH, et al. Deep learning outperformed 11 pathologists in the classification of histopathological melanoma images. Eur J Cancer. 2019;118:91–96. [PubMed: 31325876]
25. Park AJ, Ko JM, Swerlick RA. Crowdsourcing dermatology: DataDerm, big data analytics, and machine learning technology. J Am Acad Dermatol. 2018;78(3):643–644. [PubMed: 29042152]
26. Adamson AS, Smith A. Machine Learning and Health Care Disparities in Dermatology. JAMA Dermatology. 2018;154(11):1247–1248. [PubMed: 30073260]
27. Chuchu N, Takwoingi Y, Dinnes J, et al. Smartphone applications for triaging adults with skin lesions that are suspicious for melanoma. Cochrane Database Syst Rev. 2018;12:Cd013192. [PubMed: 30521685]
28. Rydon-Grange M Psychological perspective on compassion in modern healthcare settings. J Med Ethics. 2018;44(11):729–733. [PubMed: 29959187]
29. de Zulueta PC. Suffering, compassion and ‘doing good medical ethics’. J Med Ethics. 2015;41(1):87–90. [PubMed: 25516944]
30. Albers LN, Arbiser JL, Feldman RJ. Treatment of Hailey-Hailey Disease With Low-Dose Naltrexone. JAMA Dermatol. 2017;153(10):1018–1020. [PubMed: 28768313]

31. 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.
32. Wang P, Berzin TM, Glissen Brown JR, et al. Real-time automatic detection system increases colonoscopic polyp and adenoma detection rates: a prospective randomised controlled study. Gut. 2019.
33. Evans EL, Whicher D. What Should Oversight of Clinical Decision Support Systems Look Like? AMA J Ethics. 2018;20(9):E857–863. [PubMed: 30242817]
34. 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):938–947. [PubMed: 31201137]
35. Fine JD. Epidemiology of Inherited Epidermolysis Bullosa Based on Incidence and Prevalence Estimates From the National Epidermolysis Bullosa Registry. JAMA Dermatol. 2016;152(11):1231–1238. [PubMed: 27463098]
36. Aggarwal SLP. Data augmentation in dermatology image recognition using machine learning. Skin Res Technol. 2019.
37. 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. [PubMed: 30484822]
38. Han SS, Moon IJ, Lim W, et al. Keratinocytic Skin Cancer Detection on the Face Using Region-Based Convolutional Neural Network. JAMA Dermatol. 2019.
39. Hekler A, Utikal JS, Enk AH, et al. Pathologist-level classification of histopathological melanoma images with deep neural networks. Eur J Cancer. 2019;115:79–83. [PubMed: 31129383]
40. Jiang YQ, Xiong JH, Li HY, et al. Recognizing Basal Cell Carcinoma on Smartphone-Captured Digital Histopathology Images with Deep Neural Network. Br J Dermatol. 2019.
41. Lott JP, Boudreau DM, Barnhill RL, et al. Population-Based Analysis of Histologically Confirmed Melanocytic Proliferations Using Natural Language Processing. JAMA Dermatol. 2018;154(1):24–29. [PubMed: 29094145]
42. Bernardis E, Castelo-Soccio L. Quantifying Alopecia Areata via Texture Analysis to Automate the SALT Score Computation. J Investig Dermatol Symp Proc. 2018;19(1):S34–S40.
43. 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. 2017;23(2):176–185. [PubMed: 27516408]
44. Seetharam K, Kagiyama N, Sengupta PP. Application of moblile health, telemedicine and artificial intelligence to echocardiography. Echo Res Pract. 2019;6(2):R41–R52 [PubMed: 30844756]
45. Pasquali P, Sonthalia S, Moreno-Ramirez Det al. Teledermatology and its current perspective. Indian Dermatol Online J. 2020;11(1): 12–20. [PubMed: 32055502]
46. Ting DS, Gunasekeran DV, Wickahm L. et al. Next generation telemedicine platforms to screen and triage. Br J Opthalmol. 2020;104(3):299–300

Table 1:

Recent Studies in Dermatology utilizing Artificial Intelligence

Study	Objectives and key findings	Algorithm	Sample Size
Skin Cancer Classifiction			
Esteva et al.[14](2017)	Classifying images of 1.) keratinocyte carcinomas vs. benign seborrheic keratosis and 2.) malignant melanomas vs. benign nevi. ROC AUC of 0.96 for carcinomas and 0.96 for melanomas.	GoogLeNet Inception v3 model	129,450 images
Han et al.[20](2018)	Classifying 12 different skin lesions based on clinical images. ROC AUCs for BCC, SCC, melanoma were 0.96, 0.83, 0.96, respectively	ResNet	19,398 images
Haenssle et al.[15](2018)	Classifying dermatoscopic images of melanoma vs benign nevi. Compared performance to 58 dermatologists. Dermatologists had improved sensitivity and specificity with additional clinical info but still did not perform as well as CNN.	GoogLeNet Inception v3 model	not provided
Marchetti et al.[17](2018)	Reported the results of 25 deep learning algorithms trained to classify melanomas vs. benign nevi. Top 5 algorithms were combined in a fusion algorithm which achieved ROC AUC of 0.86	Variable	900 images
Brinker el al.[18](2019)	Trained a melanoma classification model on dermatoscopic images and then evaluated it on clinical images. For equal levels of sensitivity (89.4%), the model achieved a higher specificity (68.2%) than the mean of 157 dermatologists (64.4%)	ResNet	20,735 images
Yap et al.[19](2018)	Combined dermatoscopic images with macroscopic images and patient metadata to train a CNN for skin lesion classification. They report an ROC AUC of 0.866 for melanoma detection.	ResNet	2,917 cases
Aggarwal et al.[36](2019)	Deep learning algorithms were evaluated before and after data augmentation. For the five classes of lesions tested, data augmentation increased AUC by 0.132 on average	TensorFlow Inception version-3	938 images
Tschandl et al.[37](2019)	Classification of nonpigmented skin cancer, compared performance to 95 human raters. ROC AUC of 0.742. CNN had higher percentage of correct diagnosis (37.6%) compared to average human rater (33.5%) but not compared to experts (37.3%).	TensorFlow Inception version-3	7895 dermoscopic images, 5829 close up images
Han et al.[38](2019)	Using unprocessed photos, localize and diagnose keratinocytic skin cancers without manual preselection of suspicious lesions by dermatologists. ROC AUC of 0.910. F1 score of 0.831. Youden index of 0.675	Region-based Convolutional Neural Network	182,348 images
Dermatopathology			
Hekler et al.[39](2019)	A CNN was trained with images of melanomas and nevi. The CNN had a 19% discordance with the histopathologist. This is comparable to the discordance between human pathologists (25–26%)	ResNet	595 whole slide images
Olsen et al.[22](2018)	CNN was trained on whole slide images to identify basal cell carcinomas, dermal nevi, and seborrheic keratoses. The ROC AUC were 0.99, 0.97, 0.99 respectively	VGG	450 whole slide images
Hart et al.[23](2019)	Developed a CNN trained on whole slide images to distinguish between spitz and conventional melanocytic lesions. They report an accuracy of 92% for a single call from a whole slide image.	TensorFlow Inception version-3	100 whole slide images
Jiang et al.[40](2019)	Used smartphone cameras to photograph microscope ocular images instead of using whole slide images. Developed CNN to recognize BCC. ROC AUC of 0.95	GoogLeNet Inception v3 model	8,046 microscopic ocular images
Hekler et al.[24](2019)	Developed a CNN to classify histopathologic images as melanoma or nevi. CNN achieved a mean sensitivity/specificity/accuracy of 76%/60%/68%. The 11 pathologists achieved 52%/67%/59%	ResNet	695 whole slide images
Predicting Novel Risk Factors/Epidemiology			
Roffman et al.[8](2018)	Trained their model with a data set of personal health information. The study identified 13 parameters that were predictive of NMSC. The model achieved a sensitivity of 86.2% and a specificity of 62.7%	Did not specify further than neural network	2,056 NMSC cases and 460,574 control cases
Lott et al.[41](2018)	Used natural language processing to analyze EMR pathology reports of patients who underwent skin biopsy to calculate population based	NegEx	80,368 skin biopsies

Study	Objectives and key findings	Algorithm	Sample Size
	frequencies of histologically confirmed melanocytic lesions. PPV 82.4%, sensitivity 81.7%, F1 measure 0.82		
Identifying Onychomycosis			
Han et al.[7](2018)	Classified nail images as onychomycosis or not. ROC AUC of 0.98	ensemble model combining ResNet-152 and VGG-19 models	49,567 images
Quantifying Alopecia Areata			
Bernardis et al.[42](2018)	Trained a texture classification model to calculate SALT score. Their model was able to calculate SALT scores in seconds, whereas the average time reported by clinicians is 7 minutes.	Texture classification model	250 images
Automating Reflectance Confocal Microscopy			
Bozkurt et al.[43](2017)	Developed a segmentation algorithm to distinguish the stratum corneum layer and accurately calculate its thickness	Complex Wavelet Transform	15 RCM stacks w/ 30 images in each stack
Mitosis Detection			
Andres et al.[21](2017)	Identified mitotic cells and detect relevant regions in whole slide images. The study reports an accuracy of 83% and a dice coefficient of 0.72.	Random Forest Classifier	59 whole slide images

Table 2.

GLOSSARY

Term	Definition
Machine Learning	The subset of artificial intelligence in which algorithms learn without being explicitly programmed
Deep Learning	A class of machine learning which applies artificial neural networks to train its algorithms by processing multilayered networks of data
Artificial Neural Networks	Systems modeled after the human brain's neurons, which adaptively learn and optimize performance.
Convolutional Neural Networks	A type of neural network, primarily used for images, which decomposes a dataset into smaller, overlapping tiles. These smaller, overlapping tiles create a compositional hierarchy of structures within the data.
Machine Learning Vocabulary	
Supervised Learning	Learns from known outcome, similar to studying from a set of Kodachrome images with known diagnoses
Unsupervised Learning	Learns without a known outcome, utilizing groupings of properties or features
Classification	Predicts content of image as a whole – is this a photo of melanoma or seborrheic keratosis?
Segmentation	Predicts regional contents of image – which parts of this biopsy are tumor regions?
Object Detection	Finds object of interest – where are the pigmented lesions on a patient's back
Predictors	Input variables, also known as independent variables or covariates. Pixels are the predictors when using a digital image.
Response	The predicted outcome, such as classification of a lesion's diagnosis from an input image
Transfer Learning	Training approach that starts with an existing pre-trained network rather than from scratch
Data Augmentation	Artificially creating valid new data from existing data, for example rotating a picture of a melanoma
Overfitting	Occurs when a model memorizes the idiosyncratic variations of a training set and does not learn generally useful data patterns. For example, including too many variables/features may improve internal validity, but may lead to biases in external validation.
Neural Network Training Terms	
Training Data	A subset of data that is used to teach and improve a network's performance.
Test Data	A subset of data used solely for evaluating a network's performance. The test data is independent from the training data.
Ground Truth	The most accurate reference based on the most up-to-date medical understanding. This usually refers to the annotation provided by physicians in supervised learning. In circumstances where disagreements exist between experts, the ground truth could be obtained from an average or a majority vote of professional opinions.
Back Propagation	Backpropagation of errors helps to update weights in the neural network to optimize accurate predications for current and future inputs. It is not a factor requiring decision/knowledge in practical experiments. While backpropagation is important and discussed in many textbooks, in practice it is completely abstracted by software.
Epoch	One "pass" through the entire training set. A training iteration during which the network examines all possible data points. The performance will often be plotted as a function of epoch.
Batch Size	The entire training set can be divided into smaller batches with a batch size of choosing, where the batch size is the number of training examples used in each backpropagation calculation (weight update). The batch size shapes overall training speed and performance, and is influenced by the size of available graphic processing units (GPUs). A GPU is specialized computer hardware helpful for deep learning.
Neural Network Components	
Weights	Internal numbers that "instruct" a network on how to make predictions. Weights are updated during training until they reach or closely approach the maximum potential performance.
Node	A basic computational unit of a neural network. The nodes combine inputs, weights, and an activation function to produce an output. These nodes can be viewed as neuron analogues, hence the name of "neural network."
Activation Function	A non-linear function that's part of a node or potentially a layer of nodes. These functions are critical for allowing neural networks to outperform the simpler linear models. Softmax and ReLu are commonly used activation functions.

Term	Definition
Loss	This value quantifies a network's performance on a given set of data. Higher loss indicates worse performance, so the goal is to minimize the loss through additional training. Common loss functions include binary cross-entropy and categorical cross-entropy (commonly mildly misattributed as Softmax loss).
Optimization Method	Numerical routine used to update the weights so as to minimize the loss, thus improve performance. Common optimization methods include SGD, RMSProp, and Adam.
Learning Rate	The learning rate is an important parameter that influences how much weights are adjusted during training. Best learning rates are likely found through experimentation.