Will Artificial Intelligence Replace Dermatologists?

SUSAN M. SWEETTER, MD
Professor of Dermatology
Director, Pigmented Lesion & Melanoma Program
Physician Leader, Cutaneous Oncology Program
Stanford University Medical Center and Cancer Institute
Susan M. Swetter, MD
S027 Managing Patients with Melanoma or Other Melanocytic Neoplasms

DISCLOSURES
I do not have any relevant relationships with industry.
How do we know what we know?

- Rules-based systems
- Unconscious pattern matching

Systems drive the way we think:
System 1 – fast, intuitive, emotional
System 2 – slower, more deliberate and logical
Dual-process theory applied to Dermatologists

**System 1 (intuitive)** – rapid, primarily visual, operates below level of perceptible consciousness. i.e. “gut feeling”
- focus on pattern recognition: “blink, think” and “10-second rule” (Giuseppe Argenziano, Naples)
- why “ugly duckling” rule discriminates melanoma more accurately than ABCD clinical warning signs (Gaudy-Marquest C et al. JAMA Dermatol 2017)

**System 2 (analytical)** – deliberate judgment, based on conscious applications of rules acquired through learning

Transition from intuitive to analytical reasoning can hinder clinical reasoning and increase diagnostic error

Skin Cancer Facts

- Skin cancer - most common cancer in the US
- 1 in 5 Americans will develop skin cancer in their lifetime
- In 2017 –
  - estimated >91K new cases of invasive melanoma and >87K melanomas in situ in the US
  - nearly 10,000 melanoma-associated deaths
- Survival rate for melanoma is >95% if detected early

“MELANOMA WRITES ITS MESSAGE IN THE
SKIN WITH ITS OWN INK
AND FOR ALL OF US TO SEE”
-Neville Davis, Queensland, Australia

…so why is early detection so hard?
Global Health Care Accessibility

Global access to quality health care has improved in the last two decades, but the gap between countries with the most and least effective treatments has grown.

Barber R et al. *The Lancet* 2017
How do we democratize health care access?

- Data collection at scale
- Diagnostics at scale
- Diminish health disparities
- 6.3 billion smartphones globally by 2021

Can we use AI to expand access to dermatologists?

AI is changing our world

- Driverless cars
- Translation capabilities
- Mortgage lending
- Financial markets
Neural Networks

- Modeled on human brain/neurons
- Inputs cause neuron to “fire” if over threshold
- AI: inputs fire again and again to get the right output
Multi-layer Neural Networks

Teach system to preserve the correct output/answer


http://cs231n.github.io/convolutional-networks/
Multi-layer Neural Networks

Generate an output

Supervised learning

- Output value is compared with the known value of the image- (“8” dog, cat, etc.)

Backpropagation \(\rightarrow\) run it back through, adjusting weights in order to reduce error

- Does this across thousands or millions of images

http://cs231n.github.io/convolutional-networks/
Convolutional Neural Networks

- Capture relationship of pixels to each other using filter as a matrix
- Runs algorithms over and over until errors are minimized across all images
- Allows for profound level of pattern recognition beyond human brain capability

https://ujjwalkarn.me/2016/08/11/intuitive-explanation-convnets/
Advances in AI

Breakthroughs in artificial intelligence over past 5 years

Convolutional neural networks + large databases + processing power = deep learning

Krizhevsky, Sutskever, Hinton et al. ImageNet Classification with Deep Convolutional Neural Networks, NIPS 2012
Algorithms train themselves to recognize what is important and what is not, without human intervention.

- **ImageNet**: large, deep CNNs trained to classify 1.4 mm high-resolution images into 1000 different classes with low error rates.
- If given sufficient examples of huskies, chihuahuas, basal cell carcinomas or melanomas, algorithms learn the relevant patterns of these categories.
If AI can differentiate between hundreds of dog breeds, it could make a great contribution to dermatology.

Stanford Dermatologist/ Dermatopathologist Roberto Novoa, MD
January 27, 2015
Constructing the dataset:

Nearly 130,000 images:
-clinician-labeled and/or biopsy-proven from 18 different, open-access online repositories

- clinical data from Stanford Dermatology clinics

Biopsy proven images:
- University of Edinburgh library
- ISDIS ISBI challenge (dermoscopic)

Images comprised more than 2,000 dermatologic diseases
Visual Taxonomy

129K images
2K diseases
Our Objectives

Evaluate performance of deep learning algorithms (namely, a single CNN) on classification of cutaneous malignancies

Compare to dermatologists

2 main questions:

- Lesion benign or malignant?
- Biopsy/treat or reassure?
Methods

Recruited 21 board-certified dermatologists
- Stanford University
- University of Pennsylvania
- Massachusetts General Hospital/Harvard
- University of Iowa

Three 100+ question tests
a. Epidermal tumors (BCC/SCC vs SKs)
b. Clinical images of melanocytic lesions (melanoma vs benign nevi, excluded SKs)
c. Dermoscopic images of melanocytic lesions (separate dataset)
Training

Goal: to build one system that could accommodate significant variation inherent in photographic images—lighting, zoom, angle, etc. with no pre-processing or lesion segmentation.

Essentially to be able to feed any captured skin image directly into the system and output a classification.
Deep Convolutional Neural Network (CNN)

GoogleNet Inception v3 CNN architecture pretrained on ImageNet;
- the net on our dataset was then trained and fine-tuned using transfer learning;
- results in probability distribution over clinical classes of skin
- training classes defined by applying a partitioning algorithm to our taxonomy
Utilized a new dermatologist-labeled dataset of >129K clinical images, including >3300 separate dermoscopic images.
Skin cancer: through the machine’s eye
Algorithm ready for testing within 1 year

CNN outputs a malignancy probability for each image
Evaluation

Algorithm validated in 2 ways using:

i) 3-class disease partition of 1st level nodes in our taxonomy (non-neoplastic, benign neoplastic, and malignant neoplastic) – with 72% overall accuracy

ii) class disease partition with 2nd level nodes – with 55% overall accuracy

Then only biopsy-labeled images used to conclusively validate the algorithm using the same CNN for all 3 tasks:

- keratinocyte carcinoma vs SK
- melanoma vs nevus
- melanoma vs nevus on dermoscopy

Results for CNN

CNN performed at least as well as dermatologists as a whole (Limitation of dermoscopy test as most dermatologists were not experts in pigmented lesions or dermoscopy)

LETTER

Dermatologist-level classification of skin cancer with deep neural networks

André Kosmopoulos, Ben Keating, Roberta Szwarcz, elder Ken, Trusnak Man and Andrew M. Szwarcz

Skin cancer, the most common human malignancy, is primarily diagnosed visually, beginning with an initial clinical screening and followed potentially by dermatoscopic analysis, a biopsy and histopathological examination. Automated classification of skin lesions using image analysis is a challenging task owing to the large variability in the appearance of skin lesions. Deep convolutional neural networks (CNNs)1,2 show potential for general and highly variable tasks across many fine-grained object categories3,4. In this work, we demonstrate classification of skin lesions using a single CNN trained on and evaluated on images directly, using only pixels and disease labels as inputs. We train a CNN using a dataset of 120,000 high-resolution images—two orders of magnitude larger than previous datasets5—consisting of 2,023 different diseases. We test our performance against 22 board-certified dermatologists on biopsy-confirmed clinical images with no clinical history classification use case. In our study, the dermatologists surpass the CNN for both dermoscopic images and macular melanoma lesions (92.9% CNN vs 96.4% dermatologists). The CNN achieves performance parity with all tested experts across both tasks, demonstrating an artificial intelligence capable of diagnosing skin cancer with a level of competence comparable to dermatologists. Outfitted with deep neural networks, mobile devices can potentially extend the reach of dermatologists outside of the clinic. It is projected that 6 million smartphone subscriptions will exist by the year 2020 (ref. 12), and can therefore potentially provide low-cost and universal access to vital diagnostic care.

There are 1.6 million new cases of skin cancer in the United States every year. Out of the five Americans will be diagnosed with a cutaneous malignancy in their lifetime. Although melanomas represent fewer than 5% of all skin cancers in the United States, they account for approximately 70% of all skin cancer-related deaths, and are responsible for over 13,000 deaths annually in the United States alone. Early detection is critical, as the estimated 5-year survival rate for melanomas drops from 98% if detected in its earliest stages to about 14% if detected in its latest stages. We developed a computational method which allows medical practitioners and patients to proactively track skin lesions and detect cancer earlier. By creating a novel disease taxonomy, and a disease-purifying algorithm that maps individual images into defining classes, we are able to build a learning system for automated dermatology.

Current work in dermatology remains flawed because of a lack of the generalization capability of medical practitioners owing to insufficient data and a focus on standardized tasks such as dermoscopy19,20 and histological image classification20. Dermoscopy images are acquired via a specialized microscope, and histopathological images are acquired via tissue biopsy and microscopy whereby both modalities yield highly standardized images. Photographs taken at different locations and using different equipment may vary in appearance and be difficult to use in a computer vision system. For example, antisymmetric images, such as lesions that are circular but not perfectly symmetric, can be difficult to classify. Moreover, the variability in the input images can be challenging to model and can lead to poor generalization. Therefore, we have established a dataset of 20,230 high-resolution images—two orders of magnitude larger than previous datasets5—consisting of 2,023 different diseases. We train a CNN using this dataset and evaluate its performance against 22 board-certified dermatologists. The CNN achieves performance parity with all tested experts across both tasks, demonstrating an artificial intelligence capable of diagnosing skin cancer with a level of competence comparable to dermatologists. Outfitted with deep neural networks, mobile devices can potentially extend the reach of dermatologists outside of the clinic. It is projected that 6 million smartphone subscriptions will exist by the year 2020 (ref. 12), and can therefore potentially provide low-cost and universal access to vital diagnostic care.

There are 1.6 million new cases of skin cancer in the United States every year. Out of the five Americans will be diagnosed with a cutaneous malignancy in their lifetime. Although melanomas represent fewer than 5% of all skin cancers in the United States, they account for approximately 70% of all skin cancer-related deaths, and are responsible for over 13,000 deaths annually in the United States alone. Early detection is critical, as the estimated 5-year survival rate for melanomas drops from 98% if detected in its earliest stages to about 14% if detected in its latest stages. We developed a computational method which allows medical practitioners and patients to proactively track skin lesions and detect cancer earlier. By creating a novel disease taxonomy, and a disease-purifying algorithm that maps individual images into defining classes, we are able to build a learning system for automated dermatology.

Current work in dermatology remains flawed because of a lack of the generalization capability of medical practitioners owing to insufficient data and a focus on standardized tasks such as dermoscopy19,20 and histological image classification20. Dermoscopy images are acquired via a specialized microscope, and histopathological images are acquired via tissue biopsy and microscopy whereby both modalities yield highly standardized images. Photographs taken at different locations and using different equipment may vary in appearance and be difficult to use in a computer vision system. For example, antisymmetric images, such as lesions that are circular but not perfectly symmetric, can be difficult to classify. Moreover, the variability in the input images can be challenging to model and can lead to poor generalization. Therefore, we have established a dataset of 20,230 high-resolution images—two orders of magnitude larger than previous datasets5—consisting of 2,023 different diseases. We train a CNN using this dataset and evaluate its performance against 22 board-certified dermatologists. The CNN achieves performance parity with all tested experts across both tasks, demonstrating an artificial intelligence capable of diagnosing skin cancer with a level of competence comparable to dermatologists. Outfitted with deep neural networks, mobile devices can potentially extend the reach of dermatologists outside of the clinic. It is projected that 6 million smartphone subscriptions will exist by the year 2020 (ref. 12), and can therefore potentially provide low-cost and universal access to vital diagnostic care.

There are 1.6 million new cases of skin cancer in the United States every year. Out of the five Americans will be diagnosed with a cutaneous malignancy in their lifetime. Although melanomas represent fewer than 5% of all skin cancers in the United States, they account for approximately 70% of all skin cancer-related deaths, and are responsible for over 13,000 deaths annually in the United States alone. Early detection is critical, as the estimated 5-year survival rate for melanomas drops from 98% if detected in its earliest stages to about 14% if detected in its latest stages. We developed a computational method which allows medical practitioners and patients to proactively track skin lesions and detect cancer earlier. By creating a novel disease taxonomy, and a disease-purifying algorithm that maps individual images into defining classes, we are able to build a learning system for automated dermatology.

Current work in dermatology remains flawed because of a lack of the generalization capability of medical practitioners owing to insufficient data and a focus on standardized tasks such as dermoscopy19,20 and histological image classification20. Dermoscopy images are acquired via a specialized microscope, and histopathological images are acquired via tissue biopsy and microscopy whereby both modalities yield highly standardized images. Photographs taken at different locations and using different equipment may vary in appearance and be difficult to use in a computer vision system. For example, antisymmetric images, such as lesions that are circular but not perfectly symmetric, can be difficult to classify. Moreover, the variability in the input images can be challenging to model and can lead to poor generalization. Therefore, we have established a dataset of 20,230 high-resolution images—two orders of magnitude larger than previous datasets5—consisting of 2,023 different diseases. We train a CNN using this dataset and evaluate its performance against 22 board-certified dermatologists. The CNN achieves performance parity with all tested experts across both tasks, demonstrating an artificial intelligence capable of diagnosing skin cancer with a level of competence comparable to dermatologists. Outfitted with deep neural networks, mobile devices can potentially extend the reach of dermatologists outside of the clinic. It is projected that 6 million smartphone subscriptions will exist by the year 2020 (ref. 12), and can therefore potentially provide low-cost and universal access to vital diagnostic care.

There are 1.6 million new cases of skin cancer in the United States every year. Out of the five Americans will be diagnosed with a cutaneous malignancy in their lifetime. Although melanomas represent fewer than 5% of all skin cancers in the United States, they account for approximately 70% of all skin cancer-related deaths, and are responsible for over 13,000 deaths annually in the United States alone. Early detection is critical, as the estimated 5-year survival rate for melanomas drops from 98% if detected in its earliest stages to about 14% if detected in its latest stages. We developed a computational method which allows medical practitioners and patients to proactively track skin lesions and detect cancer earlier. By creating a novel disease taxonomy, and a disease-purifying algorithm that maps individual images into defining classes, we are able to build a learning system for automated dermatology.

Current work in dermatology remains flawed because of a lack of the generalization capability of medical practitioners owing to insufficient data and a focus on standardized tasks such as dermoscopy19,20 and histological image classification20. Dermoscopy images are acquired via a specialized microscope, and histopathological images are acquired via tissue biopsy and microscopy whereby both modalities yield highly standardized images. Photographs taken at different locations and using different equipment may vary in appearance and be difficult to use in a computer vision system. For example, antisymmetric images, such as lesions that are circular but not perfectly symmetric, can be difficult to classify. Moreover, the variability in the input images can be challenging to model and can lead to poor generalization. Therefore, we have established a dataset of 20,230 high-resolution images—two orders of magnitude larger than previous datasets5—consisting of 2,023 different diseases. We train a CNN using this dataset and evaluate its performance against 22 board-certified dermatologists. The CNN achieves performance parity with all tested experts across both tasks, demonstrating an artificial intelligence capable of diagnosing skin cancer with a level of competence comparable to dermatologists. Outfitted with deep neural networks, mobile devices can potentially extend the reach of dermatologists outside of the clinic. It is projected that 6 million smartphone subscriptions will exist by the year 2020 (ref. 12), and can therefore potentially provide low-cost and universal access to vital diagnostic care.
Limitations of our study

- Retrospective
- Potential spectrum bias
- Study design → 2 disease categories
- Differences between in-person exam and telederm
  - We “blink, think, and compare” and use dermoscopy to help with diagnosis
- Opportunities for bias/confounding
  - Need to ensure extensive representation by varied skin types
  - Extensive additional work is needed
    - Prospective clinical validation studies
Limitations of Deep Learning

- System is **opaque**; we don’t know why it calls an image benign or malignant

- Dots, rulers, marks, etc. may introduce **bias** into dataset

- Investigators may not be aware of them or extent

- Impact on MD cognition/learning?
Strengths

Computer can assess image data imperceptible to human eye
- Only looking at specific lesions, not whole-body or sequential change detection

Broadly applicable across disciplines
- Algorithms will improve with transfer learning
- Edges, shapes remain important
- Show a skin cancer app cat pictures and it will improve at classifying skin cancer

Capable of running on a smartphone device
- App created to study this, prospective “real-time” validation in clinical setting
Further directions
What the network is “looking at”

Malignant Melanocytic Lesion

Benign Melanocytic Lesion

Inflammatory Condition

Malignant Epidermal Lesion

Benign Epidermal Lesion

Genodermatosis

Malignant Dermal Lesion

Benign Dermal Lesion

Cutaneous Lymphoma

Saliency maps for other skin conditions
Can AI be used to quantify and monitor skin disease severity?
Context matters

A dermatologist’s clinical impression is based on contextual factors beyond visual and dermoscopic examination of a lesion in isolation
Are we comfortable with AI as black box?

A trained neural net does not necessarily mimic the decision-making approach of humans--rather it identifies its own criteria for informative patterns associated with a disease.

Decisions need to be made based on machine output of probability, and the pt needs to managed.
Is AI a Pandora’s box?

- Probably not, and it won’t replace dermatologists!
- Roles in health care and education
- Real vs imaginary dangers:
  - “AI is a fundamental existential risk for human civilization.” (Elon Musk, 2017)
  - “AI software will help us understand biology, understand how to intervene and improve lives very dramatically.” (Bill Gates, 2018)
- Regulation and liability need to be addressed proactively
Clinical validation of app in progress
Can AI help dermatologists?

- 35 Y female with atypical mole syndrome, seven cutaneous melanomas (3 melanomas in situ, 3 T1a invasive melanomas and 1 T2a SLN-positive melanoma) and eight severely dysplastic nevi diagnosed since 2016 (negative for p16 mutation)

- 5 additional biopsies in 6/2017 showed 2 severely dysplastic nevi and an atypical compound Spitz tumor
A tool for surveillance? – NOT YET

New moderately dysplastic nevi with focal severe atypia
"Doctor, I have a suspicious looking mole on my shoulder."

Bottom line: We still need dermatologists!