Using AI to improve access and accuracy of information & care in dermatology

Yuan Liu, PhD
Tech Lead Manager, Google Research & Google Health

Google Health

Disclaimer: I am an employee of Google and own Alphabet stock

01 Motivation: gap in access and accuracy of dermatological care

Skin diseases are an enormous global burden and every day millions of people turn to Google to research their skin concerns

2 billion people affected with skin disease

Half the world's population faces a critical shortage of dermatologists

10 billions of annual skin condition queries on Google Search

But describing what you have is really challenging!

People **spend hours** searching the internet and talking to strangers on forums to find out what they have.

People **arrive at the correct condition only 13% of the time**, and **nearly 3-out-4 people** who needed urgent care did not realize they required it.

“I have a plant identification app. A skin app like that would be great. It would be **so much more convenient than googling.**”
- Female survey participant (25 - 45yo)

AI-powered dermatological assistive tool that helps users to research & identify their skin concerns

Discover | **Research** | **Decide**

What is this thing?
Users take pictures and input structured information

→

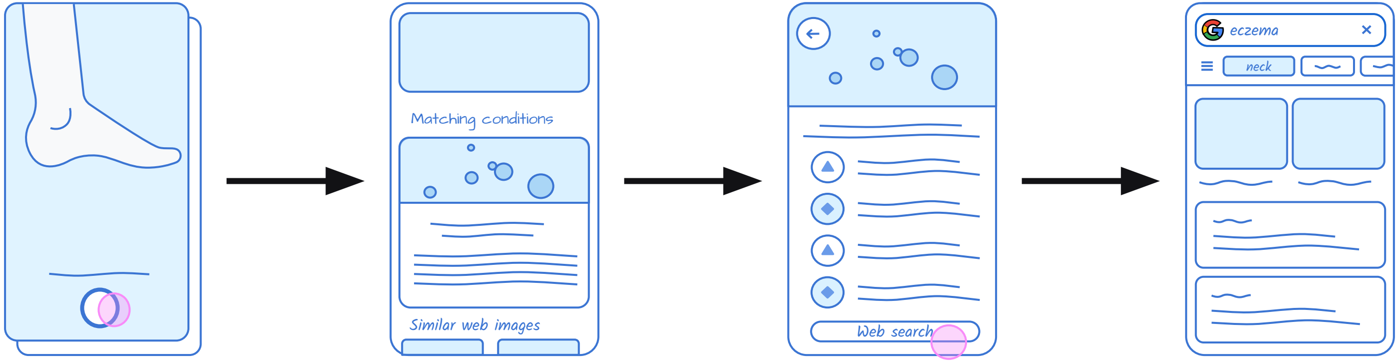
Try to find a match
Users review “matches” images and descriptions of the most salient features

→

Understand best “match”
Users learn more about condition of any potential “matches”

→

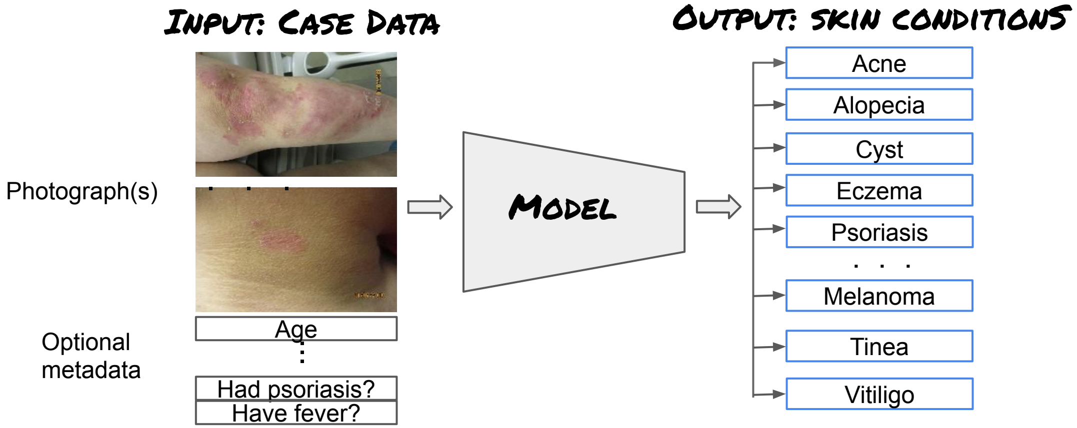
Decide what to do about it
Users can then decide on next steps or find appropriate care



02 Foundational research: the AI prototype & its clinical utility

Our overarching goal is to:

Develop an AI model to identify the most prevalent skin conditions from clinical images and metadata



Input Data: challenges in input variation

- A teledermatology dataset
  - 20k cases, 80k images
  - 17 sites, 2 US states
- Broad condition types coverage:
  - Lesion, rash, hair loss, nail infections, etc
- Different presentations per disease:
  - Skin type
  - Body part
  - Disease subtype / severity
- Image artifacts:
  - Lighting
  - Field-of-view
  - Background
- Metadata differences:
  - Missing / inconsistent metadata

Output Labels: several challenges with labeling

Expansive skin condition lexicon

SNOMED > 20k

Map subtypes/ synonyms to **a flat list of conditions**

No "absolute" ground truth

lack biopsy or definite diagnosis

Collect/predict **differential diagnosis**, a ranked list of diagnoses

Low inter-derm agreement

~60% agreement among individuals

Obtain reference diagnoses via **panels of certified dermatologists**

Highly skewed condition distribution

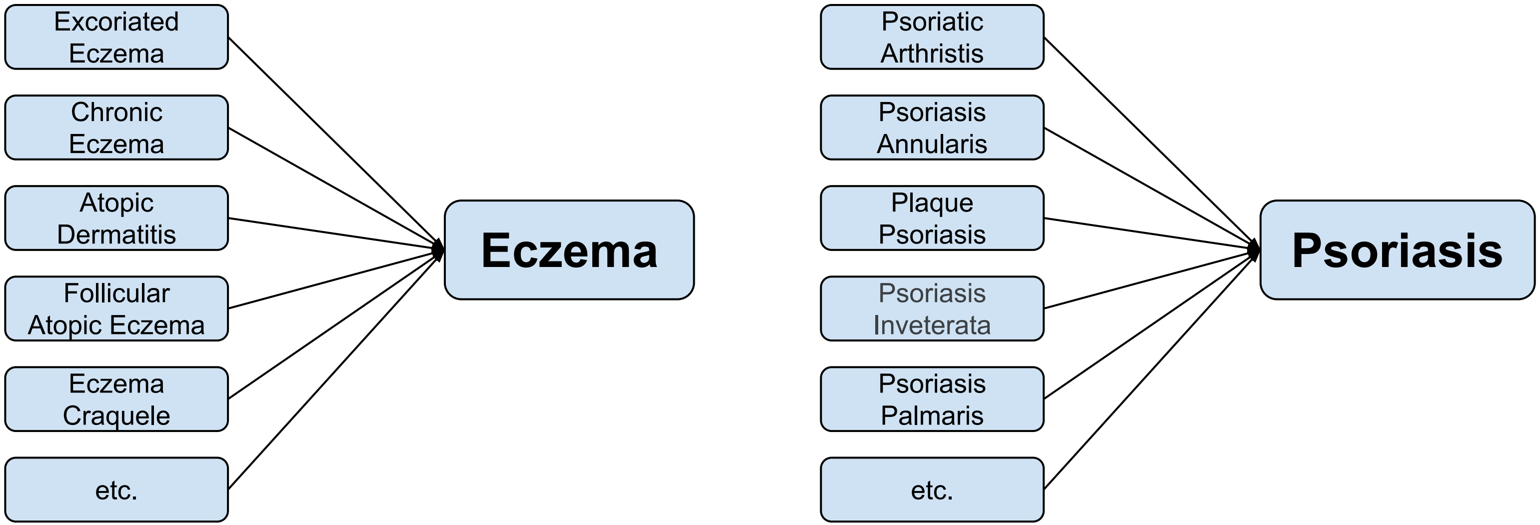
<5 examples in long tail

Support **26 conditions** most common in the dataset

Ground truth: complex labeling space

Skin condition lexicon:

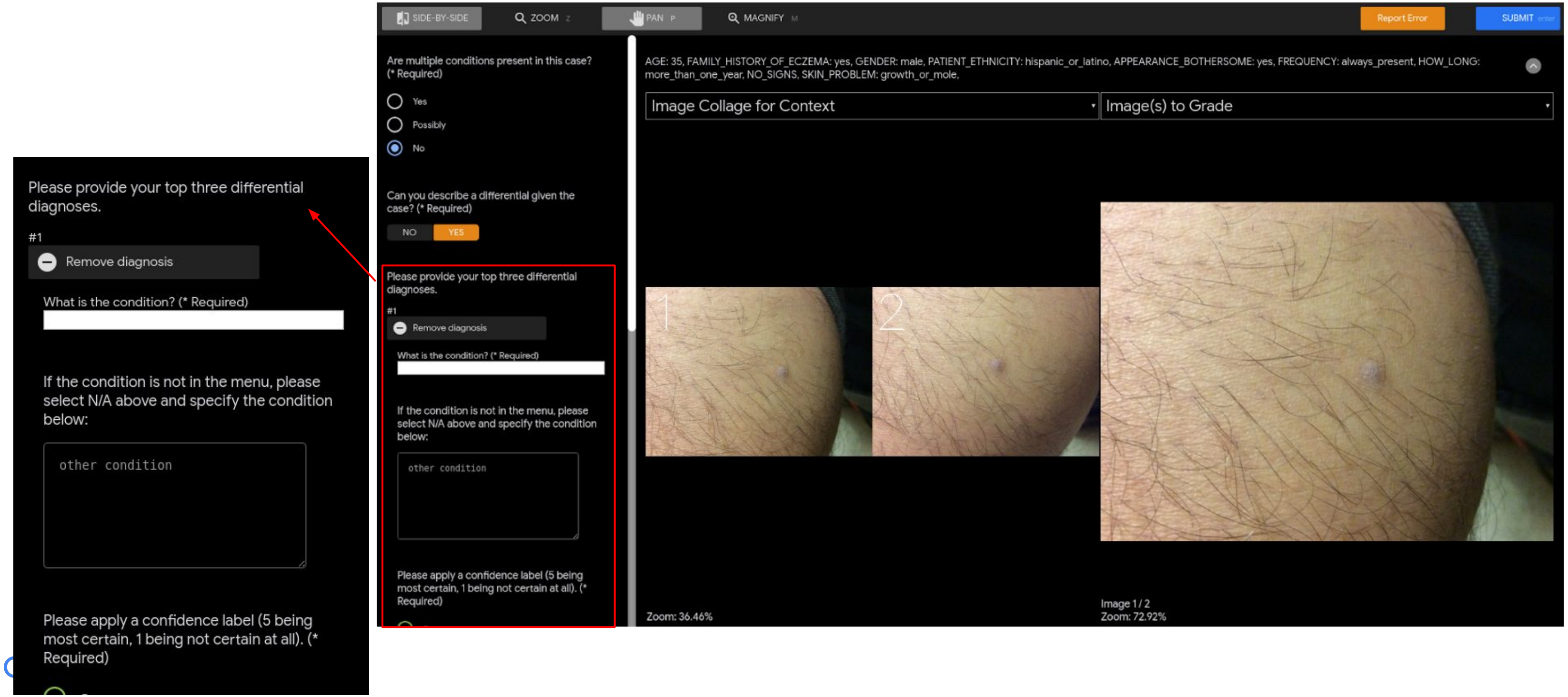
- > 20k SNOMED descriptions (5k ids) + > 1k free text entries
- Manually mapped subtypes and synonyms to a flat list of conditions



http://www.snomed.org/

Ground truth: differential diagnosis

No “absolute” ground truth: few cases have biopsy
Differential diagnosis: ranked list of diagnoses to determine clinical next steps



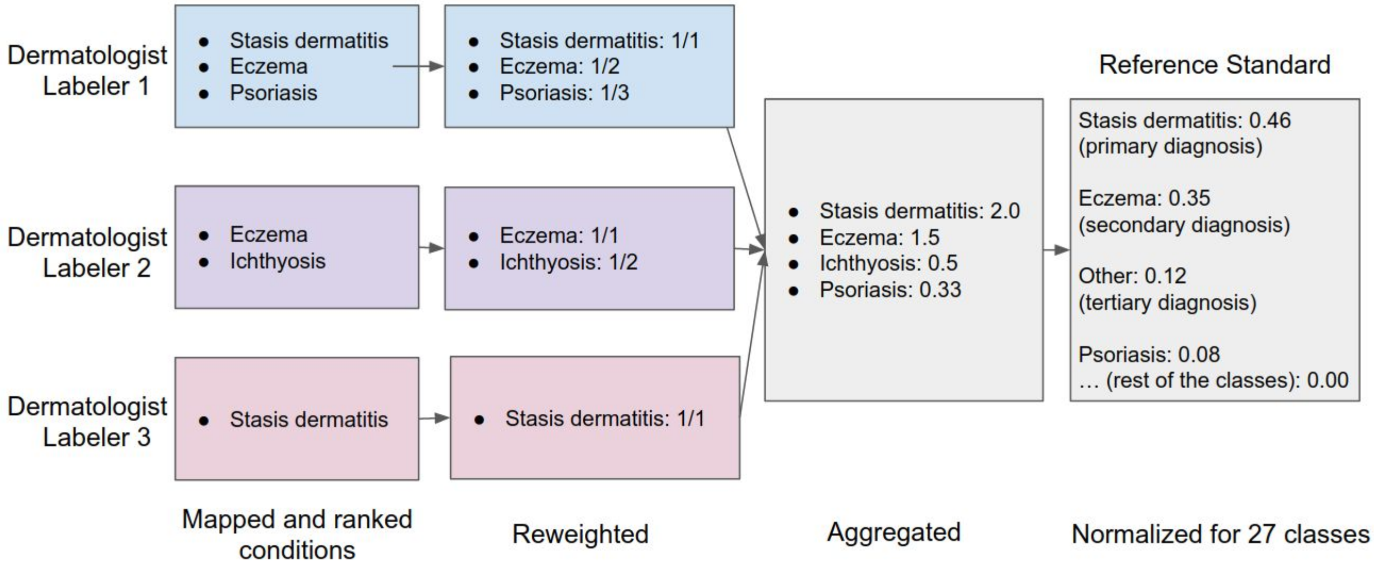
Ground truth: low inter-dermatologist agreement

Screen dermatologists via certification:

- Pass onboarding exams with score >= 0.70 top-3 agreement

Establish ground truth via collective intelligence:

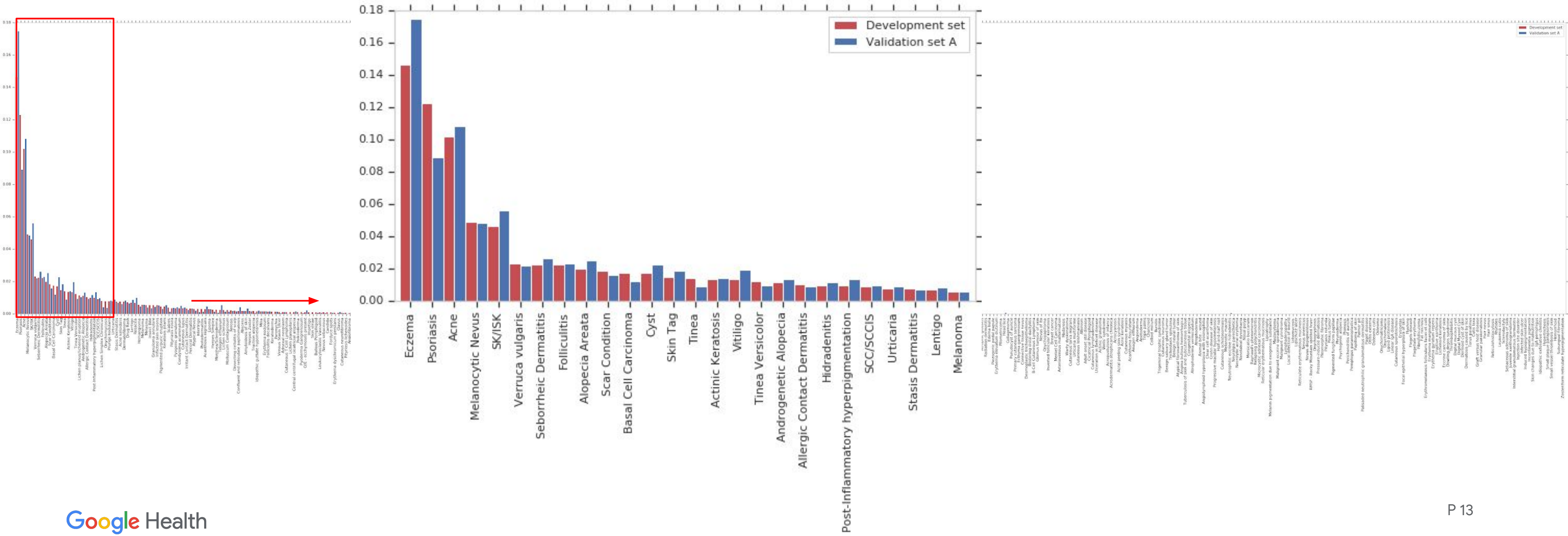
- “Position” weighted aggregation of individual dermatologist
- Reproducibility improves by 20%



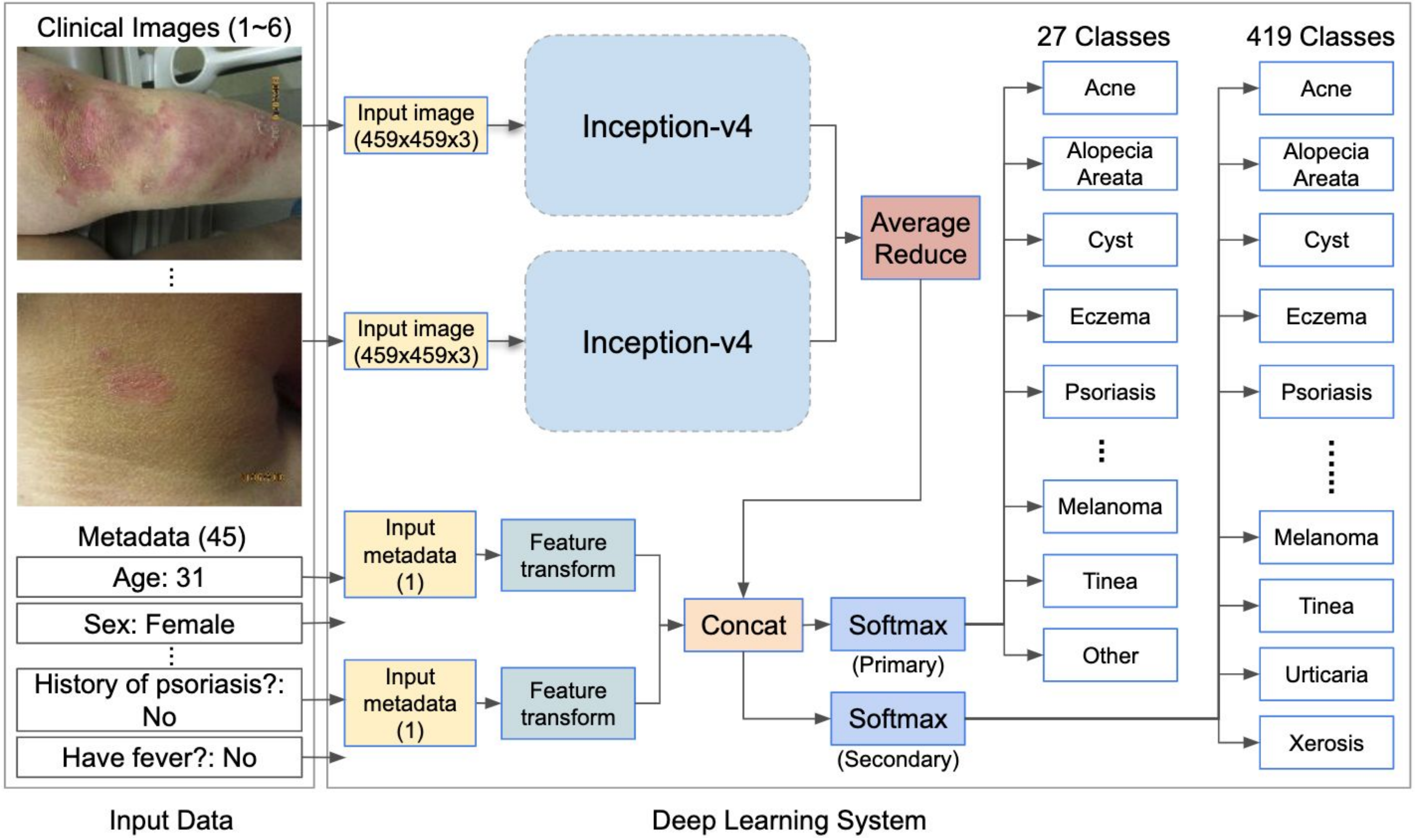
See more: Nature Medicine Paper

Ground truth: highly skewed distribution

Reduce from a flat list of parent conditions to 27 condition classes (26 + “Other”)
Still have the full flat list of conditions as a secondary prediction



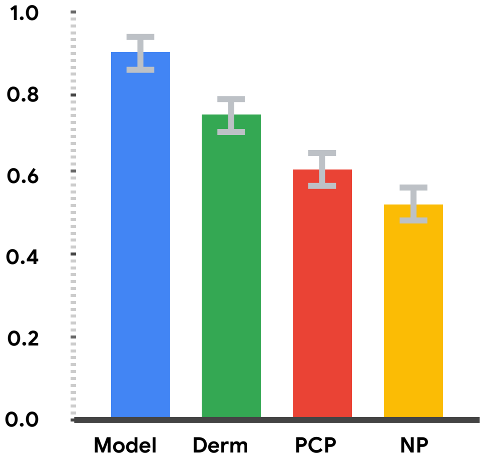
Model architecture: late fusion of images + metadata



AI model performs **comparably against tele-dermatologists**

Our model has **non-inferior diagnostic accuracy to dermatologists** across the most common 26 skin conditions,

with **top-3 accuracy** of 0.90, 0.75 (dermatologists), 0.60 (PCPs/GPs), and 0.55 (NPs)



Featured on the [cover of Nature](), June 2020 issue

Google Health

Liu et al, *Nature Medicine* 2020
https://ai.googleblog.com/2019/09/using-deep-learning-to-inform.html

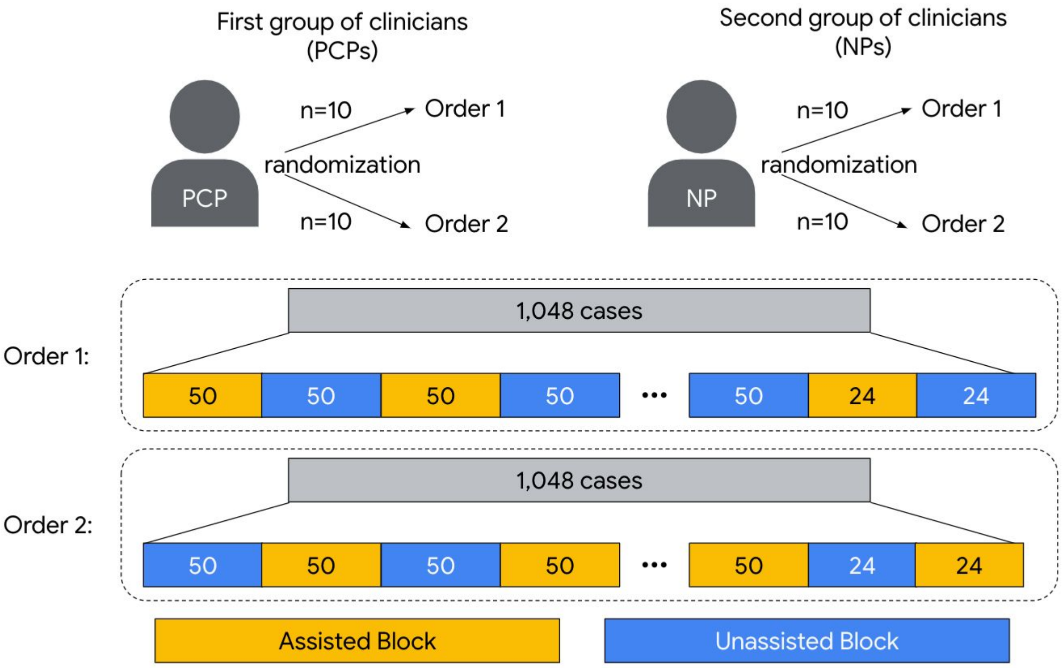
AI model can help NPs/PCPs better interpret skin cases

A multi-reader multi-case (MRMC) randomized study:

- 1,048 retrospective cases
- 120 conditions
- 40 clinicians (20 PCPs, 20 NPs)

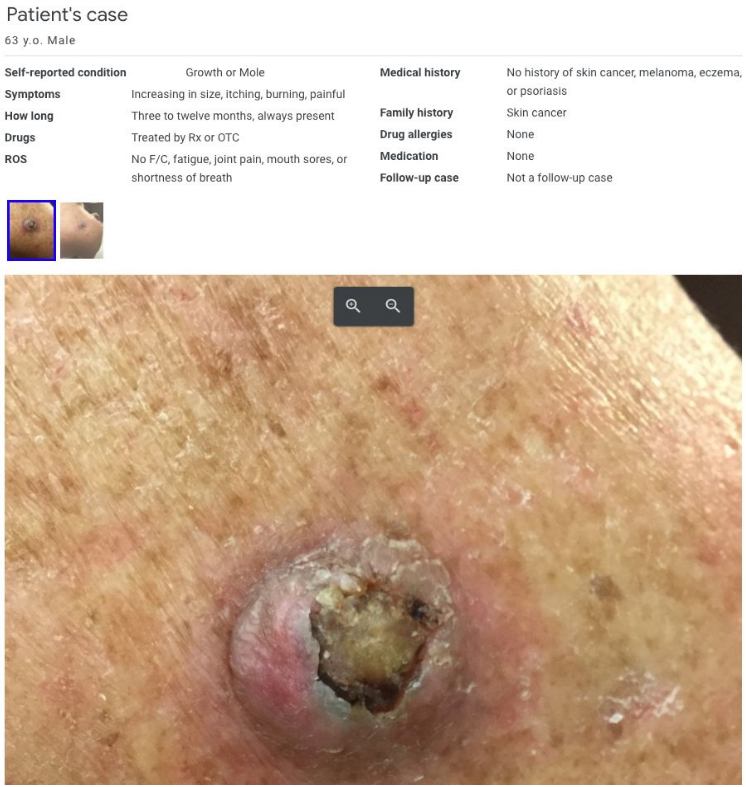
Each case was reviewed by half the clinicians with AI assistance and half without, only once per clinician.

For each clinician, the assistance modality alternated every 50 cases.

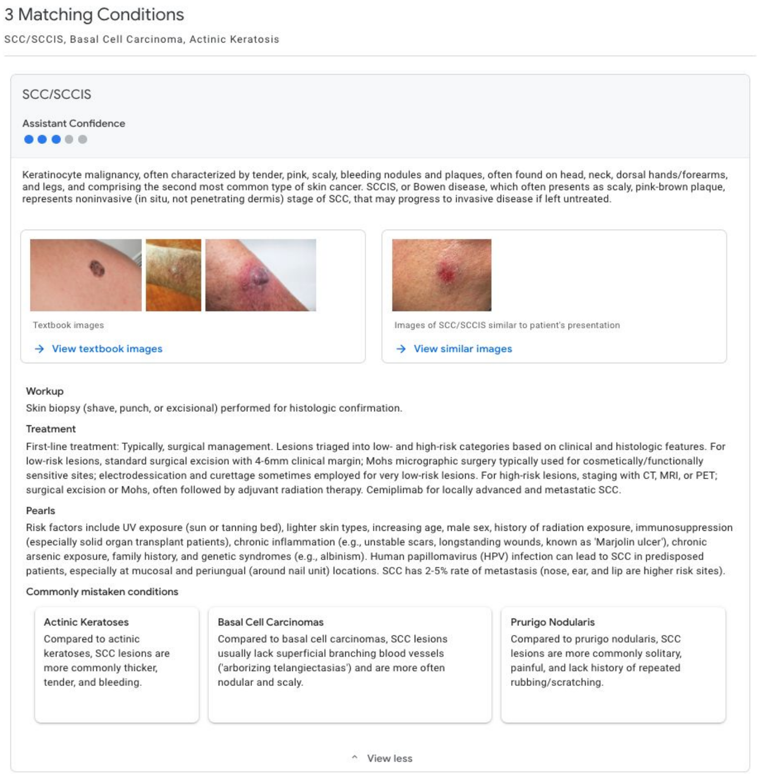


Jain et al, *JAMA Network Open* 2021
https://www.blog.google/technology/health/ai-assists-doctors-interpreting-skin-conditions/

AI model can help NPs/PCPs better interpret skin cases



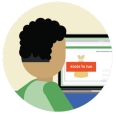
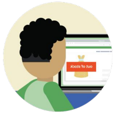
+



Jain et al, *JAMA Network Open* 2021
https://www.blog.google/technology/health/ai-assists-doctors-interpreting-skin-conditions/

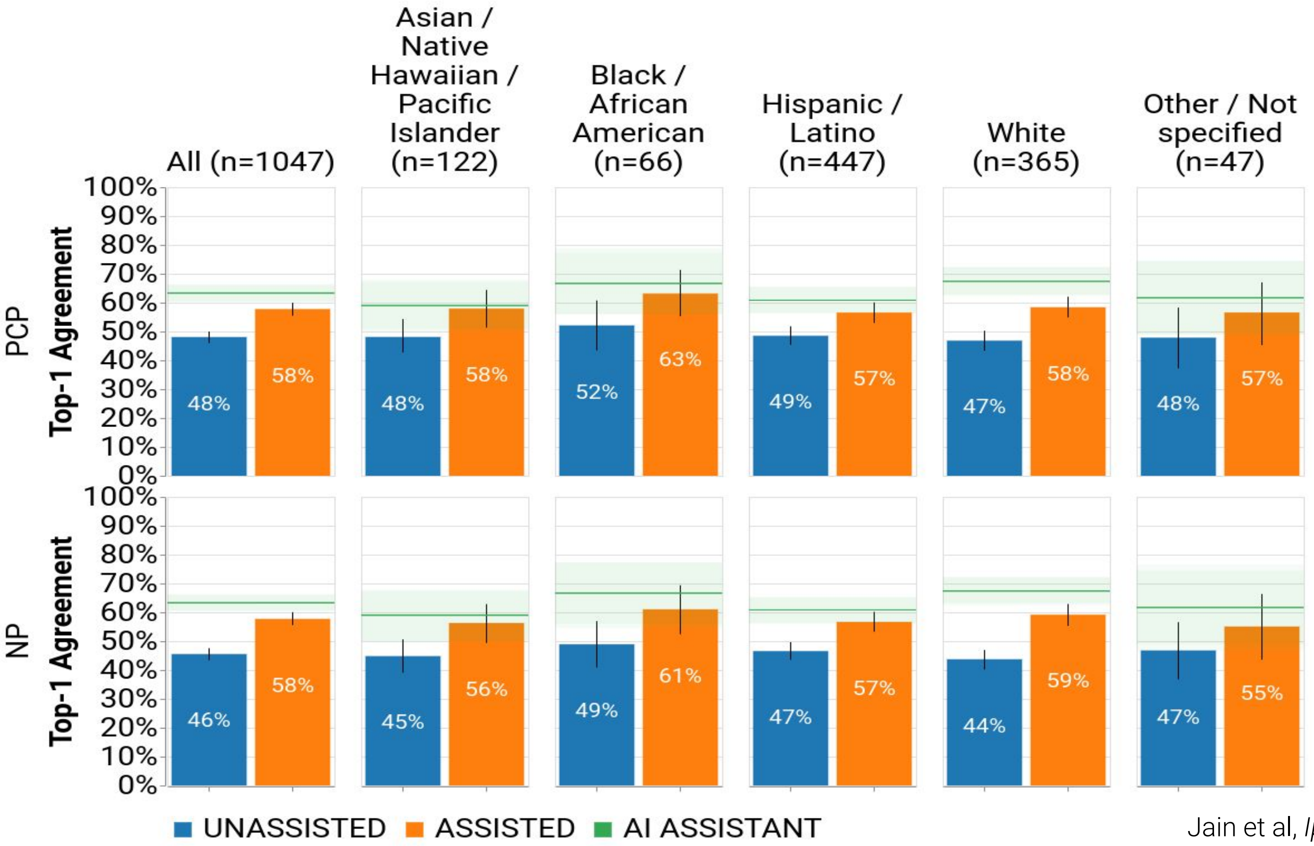
AI model can help NPs/PCPs better interpret skin cases

Non-specialist clinicians can identify the correct skin disease 20% more often and feel more confident about their assessment
No increase in likelihood to recommend biopsies or referrals to dermatologists

	Takeaway	Unassisted primary care clinician 	AI-assisted primary care clinician  AI
Primary analysis: diagnostic agreement with dermatologists	Significant increase with AI: $p < 0.001$	47%	58%
Classifying growths as benign, malignant or precancerous	Promising malignancy interpretation	62%	68%
Referrals to dermatologists	No increase	39%	36%
Desired rate of biopsy	No increase	25%	23%

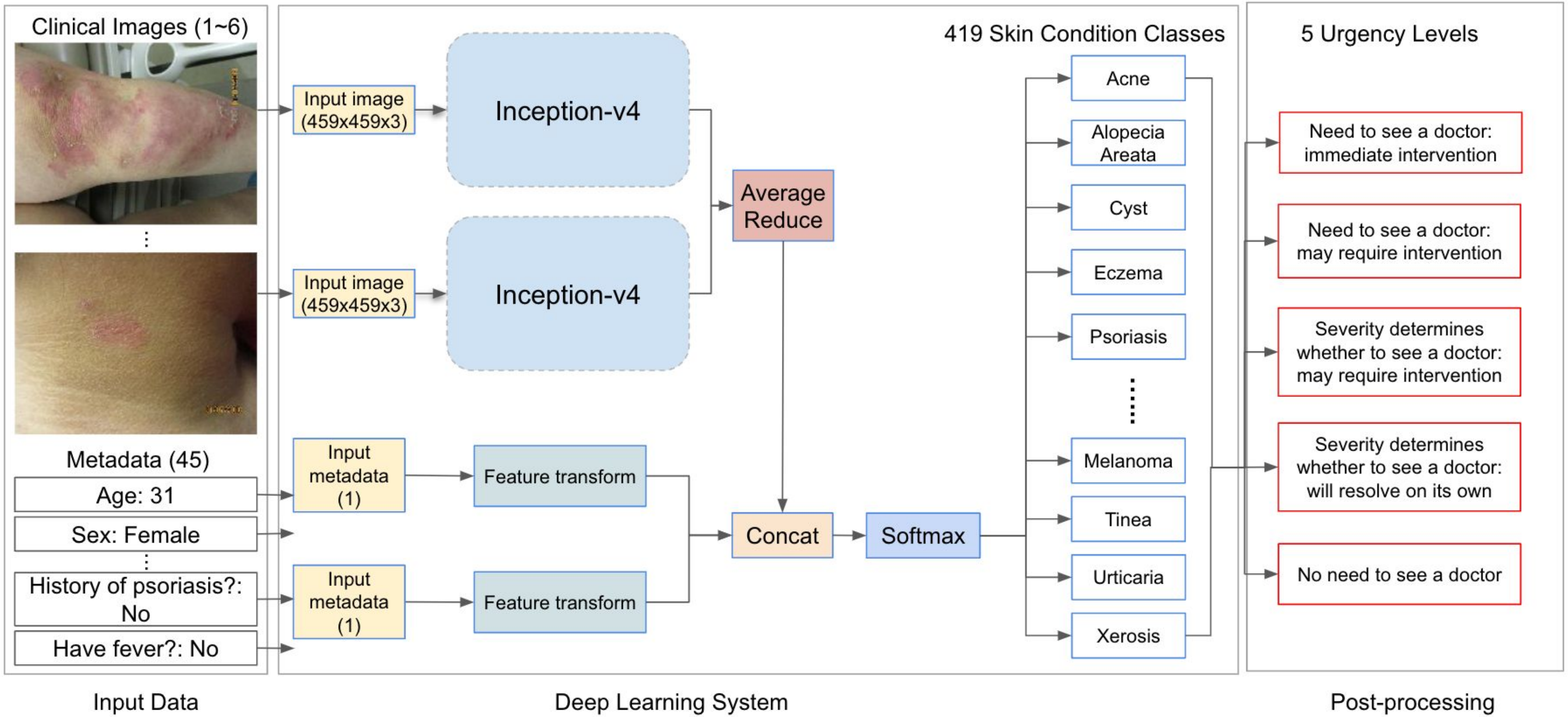
Jain et al, *JAMA Network Open* 2021
https://www.blog.google/technology/health/ai-assists-doctors-interpreting-skin-conditions/

This may help reduce existing disparities in dermatology



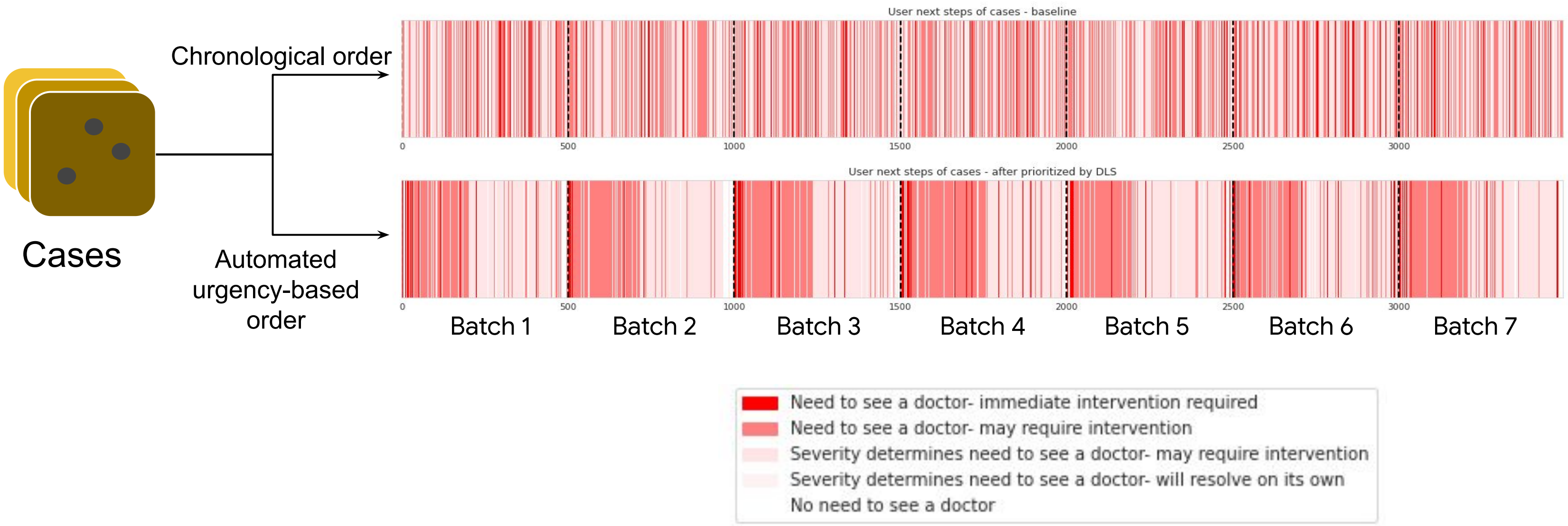
Jain et al, *Iproceedings* 2021

AI tools can also help **triage cases** in dermatology



Huang et al, *Skin Health and Disease* 2021

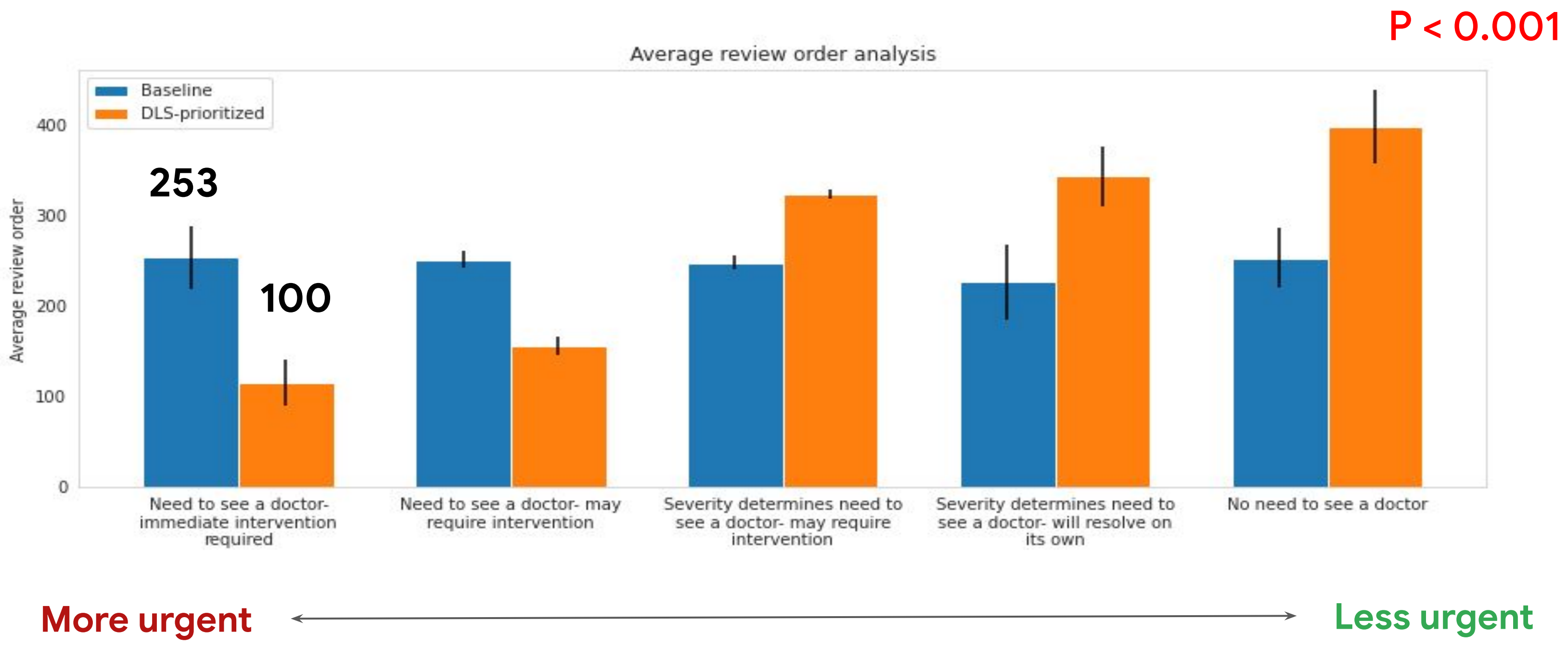
AI tools can also help triage cases in dermatology



Google Health

Huang et al, *Skin Health and Disease* 2021

AI tools can also help triage cases in dermatology



Huang et al, *Skin Health and Disease* 2021

03 Real-world translation: improvements to deploy at scale

Improving AI to be more accurate, generalizable, safe, and fair

1

Data sourcing

Acquire bigger, more diverse datasets

2

Efficient learning

Learn more efficiently from existing labeled data

3

Safety & Fairness

Ensure AI is safe and fair

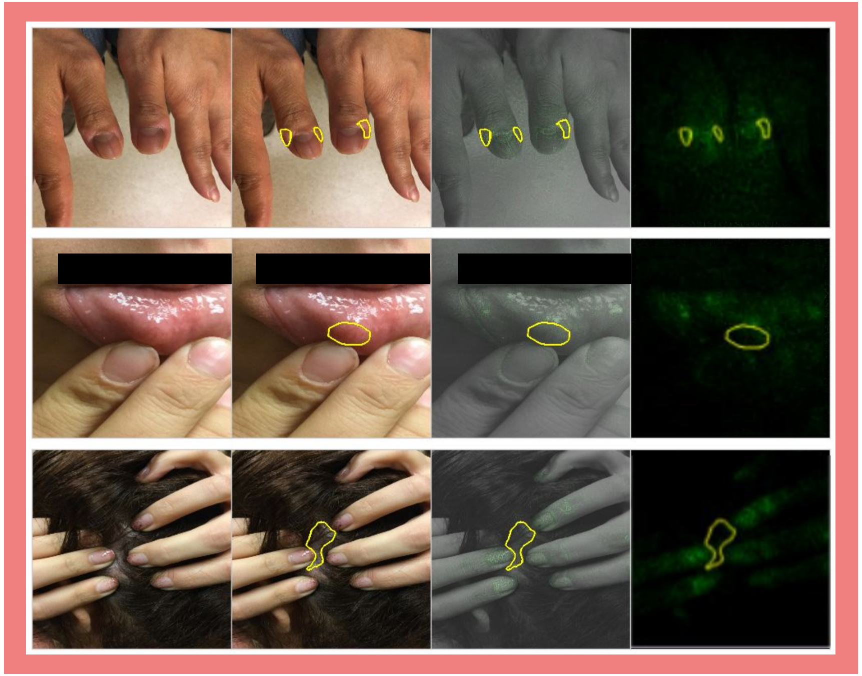
Obtain data from multiple sources

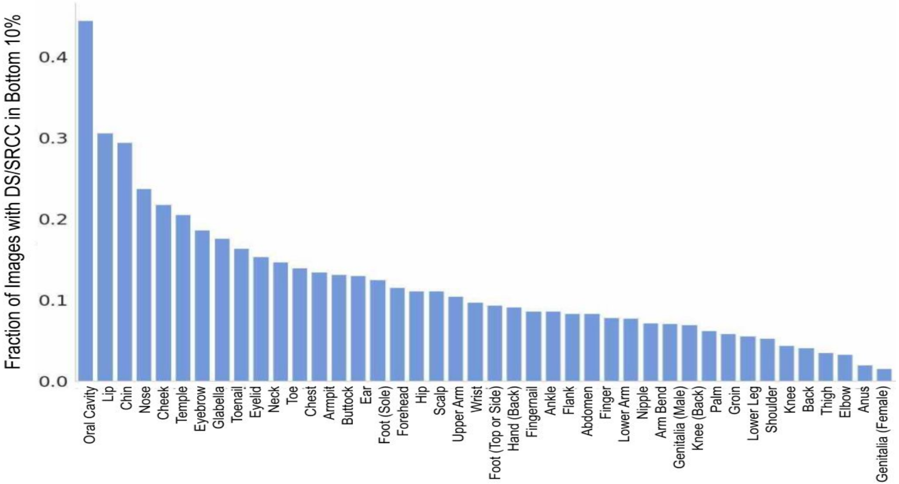
	Locations	Conditions	Demographics	Device Characteristics
A	Multiple outpatient sites throughout South America (rural and urban)	Non-urgent, common conditions treated in clinic or via telemedicine	Age mostly <=75, both male and female, diverse skin types	Clinic captured, unknown device
B	Academic outpatient & inpatient sites in Europe	Wide variety of conditions	Age mostly 25-85, both male and female, mostly lighter skin types	Clinic captured, digital camera
C	Multiple outpatient and retail sites in US (suburban, rural)	Non-urgent, common conditions referred to teledermatology	Age mostly 13-90+, both male and female, mostly lighter skin types	Assistant-captured, iPad + Canon; user-captured, various phones
D	Multiple outpatient sites throughout Australia (rural and urban)	Mostly malignancies, some benign	Age mostly 25-85, both male and female, mostly lighter skin types	Clinic captured, hand-held cameras
E	Multiple sites within US (urban)	Primarily healthy skin	Age mostly 20-50, both male and female, mostly light to brown skin types	User-captured, various phones
F	Major academic center in US (urban)	Wide variety of conditions	Age mostly 25-85, both male and female, diverse skin types	Clinic captured, unknown device (each uses their own)
G	Multiple sites within US (urban)	A wide variety of non skin images	N/A	User-captured, various phones

...... and more to come

Guide data collection via saliency analysis

Low agreement: Incorrectly Classified Images



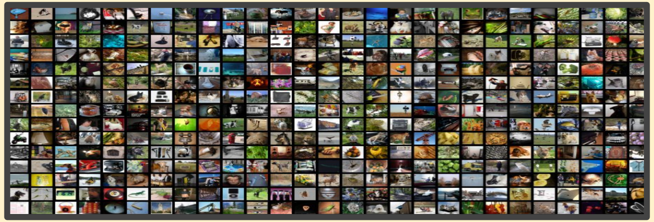


Singh et al, *CVPR ISIC* 2020

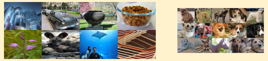
Leverage MORE NON-MEDICAL data with transfer learning

All the visual data on the internet - O(~100s of billions)

Large curated internet dataset 1 - O(300 million)



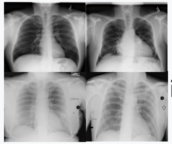
Moderate curated internet dataset 2 - O(1 million)



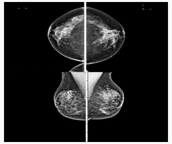
Big Transfer (BiT)

Labeled medical images

Skin condition diagnosis



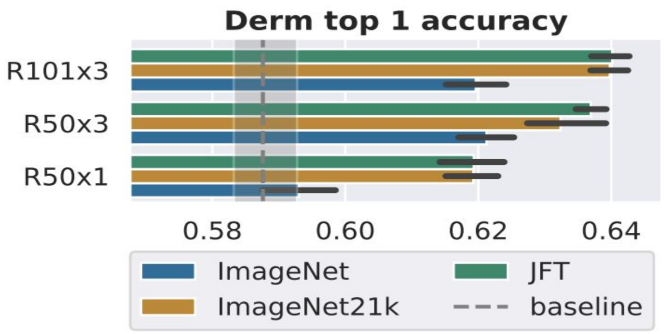
Chest X-Ray interpretation

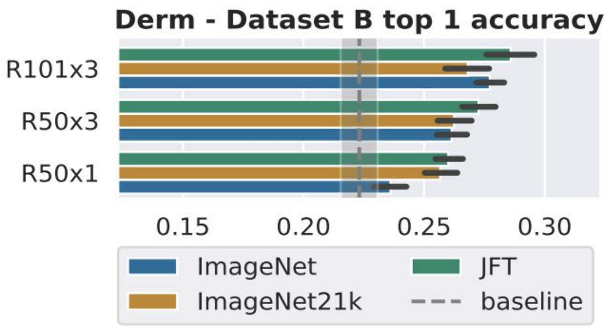


Breast cancer detection

Leverage MORE NON-MEDICAL data with transfer learning

Simple to use, drop in replacement for existing feature extraction backbones



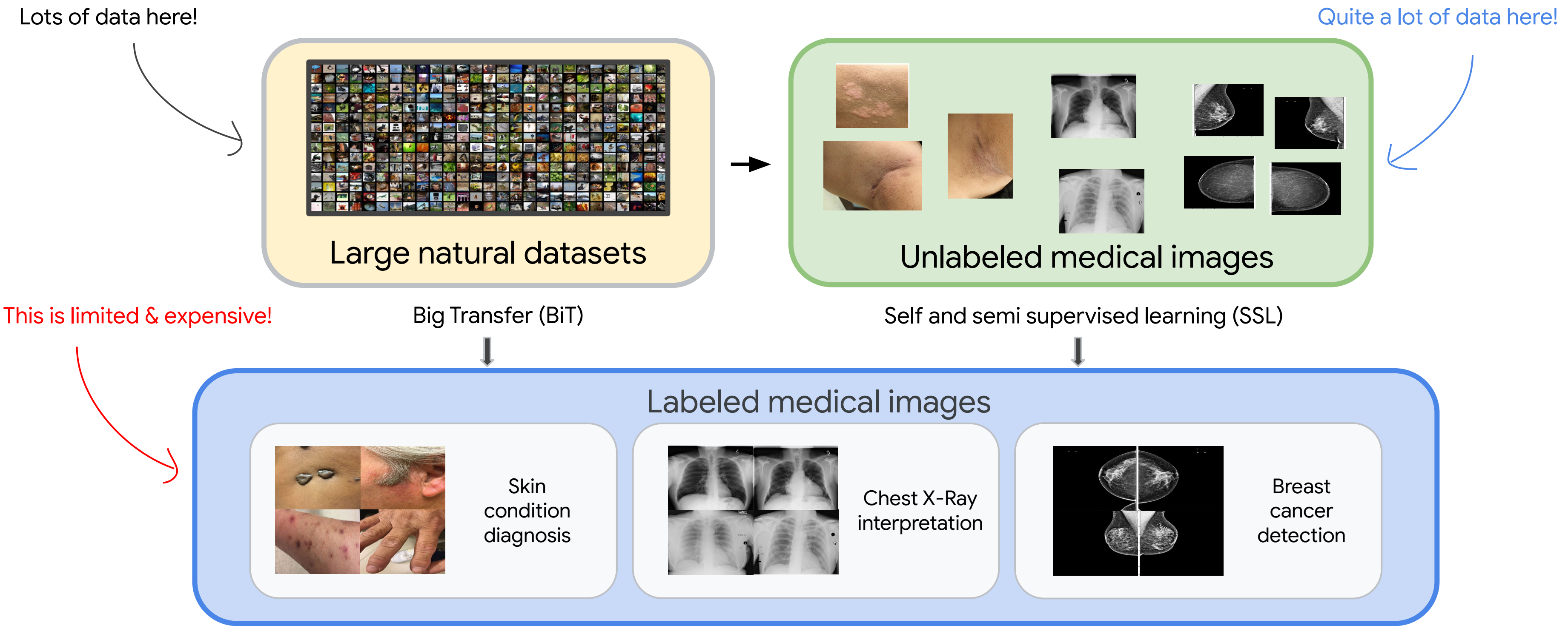


Significant improvement in task performance

BiT Models significantly more robust to distribution shifts!

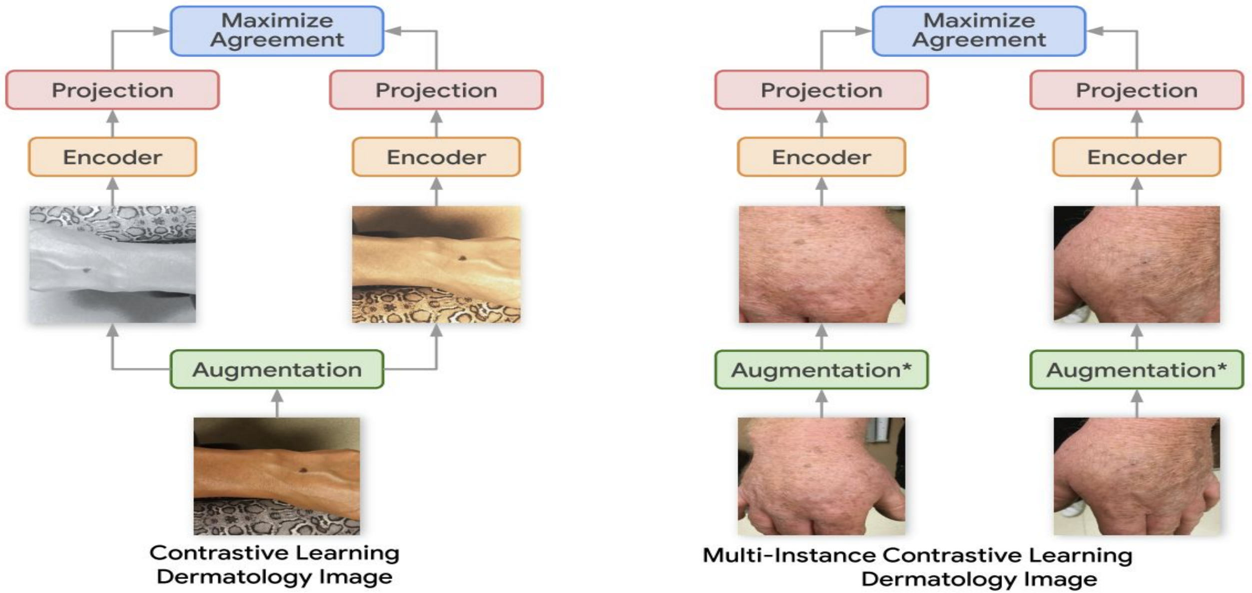
Mustafa et al, *arXiv* 2021

Leverage UNLABELED data with self-supervised learning



Leverage UNLABELED data with self-supervised learning

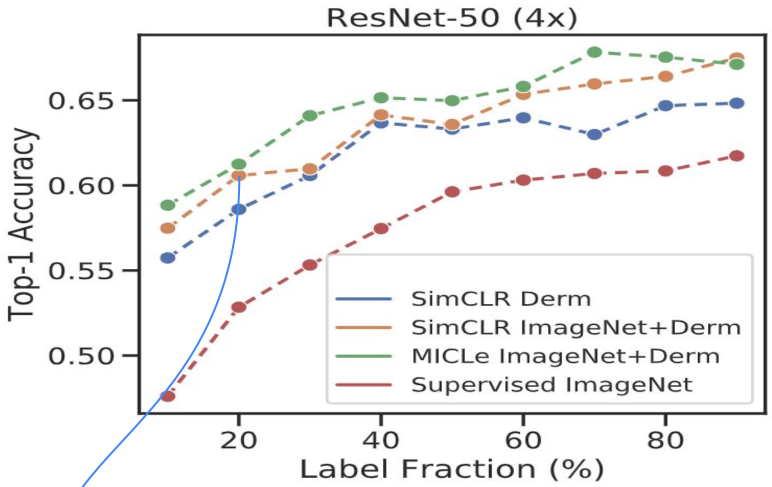
Maximizes agreement between augmented views of the same image, or images of the same case

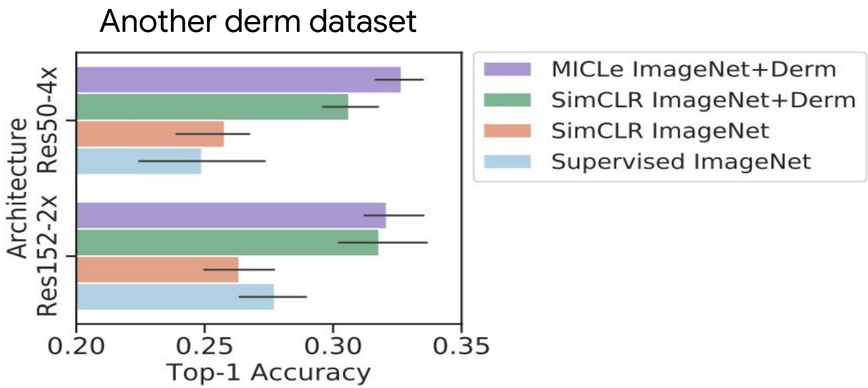


SimCLR and MICLe (multi-instance contrastive learning)

Aziz et al, *ICCV* 2021

Leverage UNLABELED data with self-supervised learning





Self supervised pre-training outperforms supervised pre-training

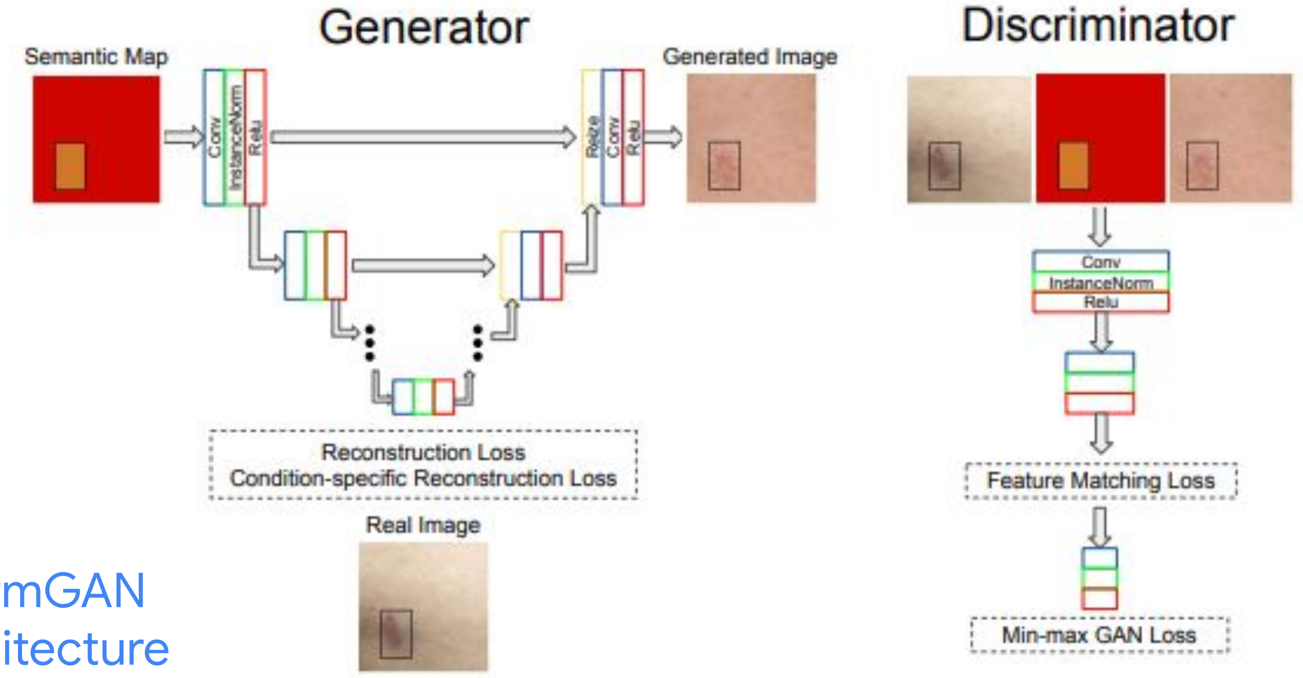
SSL pretrained models significantly more robust to distribution shifts!

MICLe can reach baseline performance using only 20% of the training data!!

Google Health

Aziz et al, *ICCV* 2021

Synthesize dermatology images using generative learning



DermGAN Architecture

Ghorbani et al, *NeurIPS ML4H* 2019
https://ai.googleblog.com/2020/02/generating-diverse-synthetic-medical.html

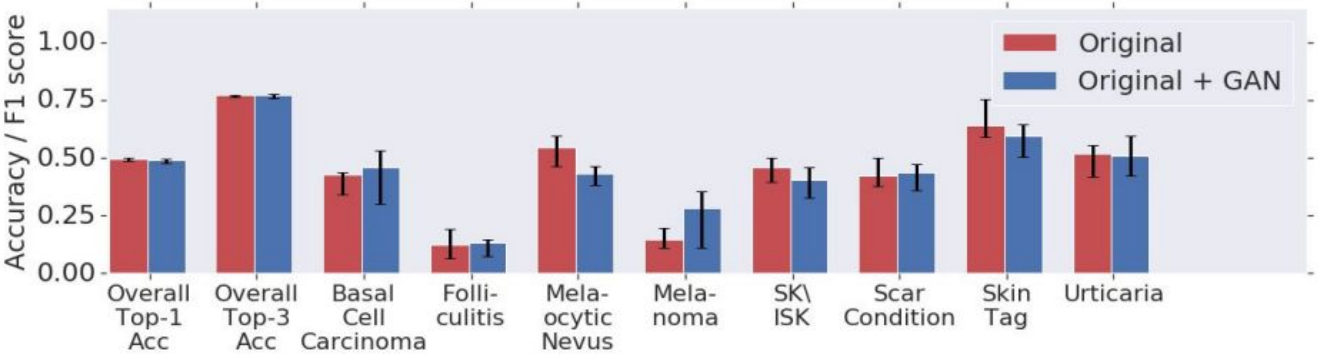
Synthesize dermatology images using generative learning

Fake

Real

Vary by skin type

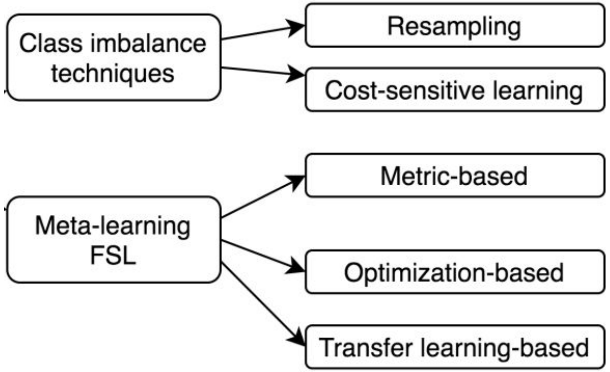
Vary by ROI size



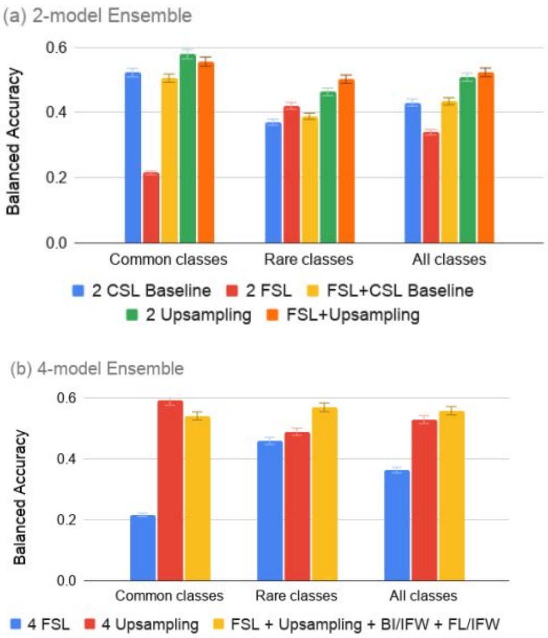
Ghorbani et al, *NeurIPS ML4H* 2019
https://ai.googleblog.com/2020/02/generating-diverse-synthetic-medical.html

Improve long-tail recognition with few-shot learning

Use an ensemble of few shot learning and conventional supervised learning models



- Upsampling (Chawla et al. 2012)
- Loss reweighting (Lin et al. 2017)
- Focal loss (Lin et al. 2017)
- Matching networks (Vinyals et al. 2016)
- Prototypical networks (Snell et al. 2017)
- Relation networks (Sung et al. 2018)
- MAML (Finn et al. 2017)
- Proto-MAML (Triantafillou et al. 2019)
- Baseline++ (Chen et al. 2019)



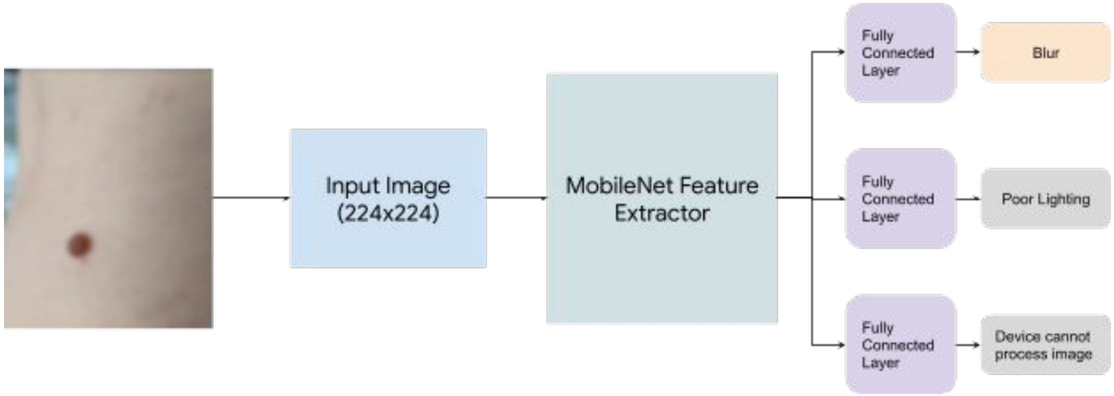
Weng et al, *NeurIPS ML4H* 2020

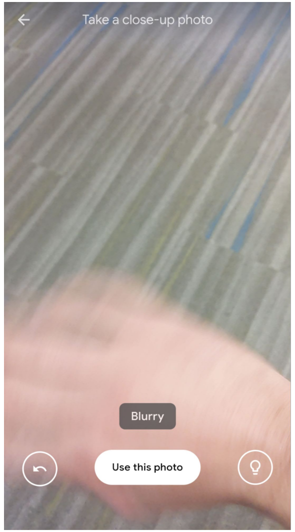
Guide intake to ensure sufficient input quality

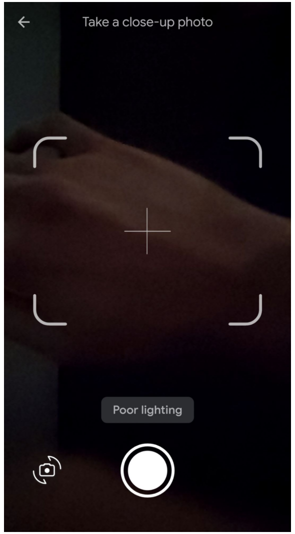
In teledermatology, up to 10% of images submitted by everyday users are so suboptimal that dermatologists cannot render a confident interpretation:

- Common quality reasons: **blurry** and **bad lighting**
- Real world adversarial use case: "**non-skin**"

Goal is to filter out **far out-of-distribution (OOD)** examples





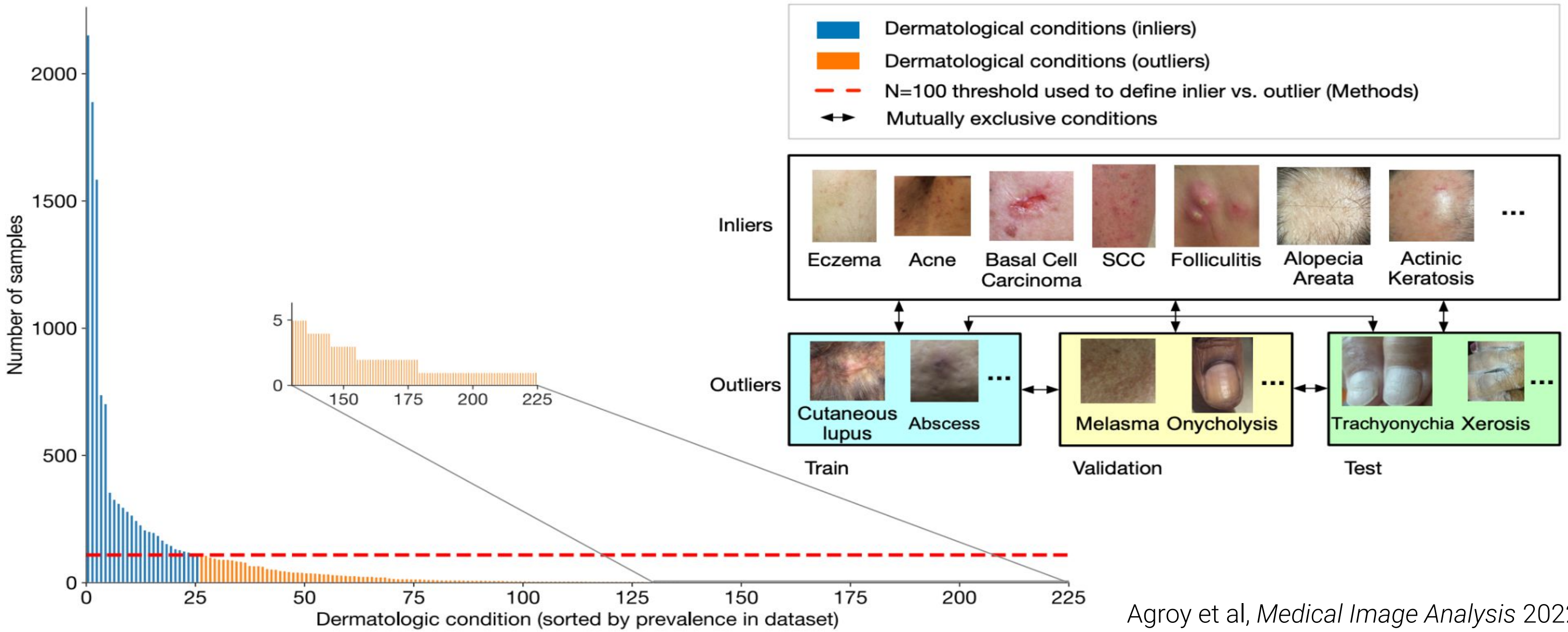


https://blog.tensorflow.org/2021/10/how-DermAssist-uses-TensorFlowJS.html

Teach AI to know when it doesn't know

Near OOD: many conditions in the long tail which AI hasn't encountered in training

Need to abstain from making predictions when encountering them in the wild

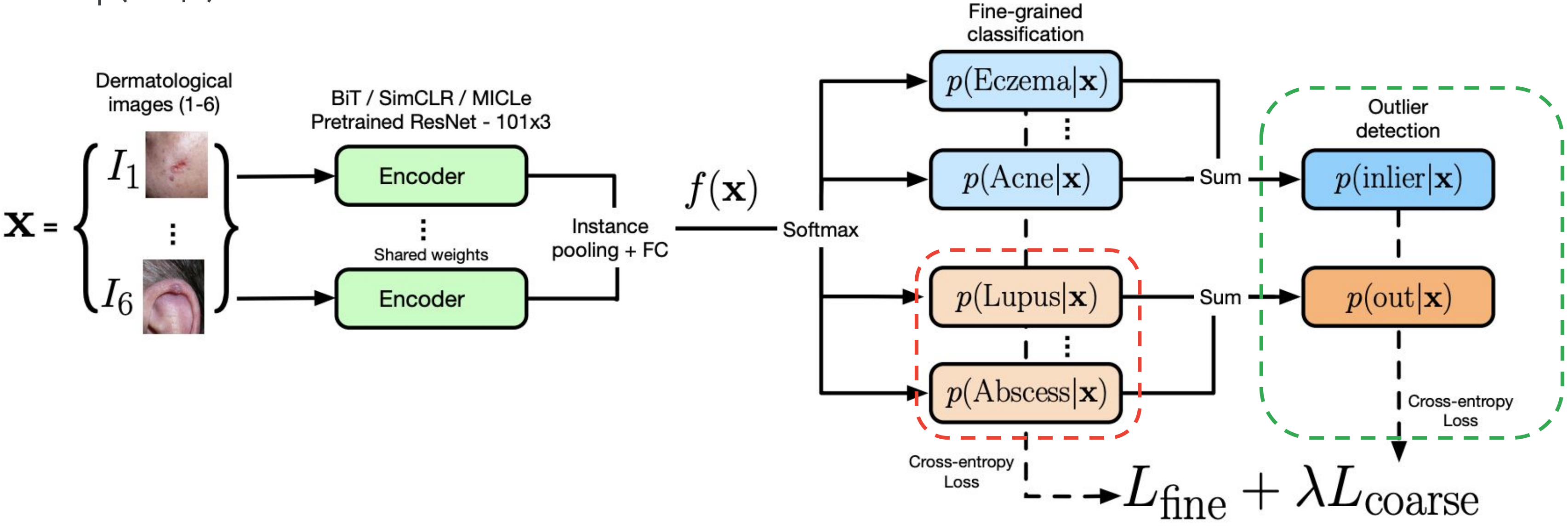


Agroy et al, *Medical Image Analysis* 2022

Teach AI to know when it doesn't know

Hierarchical Outlier Detection (HOD) Loss:

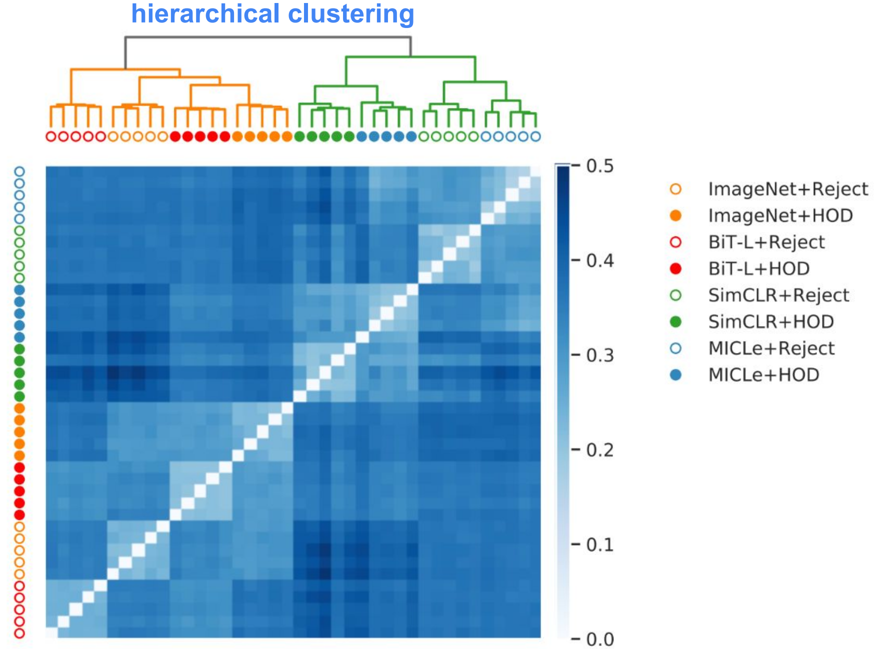
- Multiple Abstention classes (Expanded fine-grained training outlier classes)
- High-level coarse inlier vs. outlier loss
- p(out|**x**) is used as the OOD score

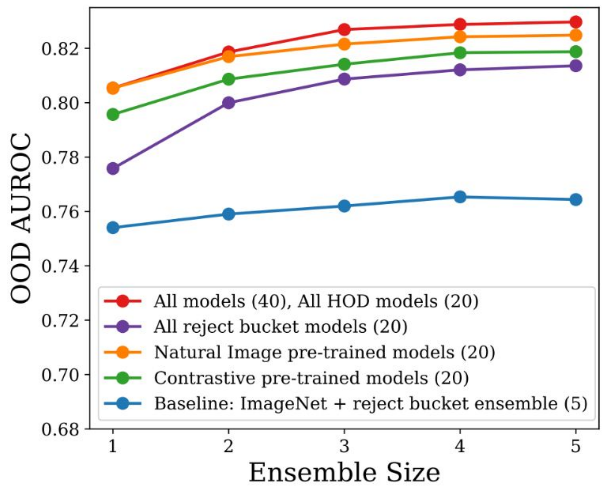


Agroy et al, *Medical Image Analysis* 2022

Teach AI to know when it doesn't know

Complementarity in learnt representations -> more diverse ensembles -> better predictive uncertainty quantification





Agroy et al, *Medical Image Analysis* 2022

Evaluate fairness under distribution shift

Fairness considerations need to be built into the entire process: problem definition, data collection, algorithm development, and post-deployment evaluation

A largely fair model may exhibit disparities in performance when deployed in the real world

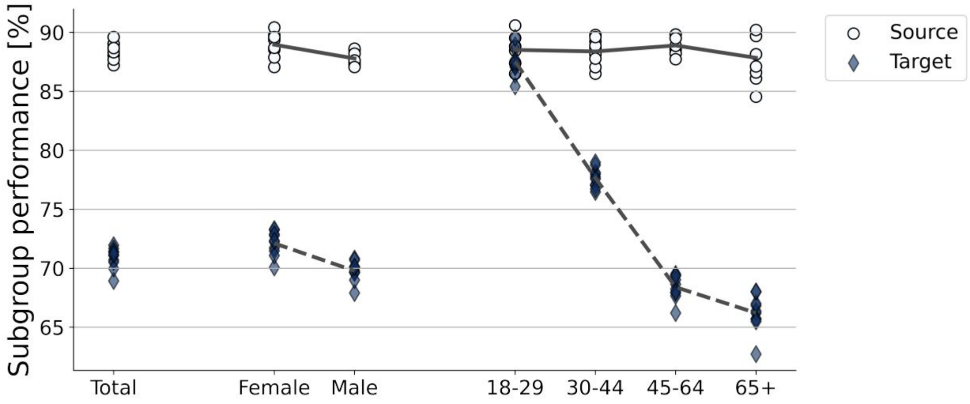


Fig. 4. Model performance in dermatology, as estimated via Top-3 accuracy (in %). The plot displays the total performance, as well as performance stratified by sex and by age on the source (circles with plain line) and target (diamonds with dashed line) data. Each marker represents one replicate of the model.

Shrouff et al, *NeurIPS* 2022

04 Where we are

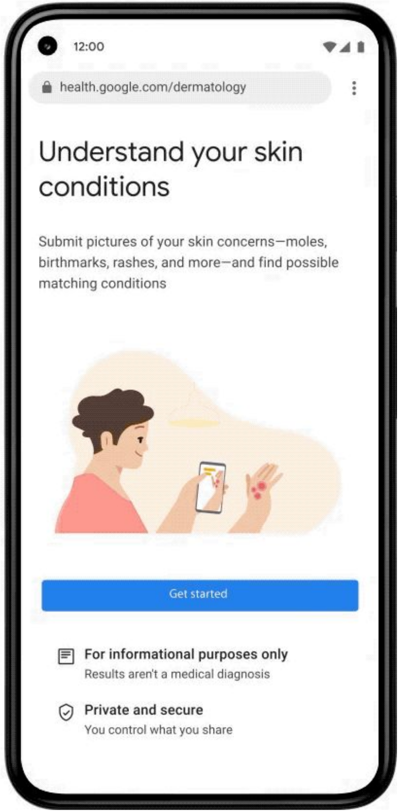
DermAssist and beyond

We have developed DermAssist, an AI-powered, dermatology assistive tool, with the AI model covering up to **288 conditions** by training the model on millions more images and using advanced AI technologies.

This informational tool has obtained **CE Mark** as a **Class I medical device in the EU**.*

We continue to learn how best we can leverage AI to improve the world's access to accurate dermatological information and care.

*This product has not been evaluated by the U.S. FDA for safety or efficacy.



https://www.blog.google/technology/health/ai-dermatology-preview-io-2021/

A list of publications

Journals:

- Huang, S., et al., Machine learning for clinical operations improvement via case triaging. *Skin Health and Disease* (2021).
- Guha Roy, A., et al., Does your dermatology classifier know what it doesn't know? Detecting the long-tail of unseen conditions. *Med. Image Analysis* (2021).
- Jain, A., et al., Development and Assessment of an Artificial Intelligence–Based Tool for Skin Condition Diagnosis by Primary Care Physicians and Nurse Practitioners in Teledermatology Practices. *JAMA Netw Open* (2021).
- Liu, Y., et al., A deep learning system for differential diagnosis of skin diseases. *Nat. Med.* (2020).
- D'Amour, A., et al., Underspecification Presents Challenges for Credibility in Modern Machine Learning. JMLR. arXiv [cs.LG] (2020).
- Eng, C., Liu, Y. & Bhatnagar, R. Measuring clinician-machine agreement in differential diagnoses for dermatology. *Br. J. Dermatol.* (2019).

Conferences:

- Schrouff, J., et al., Maintaining fairness across distribution shift: do we have viable solutions for real-world applications?. *NeurIPS*. arXiv [cs.LG] (2022).
- Jain, A., et al., Race- and Ethnicity-Stratified Analysis of an Artificial Intelligence–Based Tool for Skin Condition Diagnosis by Primary Care Physicians and Nurse Practitioners. *Iproceedings* (2022).
- Azizi, S., et al., Big Self-Supervised Models Advance Medical Image Classification. *ICCV* (2021).
- Mustafa, B., et al., Supervised Transfer Learning at Scale for Medical Imaging. arXiv [cs.CV] (2021).
- Weng, W.-H., et al., Addressing the Real-world Class Imbalance Problem in Dermatology. *Machine Learning for Health NeurIPS Workshop (ML4H)* (2020).
- Singh, N., et al., Agreement Between Saliency Maps and Human-Labeled Regions of Interest: Applications to Skin Disease Classification. *Skin Image Analysis CVPR Workshop (ISIC)* (2020).
- Ghorbani, A., et al., DermGAN: Synthetic Generation of Clinical Skin Images with Pathology. *Machine Learning for Health NeurIPS Workshop (ML4H)*, (2019).

Blog posts:

- "Ask a Techspert: What does AI do when it doesn't know?" by Iz Conroy | Google Keyword Blog | 08-Feb-2022
- "Does Your Medical Image Classifier Know What It Doesn't Know?" by Abhijit Guha Roy & Jie Ren | Google AI Blog | 27-Jan-2022
- "How Underspecification Presents Challenges for Machine Learning" by Alex D'Amour and Katherine Heller | Google AI Blog | 18-Oct-2021
- "Self-Supervised Learning Advances Medical Image Classification" by Shekoofeh Azizi | Google AI Blog | 13-Oct-2021
- "How DermAssist uses TensorFlow.js for on-device image quality checks" by Miles Hutson & Aaron Loh | TensorFlow Blog | 11-Oct-2021
- "Using AI to help find answers to common skin conditions" by Peggy Bui & Yuan Liu | Google Keyword Blog | 18-May-2021
- "AI assists doctors in interpreting skin conditions" by Ayush Jain & Peggy Bui | Google Keyword Blog | 28-Apr-2021
- "Generating Diverse Synthetic Medical Image Data for Training Machine Learning Models" by Timo Kohlberger & Yuan Liu | Google AI Blog | 19-Feb-2020
- "Using Deep Learning to Inform Differential Diagnoses of Skin Diseases" by Yuan Liu & Peggy Bui | Google AI Blog | 12-Sep-2019

Thank you

For more, please reach out to:
yuanliu@google.com or
dermatology-research@google.com