Evaluating DNNs in Dermatology with the Fitzpatrick 17k dataset

Matt Groh, Caleb Harris, Luis Soenksen, Felix Lau, Rachel Han, Aerin Kim, Arash Koochek, Omar Badri

Sixth ISIC Skin Image Analysis Workshop
CVPR 2021 Virtual
June 19, 2021

The New York Times

Accou

Dermatology Has a Problem With Skin Color

Common conditions often manifest differently on dark skin. Yet physicians are trained mostly to diagnose them on white skin.

CONTRIBUTE

STAT

TOPICS OPINION TEAM EVENTS NEWSLETTERS REPORTS VIDEO PODCA

TRENDING: CORONAVIRUS FIRST OPINION BIOTECH HEALTH

HEALTH

Dermatology faces a reckoning: Lack of darker skin in textbooks and journals harms care for patients of color

By Usha Lee McFarling July 21, 2020

Reprints

Roadmap for today's talk

- Motivation – what's at stake for skin image analysis?
- The Fitzpatrick 17k dataset –how can we characterize the data?
- Evaluating Training Deep Neural Networks
- Comparing Fitzpatrick labels with ITA
- Discussion

Lack of Publicly Available Datasets with Skin Type Labels

Derm 7 pt ❌

Dermofit Image Library ❌

ISIC 2018, ISIC 2019, ISIC 2020 ❌

MED-NODE ❌

PH2 ❌

SD-128, SD-198, SD-260 ❌

Lack of Publicly Available Datasets with Skin Type Labels

Derm 7 pt ❌

Dermofit Image Library ❌

ISIC 2018, ISIC 2019, ISIC 2020 ❌

MED-NODE ❌

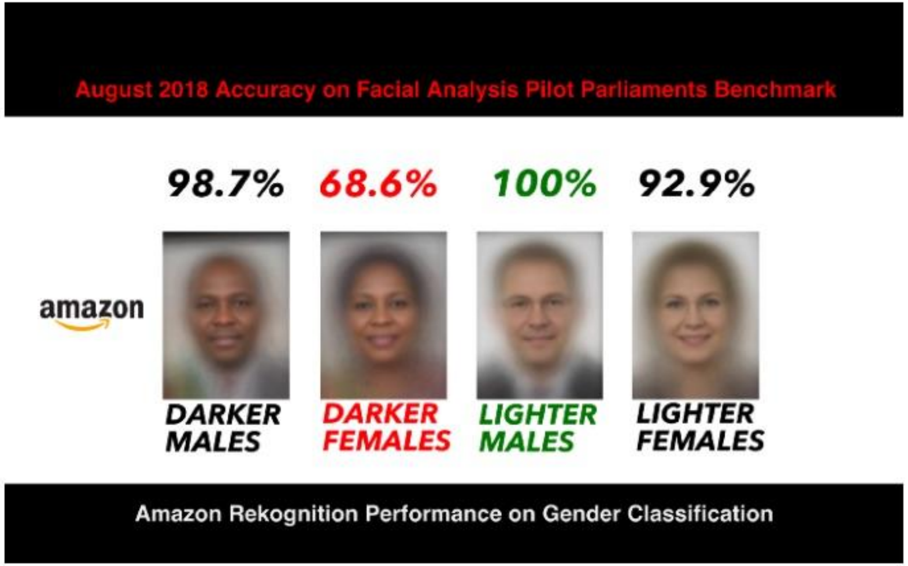
PH2 ❌

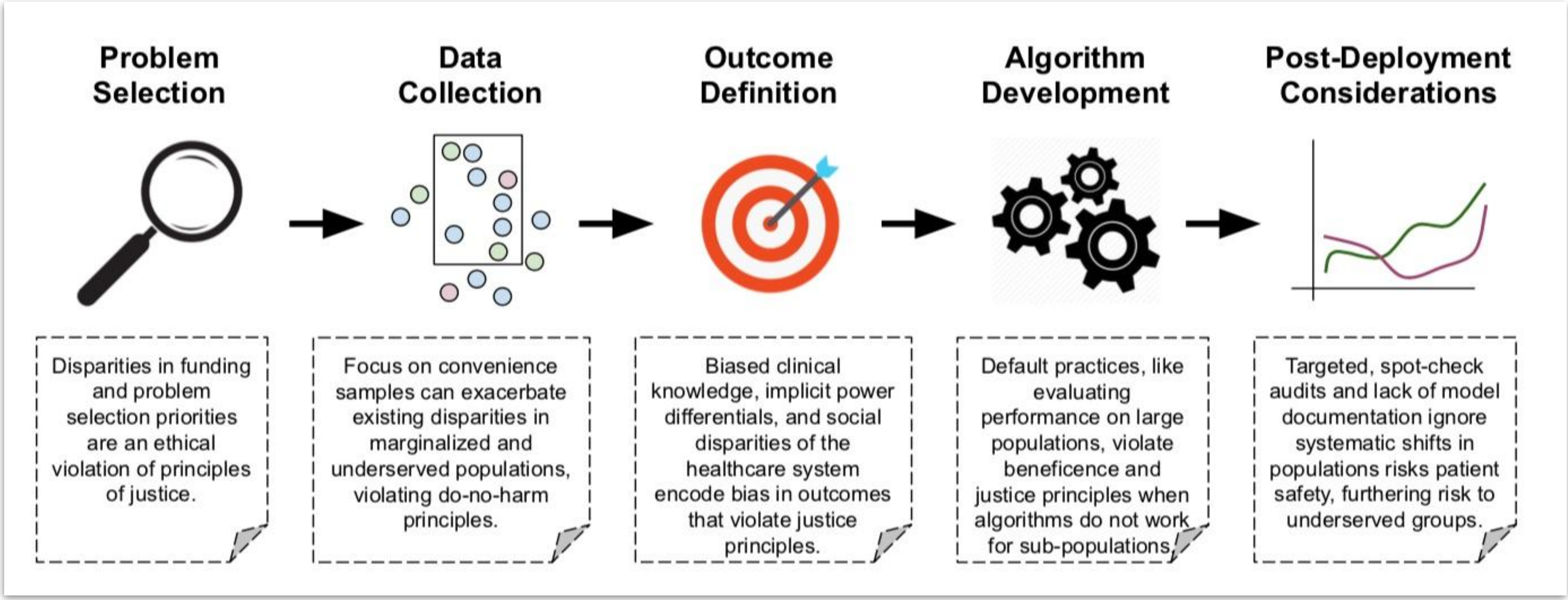
SD-128, SD-198, SD-260 ❌

PAD-UFES-20 ✅ (579/1,373 patients have data on Fitzpatrick skin type)

Response: Racial and Gender bias in Amazon Rekognition — Commercial AI System for Analyzing Faces.

Joy Buolamwini Jan 25, 2019 · 15 min read





Citation: Chen et al 2020 Ethical Machine Learning in Health Care

Fitzpatrick 17k

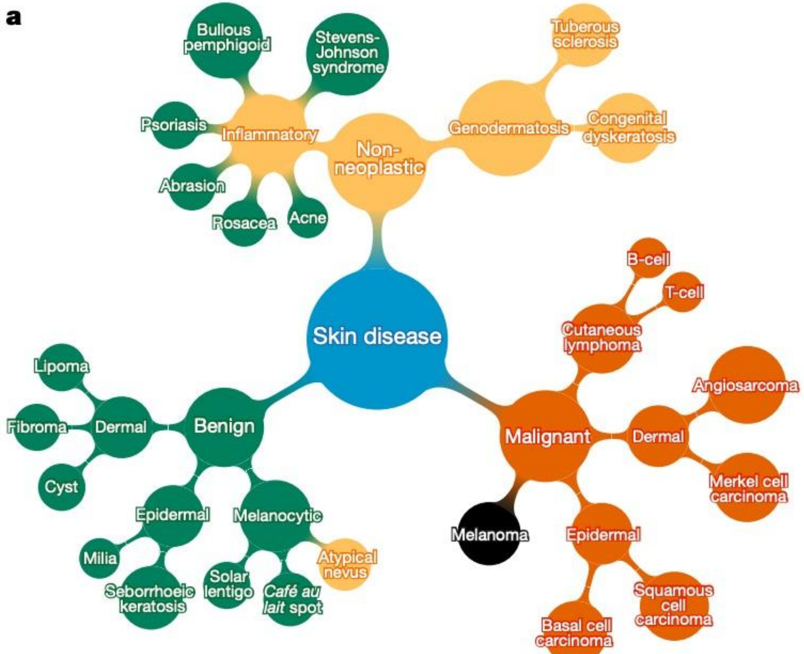
16,577 clinical images labeled with skin conditions and Fitzpatrick skin types

12,672 images from DermaAmin and 3,905 images from Atlas Dermatologico

	A	B	C	D	E	F	G
1	md5hash	fitzpatrick	label	nine_partition_label	three_partition_label	qc	url
2	5e82a45bc5d78bd24ae9202d194423f8	3	drug induced pigmentary changes	inflammatory	non-neoplastic		https://www.dermaamin.com/site/images/clinical-pic/m/minocycline-pigmentation/minocycline-pigmentation1.jpg
3	fa2911a9b13b6f8af79cb700937cc14f	1	photodermatoses	inflammatory	non-neoplastic		https://www.dermaamin.com/site/images/clinical-pic/p/photosensitivity/photosensitivity18.jpg
4	d2bac3c9e4499032ca8e9b07c7d3bc40	2	dermatofibroma	benign dermal	benign		https://www.dermaamin.com/site/images/clinical-pic/d/dermatofibroma/dermatofibroma71.jpg
5	0a94359e7eaacd7178e06b2823777789	1	psoriasis	inflammatory	non-neoplastic		https://www.dermaamin.com/site/images/clinical-pic/p/psoriasis/psoriasis38.jpg
6	a39ec3b1f22				tic		https://www.dermaamin.com/site/images/clinical-pic/p/psoriasis-scalp/psoriasis-scalp20.jpg
7	45f7fe0e102						https://www.dermaamin.com/site/images/clinical-pic/k/kaposis-sarcoma/kaposis-sarcoma4.jpg
8	6c395be9325				tic		https://www.dermaamin.com/site/images/clinical-pic/s/sweet-syndrome/sweet-syndrome98.jpg
9	9dc73230c77				tic		https://www.dermaamin.com/site/images/clinical-pic/g/granuloma_annulare/granuloma_annulare41.jpg
10	f23937e86de				tic		https://www.dermaamin.com/site/images/clinical-pic/L/larva-migrans/larva-migrans88.jpg
11	09d46db9589				tic		https://www.dermaamin.com/site/images/clinical-pic/a/allergic_contact_dermatitis/allergic_contact_dermatitis114.jpg
12	9bc21ae9502				tic		https://www.dermaamin.com/site/images/clinical-pic/n/necrobiois-lipoidica-diabeticorum/necrobiois-lipoidica-diabeticorum88.jpg
13	e702b1a7dc4				tic		https://www.dermaamin.com/site/images/clinical-pic/s/sweet-syndrome/sweet-syndrome50.jpg
14	ddcad677b7b				tic		https://www.dermaamin.com/site/images/clinical-pic/h/hidradenitis_suppurativa/hidradenitis_suppurativa50.jpg
15	b87804452f6						https://www.dermaamin.com/site/images/clinical-pic/L/lmm/lmm6.jpg
16	d1fb87ee7ee				tic	1 Diagnostic	https://www.dermaamin.com/site/images/clinical-pic/a/acne_vulgaris/acne_vulgaris150.jpg
17	8438db40abd				tic		https://www.dermaamin.com/site/images/clinical-pic/n/necrobiois-lipoidica-diabeticorum/necrobiois-lipoidica-diabeticorum7.jpg
18	2d57e08861b				tic		https://www.dermaamin.com/site/images/clinical-pic/s/sarcoidosis-of-the-skin-plaque-form/sarcoidosis-of-the-skin-plaque-form15.jpg
19	1e119546f5b				tic		https://www.dermaamin.com/site/images/clinical-pic/X/xeroderma-pigmentosum/xeroderma-pigmentosum13.jpg
20	4c3f795cf8eb						https://www.dermaamin.com/site/images/clinical-pic/m/melanoma/melanoma17.jpg
21	99247c9fe48						https://www.dermaamin.com/site/images/clinical-pic/d/dermatofibroma/dermatofibroma13.jpg
22	b09233673fc						https://www.dermaamin.com/site/images/clinical-pic/a/actinic_keratoses/actinic_keratoses83.jpg
23	449d63bec3a				tic		https://www.dermaamin.com/site/images/clinical-pic/L/localized-scleroderma/localized-scleroderma3.jpg
24	7a06b6baf51				tic		https://www.dermaamin.com/site/images/clinical-pic/h/hidradenitis_suppurativa/hidradenitis_suppurativa18.jpg
25	fb9640a13e0						https://www.dermaamin.com/site/images/clinical-pic/s/syringoma/syringoma33.jpg

Skin Condition Labels

3 high-level categories, 9 mid-level categories, 114 low-level categories



Citation: Esteva et al 2017 Dermatologist-level classification of skin cancer

Fitzpatrick Skin Type Labels

The Fitzpatrick Scale

Select one of seven choices: 1, 2, 3, 4, 5, 6, and unknown.

Fitzpatrick Skin Type Labels

	Accuracy	Accuracy (off-by-one)	# of Images
Type 1	49%	79%	10
Type 2	38%	84%	100
Type 3	25%	71%	98
Type 4	26%	71%	47
Type 5	34%	85%	44
Type 6	59%	83%	13

Table 2. Accuracy of human annotators relative to the gold standard dataset of 312 Fitzpatrick skin type annotations provided by a board-certified dermatologist.

Data Distribution

	Non-Neoplastic	Benign	Malignant
# Images	12,080	2,234	2,263
Type 1	17.0%	19.9%	20.2%
Type 2	28.1%	30.0%	32.8%
Type 3	19.7%	21.2%	20.2%
Type 4	17.5%	16.4%	13.3%
Type 5	10.1%	7.1%	6.5%
Type 6	4.4%	2.0%	2.7%
Unknown	3.2%	3.3%	4.6%

Table 1. Distribution of skin conditions in *Fitzpatrick 17k* by Fitzpatrick skin type and high level skin condition categorization.

```python
dataloaders, dataset_sizes = custom_load(
    256,
    20,
    "{}".format(train_path),
    "{}".format(test_path))
model_ft = models.vgg16(pretrained=True)
for param in model_ft.parameters():
    param.requires_grad = False
model_ft.classifier[6] = nn.Sequential(
            nn.Linear(4096, 256),
            nn.ReLU(),
            nn.Dropout(0.4),
            nn.Linear(256, len(label_codes)),
            nn.LogSoftmax(dim=1))
```

```python
transform=transforms.Compose([
    transforms.ToPILImage(),
    transforms.RandomResizedCrop(size=256, scale=(0.8, 1.0)),
    transforms.RandomRotation(degrees=15),
    transforms.ColorJitter(),
    transforms.RandomHorizontalFlip(),
    transforms.CenterCrop(size=224),  # Image net standards
    transforms.ToTensor(),
    transforms.Normalize([0.485, 0.456, 0.406],
                         [0.229, 0.224, 0.225])
    ])
```

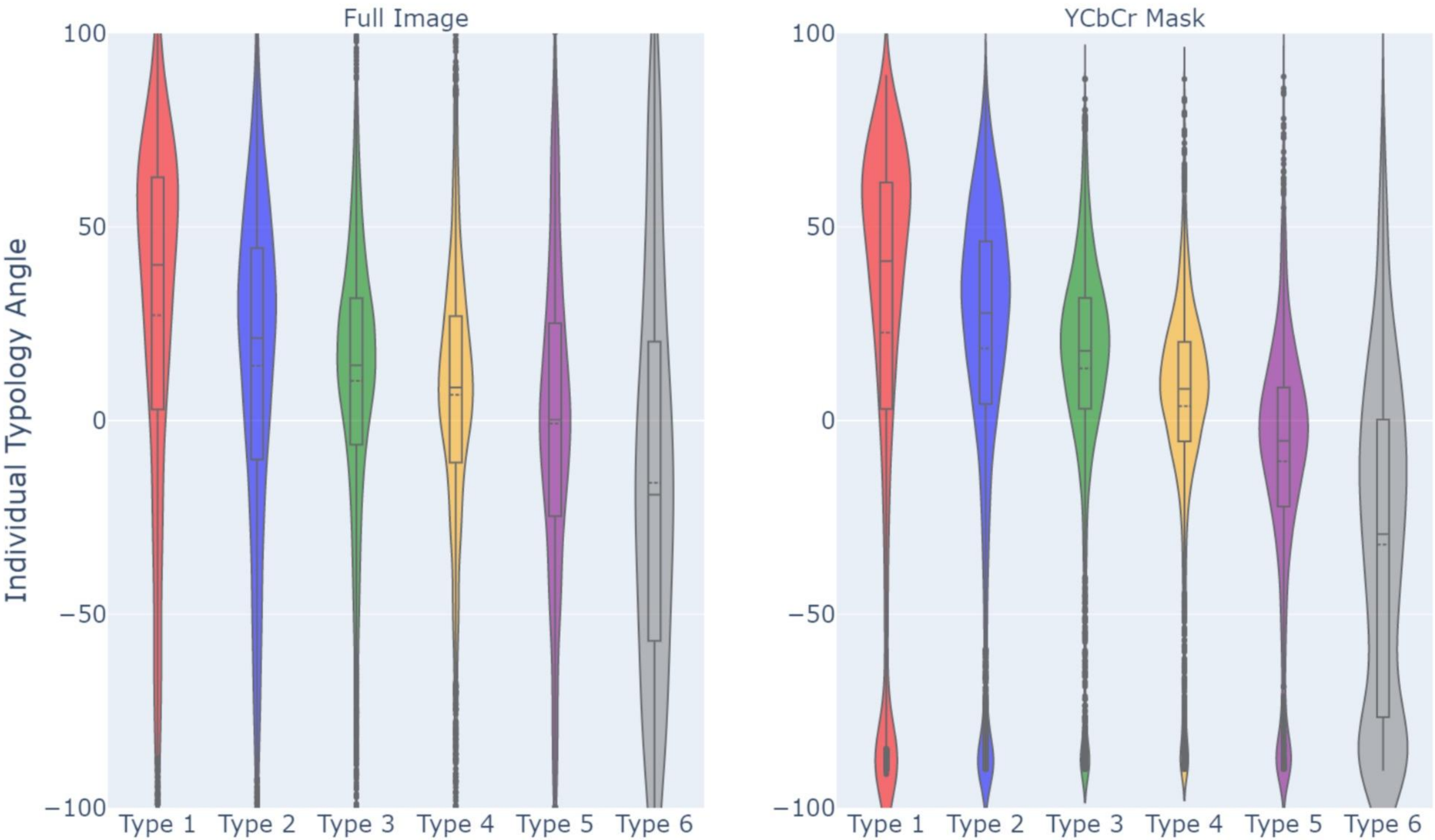
Holdout Set	Verified	Random	Source A	Source B	Fitz 3-6	Fitz 1-2 & 5-6	Fitz 1-4
# Train Images	16,229	12,751	12,672	3,905	7,755	6,089	2,168
# Test Images	348	3,826	3,905	12,672	8,257	10,488	14,409
Overall	26.7%	20.2%	27.4%	11.4%	13.8%	13.4%	7.7%
Type 1	15.1%	15.8%	40.1%	6.6%	-	10.0%	4.4%
Type 2	23.9%	16.9%	27.7%	8.6%	-	13.0%	5.5%
Type 3	27.9%	22.2%	25.3%	13.7%	15.9%	-	9.1%
Type 4	30.9%	24.1%	26.2%	17.1%	14.2%	-	12.9%
Type 5	37.2%	28.9%	28.4%	17.6%	10.1%	21.1%	-
Type 6	28.2%	15.5%	25.7%	14.9%	9.0%	12.1%	-

Holdout Set	Verified	Random	Source A	Source B	Fitz 3-6	Fitz 1-2 & 5-6	Fitz 1-4
# Train Images	16,229	12,751	12,672	3,905	7,755	6,089	2,168
# Test Images	348	3,826	3,905	12,672	8,257	10,488	14,409
Overall	26.7%	20.2%	27.4%	11.4%	13.8%	13.4%	7.7%
Type 1	15.1%	15.8%	40.1%	6.6%	-	10.0%	4.4%
Type 2	23.9%	16.9%	27.7%	8.6%	-	13.0%	5.5%
Type 3	27.9%	22.2%	25.3%	13.7%	15.9%	-	9.1%
Type 4	30.9%	24.1%	26.2%	17.1%	14.2%	-	12.9%
Type 5	37.2%	28.9%	28.4%	17.6%	10.1%	21.1%	-
Type 6	28.2%	15.5%	25.7%	14.9%	9.0%	12.1%	-

Holdout Set	Verified	Random	Source A	Source B	Fitz 3-6	Fitz 1-2 & 5-6	Fitz 1-4
# Train Images	16,229	12,751	12,672	3,905	7,755	6,089	2,168
# Test Images	348	3,826	3,905	12,672	8,257	10,488	14,409
Overall	26.7%	20.2%	27.4%	11.4%	13.8%	13.4%	7.7%
Type 1	15.1%	15.8%	40.1%	6.6%	-	10.0%	4.4%
Type 2	23.9%	16.9%	27.7%	8.6%	-	13.0%	5.5%
Type 3	27.9%	22.2%	25.3%	13.7%	15.9%	-	9.1%
Type 4	30.9%	24.1%	26.2%	17.1%	14.2%	-	12.9%
Type 5	37.2%	28.9%	28.4%	17.6%	10.1%	21.1%	-
Type 6	28.2%	15.5%	25.7%	14.9%	9.0%	12.1%	-

Skin Type Classification Systems

- Visual
  - Fitzpatrick skin type, Glogau wrinkle scale, Goldman world classification of skin types, Roberts skin type classification system, Taylor hyperpigmentation scale, von Luschan chromatic scale
- Self-reported
  - Baumann skin type, Fanous classification, Kawada skin classification system for Japanese individuals, Lancer ethnicity scale, Modified Fitzpatrick skin type, Willis and Earles scale
- Algorithmic
  - Individual Typology Angle (ITA)
- Spectrophotometer
  - Melanin Index (requires a spectrophotometer)



Full Image
YCbCr Mask
Individual Typology Angle
100
50
0
−50
−100
Type 1
Type 2
Type 3
Type 4
Type 5
Type 6

Figure 3. 18 images plot arranged based on ITA values and Fitzpatrick labels.

Figure 3. 18 images plot arranged based on ITA values and Fitzpatrick labels.

Take-aways

(1) Dark skin is underrepresented in many aspects of dermatology
(2) A deep neural network trained to classify skin conditions does better on skin types similar to the ones upon which it was trained
(3) Automated methods for calculating skin type can be noisy

Thanks!

Please, feel free to reach out to us!

groh@mit.edu
harris@mit.edu
soenksen@mit.edu
felix.lau@scale.com
rachel.han@scale.com
aerin.kim@scale.com
arash.koochek@bannerhealth.com
obadri@gmail.com