Fitzpatrick 17k dataset team

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

Open Data
@ MIT
Evaluating Deep Neural Networks Trained on Clinical Images in Dermatology with the Fitzpatrick 17k Dataset

Matthew Groh
MIT Media Lab
Cambridge, MA
groh@mit.edu

Caleb Harris
MIT Media Lab
Cambridge, MA
harrisc@mit.edu

Luis Soenksen
MIT, Harvard University
Cambridge, MA
soenksen@mit.edu

Felix Lau
Scale
San Francisco, CA
felix.lau@scale.com

Rachel Han
Scale
San Francisco, CA
rachel.han@scale.com

Aerin Kim
Scale
San Francisco, CA
aerin.kim@scale.com

Arash Koochek
Banner Health
Phoenix, AZ
arash.koochek@bannerhealth.com

Omar Badri
Northeast Dermatology Associates
Beverly, MA
obadri@gmail.com
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.

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
Lack of Publicly Available Dermatology Images 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 Dermatology Images 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)
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
Enabling Future Research

- Enables algorithmic audits and explores selection bias in training datasets and algorithmic bias in models

- 114 researchers have requested to download the dataset since June 2021

- Cited in 20+ articles including articles in Science Advances, Nature Communications, NPJ Digital Medicine, JAMA Dermatology, International Conference on Machine Learning, and elsewhere