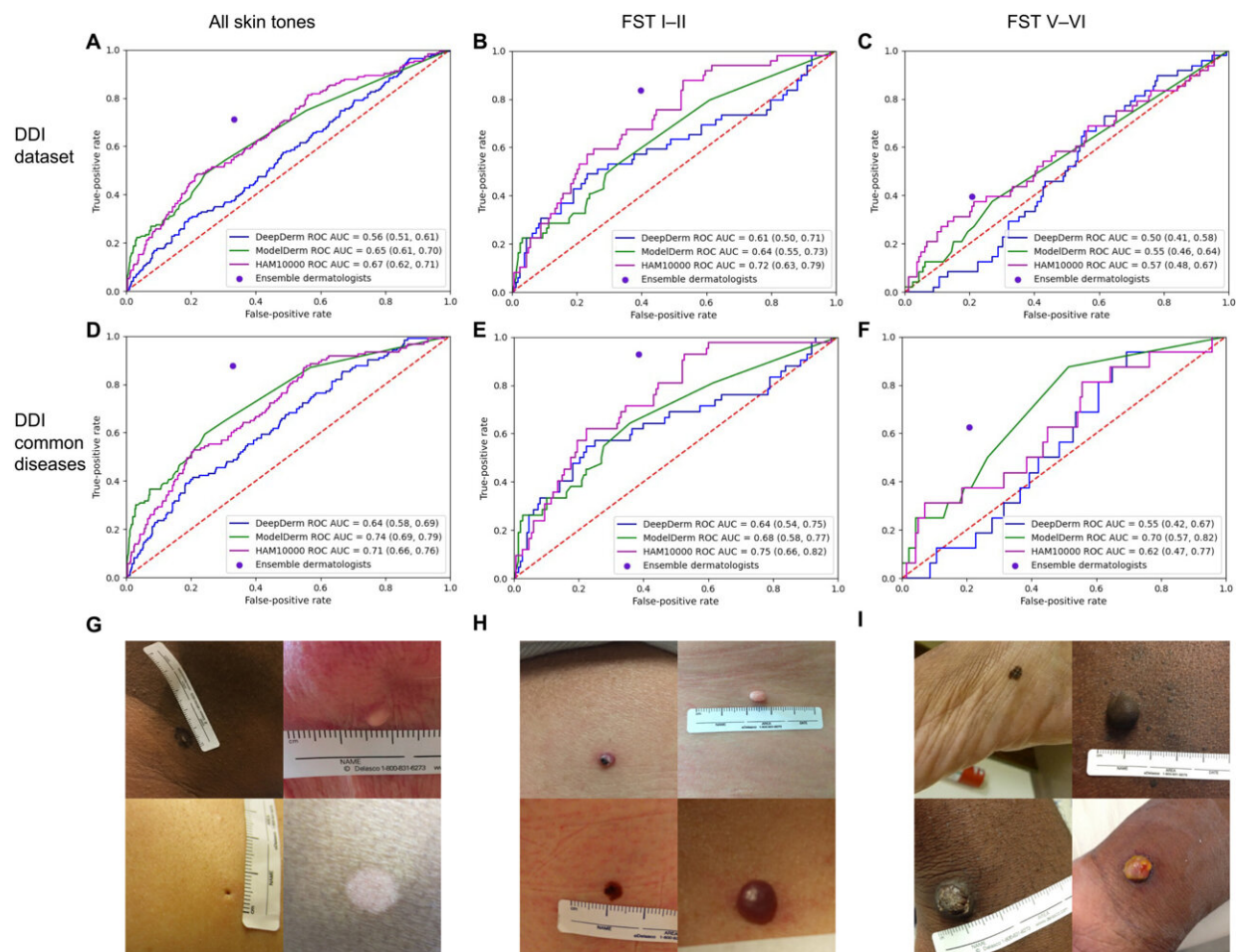# Training physicians and algorithms in dermatology diversity

September 15 2022, by Hanae Armitage

DDI dataset and algorithm performance. Row 1: Performance of all three AI models and the majority vote of an ensemble of dermatologists on the entire DDI dataset (A), FST I–II (B), and FST V–VI (C). Row 2: Performance across the DDI common diseases dataset with the performance of all algorithms and ensemble of dermatologists on the entire DDI common diseases dataset (D), FST

I–II (E), and FST V–VI (F). Row 3: Example images from the entire DDI dataset for all skin tones (G), FST I–II (H), and FST V–VI (I). Photo Credit: DDI dataset, Stanford School of Medicine. Credit: *Science Advances* (2022). DOI: 10.1126/sciadv.abq6147

There's a long-standing challenge in dermatology: textbooks, databases, journals and lectures are largely bereft of images that feature darker skin.

Their absence can cause gaps in clinical expertise and in diagnosis, as symptoms of a disease don't necessarily appear the same on all skin tones. Physicians trained to identify signs of illness on lighter shades can overlook them in people with a darker complexion, and algorithms trained on a sea of beige pictures may miss signs of disease when evaluating images from a patient with brown skin.

"Algorithms are only as good as the data on which they are based," said James Zou, Ph.D., Stanford Medicine assistant professor of biomedical data science and a machine learning expert. "A massive, open database cataloging dermatological images from people of color could help doctors assess whether these algorithms function accurately on all skin colors."

He and others at Stanford have been compiling such a database for years.

Using a preliminary version of that database, Zou; Roxana Daneshjou, MD, Ph.D., a practicing dermatologist at Stanford Medicine; graduate student Kailas Vodrahalli; and others conducted a study, published Aug. 12 in *Science Advances*, to evaluate algorithms used in dermatology.

The group dug into potential bias by pulling previously described

algorithms and testing their accuracy on diverse skin images. The result was somewhat predictable: Algorithms trained on homogenous, light-skinned images accurately detected dermatological disease in that shade of skin. But not so much for darker tones.

The good news? Daneshjou, Zou and clinical associate professor of dermatology Albert Chiou, MD, who shares co-senior authorship of the study with Zou, have devised with a way to right the wrongs of biased algorithms.

"People who are in the business of creating algorithms need to be aware of this problem and make sure they're testing their [algorithm](#) on all sorts of diverse skin tones," said Daneshjou, the lead author of the study. "It just emphasizes the importance of having diverse teams where both physicians and machine-learning experts from [diverse backgrounds](#) are involved."

A database of diverse skin tones

The amount of new clinically relevant algorithms has seen a dramatic uptick in the past several years, thanks to more data availability. But availability and diversity aren't the same things—and most databases of skin images are still predominately beige.

"Erythema or redness, for instance, is a feature of disease that appears different on dark versus light skin," Daneshjou said. "That's why it's important for physicians—and algorithms—to know the differences in what they're looking for."

The new, more diverse database, created by Daneshjou and others at the Stanford Center for Artificial Intelligence in Medicine and Imaging, includes a wide repository of medical images of patients who aren't identified.

When they tested the accuracy of various published algorithms on the database, they saw much poorer algorithmic performance on images of black and brown skin. "But when we took a subset of our diverse data and finetuned the algorithms, we were able to close the gap in performance," Daneshjou said.

Daneshjou and Zou have made the diverse database available to scientists who want to use their data to fine-tune their algorithms or test for biases. And, Zou said, the database could be helpful for the general public.

"Often, people will spot something—a mole for instance—and want to look up previous cases on the internet," he said. "This could be a valuable resource for patients who might not otherwise be able to find images of skin that looks like theirs."

More information: Roxana Daneshjou et al, Disparities in dermatology AI performance on a diverse, curated clinical image set, *Science Advances* (2022). DOI: 10.1126/sciadv.abq6147

Provided by Stanford University