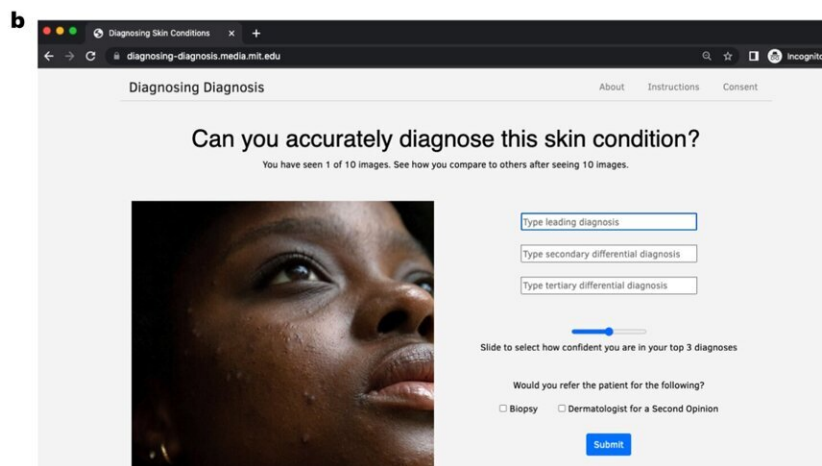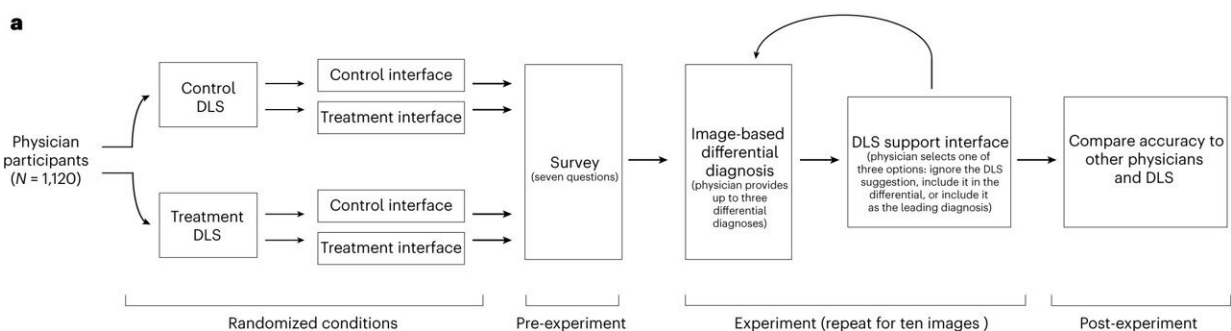# Doctors have more difficulty diagnosing disease when looking at images of darker skin: AI may offer a solution

February 5 2024, by Anne Trafton

Experimental design flowchart and user interface. a, Flowchart describing the experimental design, including the number of participants, the randomly assigned conditions, the pre-experiment survey, the experiment, and the final stage, in which participants can see how they compare to other participants. b, Screenshot of the user interface for the image-based differential diagnosis portion of the experiment. Panel b reproduced with permission from Refinery29.

When diagnosing skin diseases based solely on images of a patient's skin, doctors do not perform as well when the patient has darker skin, according to a new study from MIT researchers.

The study, which included more than 1,000 dermatologists and general practitioners, found that dermatologists accurately characterized about 38% of the images they saw, but only 34% of those that showed darker skin. General practitioners, who were less accurate overall, showed a similar decrease in accuracy with darker skin.

The research team also found that assistance from an artificial intelligence algorithm could improve doctors' accuracy, although those improvements were greater when diagnosing patients with lighter skin.

While this is the first study to demonstrate physician diagnostic disparities across skin tone, other studies have found that the images used in dermatology textbooks and training materials predominantly feature lighter skin tones. That may be one factor contributing to the discrepancy, the MIT team says, along with the possibility that some doctors may have less experience in treating patients with darker skin.

"Probably no doctor is intending to do worse on any type of person, but it might be the fact that you don't have all the knowledge and the experience, and therefore on certain groups of people, you might do worse," says Matt Groh Ph.D. '23, an assistant professor at the Northwestern University Kellogg School of Management. "This is one of those situations where you need empirical evidence to help people figure out how you might want to change policies around dermatology education."

Groh is the lead author of the study, which is [published](link) today (Feb. 5) in *Nature Medicine*. Rosalind Picard, an MIT professor of media arts and sciences, is the senior author of the paper.

Diagnostic discrepancies

Several years ago, an MIT study led by Joy Buolamwini Ph.D. '22 found that facial-analysis programs had much higher error rates when predicting the gender of darker skinned people. That finding inspired Groh, who studies human-AI collaboration, to look into whether AI models, and possibly doctors themselves, might have difficulty diagnosing skin diseases on darker shades of skin—and whether those diagnostic abilities could be improved.

"This seemed like a great opportunity to identify whether there's a social problem going on and how we might want fix that, and also identify how to best build AI assistance into medical decision-making," Groh says. "I'm very interested in how we can apply machine learning to real-world problems, specifically around how to help experts be better at their jobs. Medicine is a space where people are making really important decisions, and if we could improve their decision-making, we could improve patient outcomes."

To assess doctors' diagnostic accuracy, the researchers compiled an array of 364 images from dermatology textbooks and other sources, representing 46 skin diseases across many shades of skin.

Most of these images depicted one of eight inflammatory skin diseases, including atopic dermatitis, Lyme disease, and secondary syphilis, as well as a rare form of cancer called cutaneous T-cell lymphoma (CTCL), which can appear similar to an inflammatory skin condition. Many of these diseases, including Lyme disease, can present differently on dark and light skin.

The research team recruited subjects for the study through Sermo, a social networking site for doctors. The total study group included 389 board-certified dermatologists, 116 dermatology residents, 459 general practitioners, and 154 other types of doctors.

Each of the study participants was shown 10 of the images and asked for their top three predictions for what disease each image might represent. They were also asked if they would refer the patient for a biopsy. In addition, the general practitioners were asked if they would refer the patient to a dermatologist.
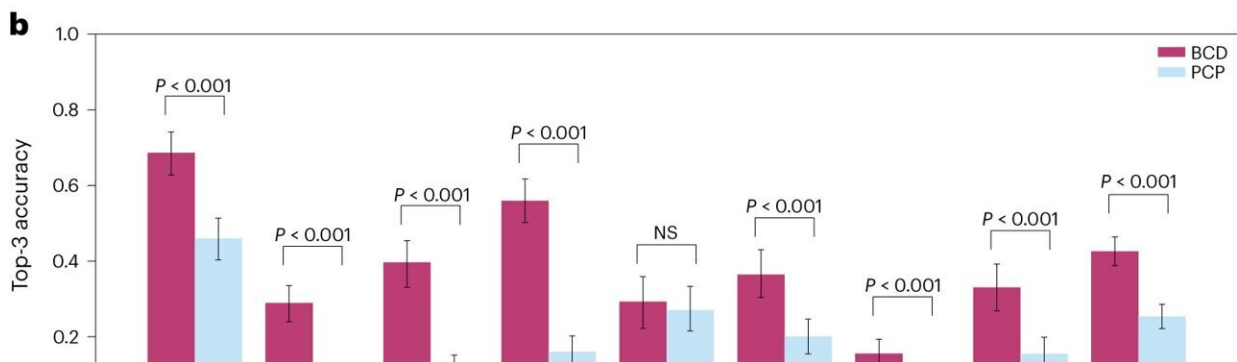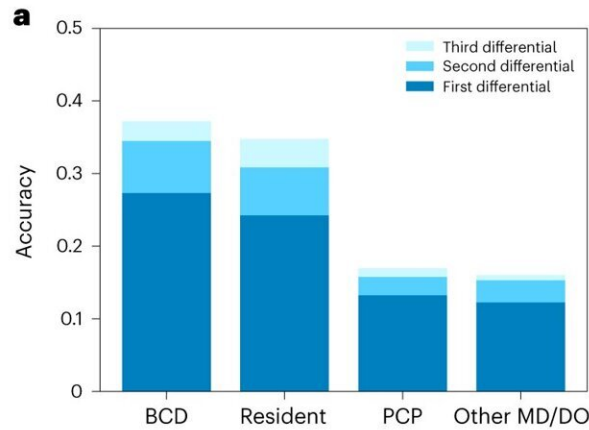
Diagnostic accuracy across skin diseases. a, Diagnostic accuracy of physician participants on the eight main skin diseases. "Resident" refers strictly to dermatology residents. Other MD/DO refers to other physicians who have a doctor of medicine or doctor of osteopathic medicine degree. b, Top-3

"This is not as comprehensive as in-person triage, where the doctor can examine the skin from different angles and control the lighting," Picard says. "However, skin images are more scalable for online triage, and they are easy to input into a machine-learning algorithm, which can estimate likely diagnoses speedily."

The researchers found that, not surprisingly, specialists in dermatology had higher accuracy rates: They classified 38% of the images correctly, compared to 19% for general practitioners.

Both of these groups lost about four percentage points in accuracy when trying to diagnose skin conditions based on images of darker skin—a statistically significant drop. Dermatologists were also less likely to refer darker skin images of CTCL for biopsy, but more likely to refer them for biopsy for noncancerous skin conditions.

"This study demonstrates clearly that there is a disparity in diagnosis of skin conditions in dark skin. This disparity is not surprising; however, I have not seen it demonstrated in the literature such a robust way. Further research should be performed to try and determine more precisely what the causative and mitigating factors of this disparity might be," says Jenna Lester, an associate professor of dermatology and director of the Skin of Color Program at the University of California at San Francisco, who was not involved in the study.

A boost from AI

After evaluating how doctors performed on their own, the researchers also gave them additional images to analyze with assistance from an AI algorithm the researchers had developed. The researchers trained this algorithm on about 30,000 images, asking it to classify the images as one of the eight diseases that most of the images represented, plus a ninth category of "other."

This algorithm had an accuracy rate of about 47%. The researchers also created another version of the algorithm with an artificially inflated success rate of 84%, allowing them to evaluate whether the accuracy of the model would influence doctors' likelihood to take its recommendations.

"This allows us to evaluate AI assistance with models that are currently the best we can do, and with AI assistance that could be more accurate, maybe five years from now, with better data and models," Groh says.

Both of these classifiers are equally accurate on light and dark skin. The researchers found that using either of these AI algorithms improved accuracy for both dermatologists (up to 60%) and general practitioners (up to 47% percent).

They also found that doctors were more likely to take suggestions from the higher-accuracy algorithm after it provided a few correct answers, but they rarely incorporated AI suggestions that were incorrect. This suggests that the doctors are highly skilled at ruling out diseases and won't take AI suggestions for a disease they have already ruled out, Groh says.

"They're pretty good at not taking AI advice when the AI is wrong and the physicians are right. That's something that is useful to know," he says.

While dermatologists using AI assistance showed similar increases in accuracy when looking at images of light or dark skin, general practitioners showed greater improvement on images of lighter skin than darker skin.

"This study allows us to see not only how AI assistance influences, but how it influences across levels of expertise," Groh says. "What might be going on there is that the PCPs don't have as much experience, so they don't know if they should rule a disease out or not because they aren't as deep into the details of how different skin diseases might look on different shades of skin."

The researchers hope that their findings will help stimulate medical schools and textbooks to incorporate more training on patients with darker skin. The findings could also help to guide the deployment of AI assistance programs for dermatology, which many companies are now developing.

  More information: Matthew Groh et al, Deep learning-aided decision support for diagnosis of skin disease across skin tones, *Nature Medicine* (2024). DOI: 10.1038/s41591-023-02728-3

This story is republished courtesy of MIT News (web.mit.edu/newsoffice/), a popular site that covers news about MIT research, innovation and teaching.

Provided by Massachusetts Institute of Technology