Online Supplementary Figures

Dissection of medical AI reasoning processes via physician and generative-AI collaboration

Alex J. DeGrave[1,2], Zhuo Ran Cai[3], Joseph D. Janizek[1,2], Roxana Daneshjou[4,5,*], and Su-In Lee[1,*]

[1]Paul G. Allen School of Computer Science and Engineering, University of Washington
[2]Medical Scientist Training Program, University of Washington
[3]Program for Clinical Research and Technology, Stanford University
[4]Department of Dermatology, Stanford School of Medicine
[5]Department of Biomedical Data Science, Stanford School of Medicine
[*]indicates co-senior authorship

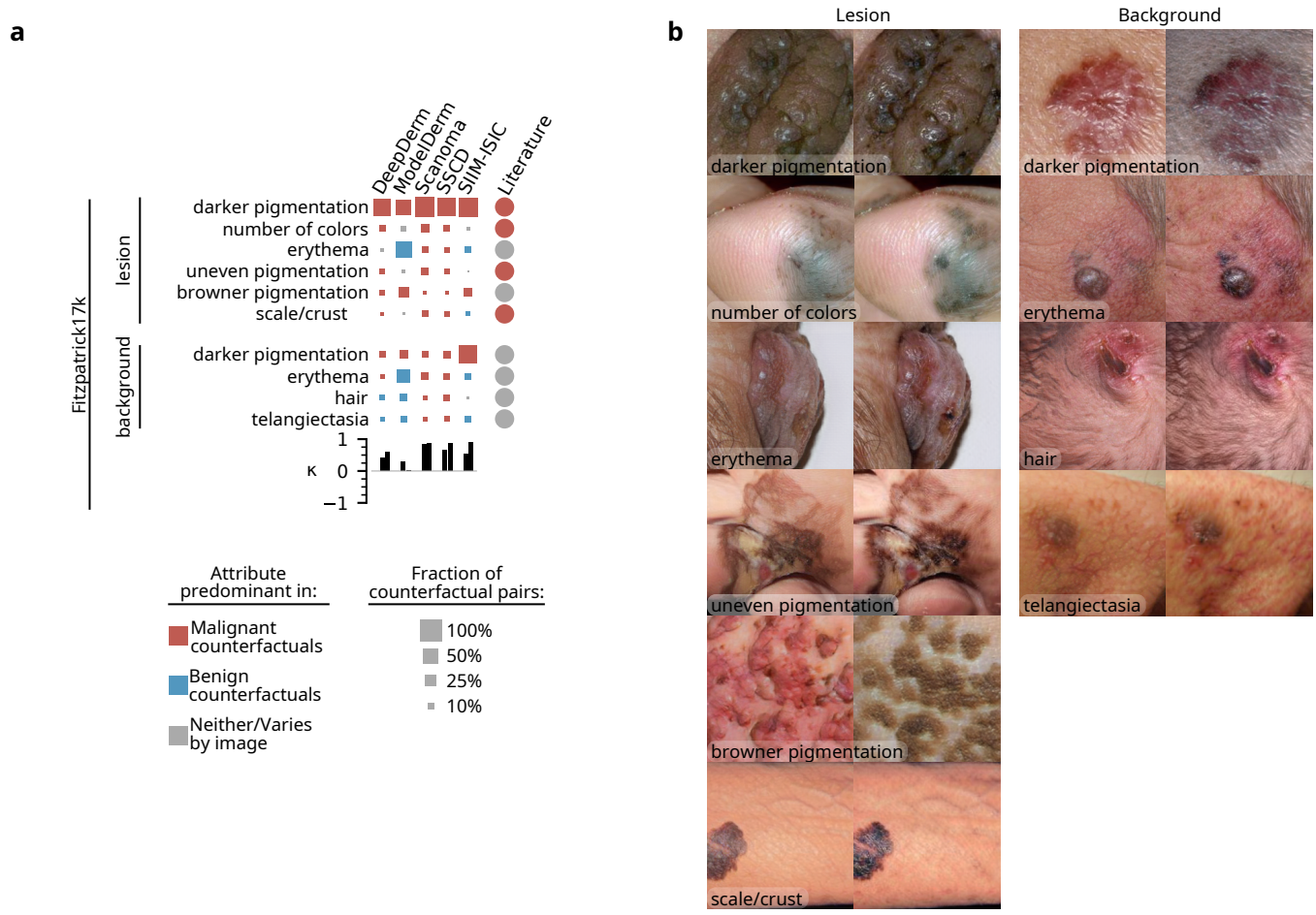

Fig. 1 | Examples of counterfactuals generated from clinical images in the dataset Fitzpatrick17k. **a**, Attributes identified by our joint expert-XAI auditing procedure as key influences on the output of dermatology AI devices (excerpted from main text Fig. 2c). **b**, Examples of counterfactuals (generated from clinical images in the dataset Fitzpatrick17k) that differ in each of the top ten attributes identified in the Fitzpatrick17k data; the attribute is present to a greater extent in the right image of each pair.

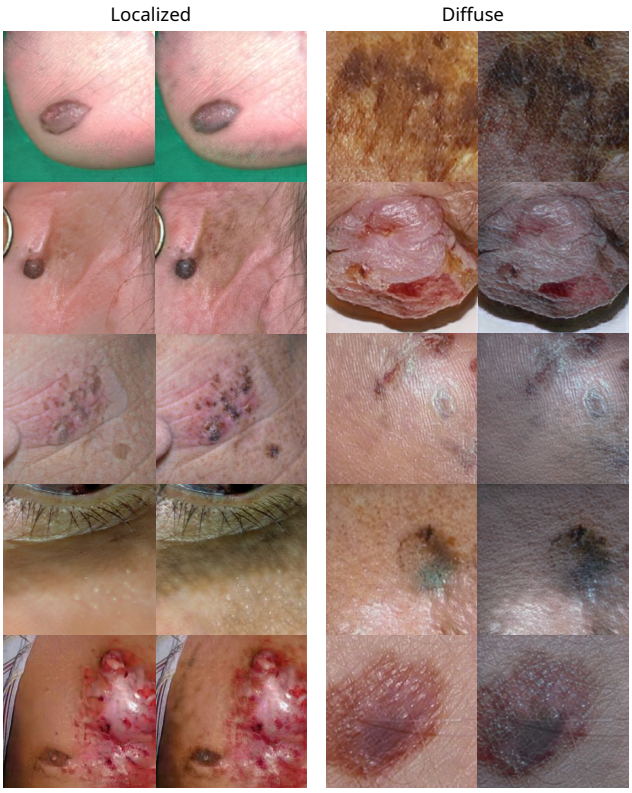

Fig. 2 | Examples of counterfactuals featuring darker pigmentation of the background skin. Annotators applied this term to describe both images with localized areas of darker pigmentation outside the primary lesion (left) and images with diffuse darker pigmentation of the background skin. In all counterfactual pairs, the counterfactual with darker pigmentation of the background skin appears on the right.