

# Bridging the gap: Addressing the Underrepresentation of Skin-of-Color in AI-powered Dermatology Tools

Zaina A. Chaudhry, MPA; Gloria Bachmann, MD

## Introduction

Artificial intelligence (AI) is revolutionizing dermatologic diagnostics by enhancing image analysis, streamlining workflows, and expanding access to care. However, many AI tools have been trained predominantly on images of lighter skin tones, raising concerns about reduced accuracy in patients with skin of color.

This underrepresentation may perpetuate diagnostic disparities and limit the equitable application of AI in dermatology. The goal of this project is to evaluate how racial and skin tone bias affects AI performance and to highlight innovative strategies aimed at promoting more inclusive and clinically robust tools.

## Methodology

A targeted literature review was conducted using peer-reviewed sources from academic databases, guided by search terms including: *dermatology*, *artificial intelligence*, *machine learning*, *algorithmic bias*, and *skin tone*. Selection criteria prioritized recent, clinically relevant studies with rigorous methodology and strong implications for dermatologic practice.

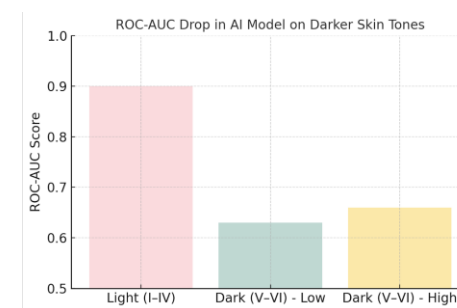
Three high-impact publications were identified and critically evaluated:

- 1. Disparities in AI Performance on a Diverse Dermatology Image Set** – assessing model accuracy across Fitzpatrick skin types.
- 2. Representations of Skin Tone and Sex in Generative AI** – analyzing bias in synthetic dermatologic imagery.
- 3. Enhancing Skin Color Diversity in Cancer Detection** – utilizing deep learning to address dataset imbalances.

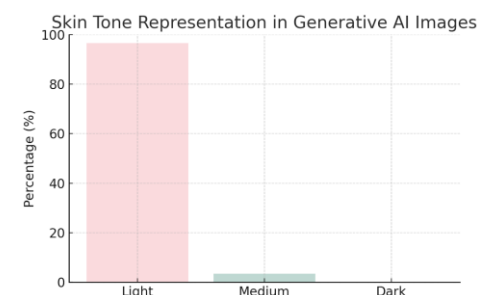
## Results

The first study analyzed the performance of AI models on the Diverse Dermatology Images (DDI) dataset, which included 656 biopsy-confirmed images across all Fitzpatrick skin types, emphasizing darker tones (V–VI). These models showed a significant drop in diagnostic accuracy for darker skin tones, with ROC-AUC scores decreasing by **27–36%** compared to lighter skin. Importantly, lesion types such as melanoma and basal cell carcinoma were more frequently misclassified in patients with skin of color.

## Results (con't)

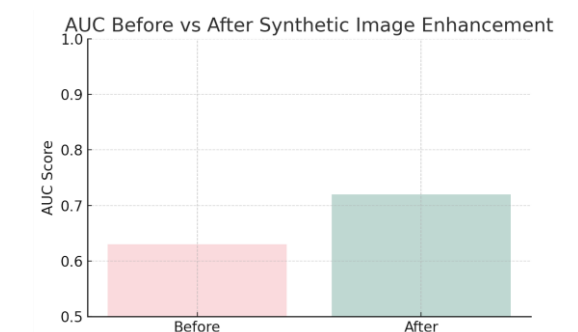
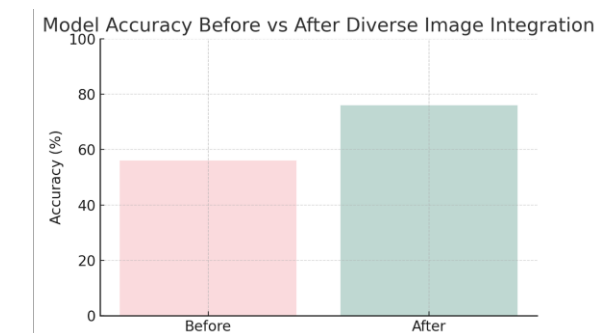


The second study, focused on generative AI tools (e.g., DALL-E 3, Midjourney), evaluated how dermatologic conditions are visually represented. Findings revealed a **critical lack of diversity: 96.6%** of generated images depicted light skin, while **0%** featured dark skin. This lack of representation raises concerns for patient education, training tools, and equitable AI development. Additionally, the models showed a **gender imbalance**, with **68%** of generated images representing male-presenting individuals.



The third study employed deep learning enhancement techniques, **style transfer (ST)** and **deep blending (DB)**, to synthetically create darker skin tone images for melanoma detection. Dermatologists found ST images **75% diagnostically useful**, and after these images were integrated into training datasets, model accuracy increased from **56% to 76%**, and AUC improved from **0.63 to 0.72**. These results underscore the clinical utility of targeted dataset diversification to reduce algorithmic bias.

## Results (con't)



## Conclusion

AI-powered dermatologic tools show immense promise, but current models remain limited by underrepresentation of skin of color in training datasets. This disparity contributes to reduced diagnostic accuracy and raises concerns about equitable care delivery.

Emerging strategies such as synthetic image generation and intentional dataset diversification demonstrate meaningful improvements in both performance and inclusivity. Moving forward, the development of dermatologic AI must be guided by principles of equity, transparency, and representation to ensure these technologies serve all patient populations effectively.