AI is emerging in dermatology for image-based diagnosis and teledermatology.

UNIVERSITY OF MIAMI
MILLER SCHOOL OF MEDICINE
DEPARTMENT of INFORMATICS
& HEALTH DATA SCIENCE

ChatGPT-4o Proves Effective in Accurately Diagnosing Dermoscopic Images of Benign and Malignant Skin Conditions

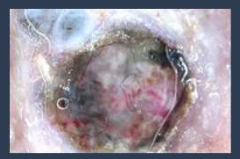
Mascha Korsch, PhD, Ryan Scheinkman, MEng, Lea Tordjman, BS, Lauren Dender, MS, Daniel Green, BS, Keyvan Nouri, MD and Azizi A. Seixas, PhD

University of Miami Miller School of Medicine

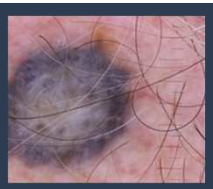
Introduction

- Background: AI is emerging in dermatology for image-based diagnosis and teledermatology, where pattern recognition is essential.
- Problem: Accuracy and bias remain concerns, especially across different skin tones. [1]
- Significance: Validated AI can improve early detection and reduce care disparities.[2]
- Objective: Evaluate ChatGPT-4o's ability to classify benign and malignant skin lesions using dermoscopic images across skin phototypes I-III, and assess accuracy, consistency, and potential clinical utility for dermatology and telehealth.
- ChatGPT-4o (OpenAI) is a multimodal generative pre-trained transformer model with over 189 million weekly users.[3]

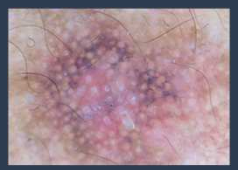
Figure 1: Four common types of Skin lesions Dermatologists assess



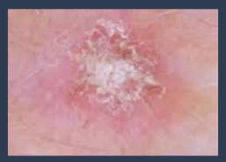
Melanoma



Blue Nevus



Actinic Keratosis



Cutaneous SCC

Methodology

- Ten skin lesions were selected for the diagnostic process: basal cell carcinoma (BCC), melanoma, invasive squamous cell carcinoma (SCC), actinic keratosis, cherry angioma (CA), solar lentigo, seborrheic keratosis, blue nevus (BN), dermatofibroma , and lichen planus.
- For each, 15 dermoscopic images from FST I-III were sourced from DermNet and Dermoscopedia, (N=150 images). Various morphologies (e.g., pigmented and nodular BCCs) were included to evaluate accuracy.
- Images were processed through ChatGPT-4o (Prompt: generate five ranked differential diagnoses from most to least likely.)
- Ratios of correctly identified images were calculated for each lesion category, focusing on both correct first identification and correct identification within the top five differential.
- Sensitivity, specificity, and accuracy were computed based on the model's top diagnosis.

Table 1: Ratios and Percentages of Correct Identification Per Lesional Dermoscopy Image

Lesion Type	Ratio of Correct 1st Identification	% Correct 1st Identification	Ratio In Top 5 Differential	% In Top 5 Differential
Basal Cell Carcinoma	11/15	73%	15/15	100%
Melanoma	15/15	100%	15/15	100%
Squamous Cell Carcinoma, Invasive	11/15	73%	15/15	100%
Actinic Keratosis	14/15	93%	15/15	100%
Cherry Angioma	8/15	53%	10/15	67%
Solar Lentigo	12/15	80%	15/15	100%
Seborrheic Keratosis	11/15	73%	15/15	100%
Blue Nevus	6/15	40%	14/15	93%
Dermatofibroma	14/15	93%	15/15	100%
Lichen Planus	15/15	100%	15/15	100%
Total	117/150	78%	144/150	96%

Results

- ChatGPT-4o correctly diagnosed 117/150 (78%) images at 1st attempt, and 144/150 (96%) within the top 5 differential diagnoses (Table 1).
- Melanoma was correctly identified on initial identification (15/15, 100% sensitivity, 91.1% specificity, 91.99% accuracy) (Table 2). **BCC and SCC** were both diagnosed upon first attempt in 11/15 cases (73% sensitivity), improving to 100% correct identification among the top 5 differentials.
- Correct initial identification was lowest for BN (6/15, 40% sensitivity) and CA (8/15, 53% sensitivity). Identification of both lesions showed significant improvement upon assessment of the top 5 differentials, (93% and 100%, respectively).
- Diagnostic accuracy exceeding 90% was achieved across all fifteen conditions (Tables 1 and Table 2).

Table 2: Sensitivity, Specificity, and Accuracy of Correct First Identification Per Lesional Dermoscopy Image

Lesion Type	Sensitivity	Specificity	Accuracy
Basal Cell Carcinoma	73%	97.03%	94.63%
Melanoma	100%	91.11%	91.99%
Squamous Cell Carcinoma, Invasive	73%	100%	97.33%
Actinic Keratosis	93%	100%	99.33%
Cherry Angioma	53%	100%	95.33%
Solar Lentigo	80%	98.50%	96.65%
Seborrheic Keratosis	73%	100%	97.33%
Blue Nevus	40%	100%	94.00%
Dermatofibroma	93%	100%	99.33%
Lichen Planus	100%	100%	100%

Conclusion

- ChatGPT-4o indicates promising diagnostic capabilities in accurately identifying various benign and malignant skin lesions in light skin using dermoscopic images.[4]
- Limitations include: small sample size, reliance on dermoscopy databases lacking consistent histopathology verification, and inclusion of a limited range of skin phototypes (I-VI).
- Nonetheless, this research provides valuable insights into the future diagnostic potential of ChatGPT-4o as a widely accessible tool in dermatology, beneficial for both educational purposes and clinical support.
- Further research should prioritize the inclusion of a diverse range of skin phototypes and larger sample sizes to comprehensively validate this AI model for dermatologic diagnosis.

References

1. Cortes J, Paravar T, Oldenburg R. Physician Opinions on Artificial Intelligence Chatbots In Dermatology: A National Online Cross-Sectional Survey of Dermatologists. *J Drugs Dermatol.* 2024;23(11):972-978. doi:10.36849/JDD.8239
2. Mahmoud NM, Soliman AM. "Early automated detection system for skin cancer diagnosis using artificial intelligent techniques". *Sci Rep.* 2024;14(1):9749. Published 2024 Apr 28. doi:10.1038/s41598-024-59783-0
3. Demand Sage. ChatGPT statistics: Key facts and figures. Published November 2024. Accessed November 4, 2024. Available from: https://www.demandsage.com/chatgpt-statistics/#:~:text=As%20of%20November%202024%2C%20ChatGPT,reached%20an%20all%2Dtime%20high.
4. Shifai N, van Doorn R, Malvehy J, Sangers TE. Can ChatGPT vision diagnose melanoma? An exploratory diagnostic accuracy study. *J Am Acad Dermatol.* 2024;90(5):1057-1059. doi:10.1016/j.jaad.2023.12.062