

## RESEARCH LETTER

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

Ryan Scheinkman, MEng<sup>1\*</sup>, Lea Tordjman, BS<sup>1\*</sup>, Lauren Dender, MS<sup>1</sup>, Daniel Green, BS<sup>1</sup>, Mascha Korsch, PhD<sup>1</sup>, and Keyvan Nouri, MD, MBA<sup>1</sup>

<sup>1</sup> Dr. Phillip Frost Department of Dermatology and Cutaneous Surgery, University of Miami Miller School of Medicine, Miami, Florida, USA

### ABSTRACT

**Introduction:** The integration of artificial intelligence (AI) and AI chatbots (AIC) into medicine is progressing rapidly, particularly within dermatology, where clinical, dermoscopic, and histological pattern recognition is essential.

**Methods:** To assess the diagnostic capabilities of ChatGPT-4o, ten skin lesions were selected: basal cell carcinoma (BCC), melanoma, invasive squamous cell carcinoma (SCC), actinic keratosis (AK), cherry angioma (CA), solar lentigo (SL), seborrheic keratosis (SK), blue nevus (BN), dermatofibroma (DF), and lichen planus (LP). For each condition, fifteen dermoscopic images were sourced from DermNet NZ and Dermoscopyedia, yielding a total of 150 images. Rank order top 5 differential diagnosis were generated for each image and sensitivity, specificity, and accuracy were computed for each.

**Results:** ChatGPT-4o correctly diagnosed 117 out of 150 (78%) images at the first attempt, and 144 out of 150 (96%) images within the top five differential diagnoses. Melanoma was correctly identified on initial diagnosis (15/15, 100% sensitivity, 91.1% specificity, 91.99% accuracy).

**Discussion:** Our research demonstrates the promising potential diagnostic capabilities of ChatGPT-4o in accurately identifying various benign and malignant skin lesions using dermoscopic images, novel findings in comparison to prior studies.

### INTRODUCTION

The integration of artificial intelligence (AI) and AI chatbots (AIC) into medicine is progressing rapidly, particularly within dermatology, where clinical, dermoscopic, and histological pattern recognition is essential. Ongoing research aims to enhance the use of teledermatology, dermoscopy, and AI-driven diagnostic systems to address disparities in dermatological care for underserved and rural populations.<sup>1</sup> While

many dermatologists anticipate the formal implementation of AIC into clinical practice, significant concerns regarding diagnostic accuracy remain, emphasizing the critical need for continuous validation of these technologies.<sup>2</sup> This study evaluates the potential of ChatGPT-4o (OpenAI), a multimodal generative pre-trained transformer AI model with over 189 million weekly users, to accurately diagnose various skin lesions through dermoscopic images.<sup>3</sup>

## METHODS

To assess the diagnostic capabilities of ChatGPT-4o, ten skin lesions were selected: basal cell carcinoma (BCC), melanoma, invasive squamous cell carcinoma (SCC), actinic keratosis (AK), cherry angioma (CA), solar lentigo (SL), seborrheic keratosis (SK), blue nevus (BN), dermatofibroma (DF), and lichen planus (LP). For each condition, fifteen dermoscopic images were sourced from DermNet NZ and Dermoscopy, yielding a total of 150 images, all Fitzpatrick phototypes I-III. Various morphologies, such as pigmented and nodular BCCs, were included to evaluate the software's accuracy in identifying different lesion subtypes under the appropriate diagnosis. Each image was processed through ChatGPT-4o with the prompt "generate a top 5 ranked list of differential diagnoses for the following dermoscopy image" to generate five ranked differential diagnoses from most to least likely, using the fixed preset ChatGPT settings and analyzed only once. Ratios of correctly diagnosed images were calculated for each skin lesion category, focusing on both correct first diagnosis and correct diagnosis within the top five differential. Sensitivity, specificity, and accuracy were computed based on the model's first diagnosis. True positives were considered images that were correctly diagnosed by ChatGPT in concordance with the expert diagnosis provided in the database for the image of each specific disease (e.g. 15/15 images of the expert diagnosed melanoma images were correctly identified as melanoma by ChatGPT representing 15 true positives/100% sensitivity). True negatives for a specific target disease considered all non-target lesion images that were treated as negatives for the target lesion (e.g. out of the 135 remaining used images that were not annotated by experts as melanoma, how

many were diagnosed as a disease other than melanoma). Closely related disease terms were treated as correct (e.g. invasive SCC, cutaneous squamous cell carcinoma, etc.).

## RESULTS

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

## DISCUSSION

Our research demonstrates the promising potential diagnostic capabilities of ChatGPT-4o in accurately identifying various benign and malignant skin lesions using dermoscopic images, novel findings in comparison to prior studies.<sup>4</sup> Limitations such as a small sample size, the reliance on dermoscopy databases lacking consistent histopathology verification, and the inclusion of a limited range of skin phototypes (I-III) must be acknowledged. Nonetheless, this

**Table 1.** Ratios and Percentages of Correct Diagnoses Per Lesional Dermoscopy Image

Lesion Type	Ratio of Correct 1 <sup>st</sup> Diagnosis	% Correct 1 <sup>st</sup> Diagnosis	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%
<b>Total</b>	<b>117/150</b>	<b>78%</b>	<b>144/150</b>	<b>96%</b>

**Table 2.** Sensitivity, Specificity, and Accuracy of Correct First Diagnosis Per Lesional Dermoscopy Image

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

research provides valuable, up-to-date 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.

**Conflict of Interest Disclosures:** None

**Funding:** None

**Corresponding Author:**

Ryan Scheinkman, MEng

348 Oak Street

Hollywood, FL, USA 33019

Email: [scheinkman@med.miami.edu](mailto:scheinkman@med.miami.edu)

---

**References:**

1. 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
2. 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
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, Malvey 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