JMIR Publications
The Leading eHealth Publisher

Exploring the analysis capabilities and clinical application potential of the Claude 3 Opus in different dermatologic images: the development of a large-scale multimodal model to assist in dermatology clinical practice

Yuwei Huang, Xu Liu, Lu Zhang, Xinyu Zhou, Min-kyu Kim, Long Lin, Xian Jiang

Submitted to: Journal of Medical Internet Research
on: June 12, 2024

Preprint
JMIR Publications

Disclaimer: © The authors. All rights reserved. This is a privileged document currently under peer-review/community review. Authors have provided JMIR Publications with an exclusive license to publish this preprint on it's website for review purposes only. While the final peer-reviewed paper may be licensed under a CC BY license on publication, at this stage authors and publisher expressively prohibit redistribution of this draft paper other than for review purposes.

Table of Contents

Preprint
JMIR Publications

Exploring the analysis capabilities and clinical application potential of the Claude 3 Opus in different dermatologic images: the development of a large-scale multimodal model to assist in dermatology clinical practice

Yuwei Huang[1*] MD; Xu Liu[1*] MD; Lu Zhang[1*] MD; Xinyu Zhou[2] MS; Min-kyu Kim[1] MS; Long Lin[3] BS; Xian Jiang[1] MD, PhD

[1]Department of Dermatology, West China Hospital of Sichuan University Laboratory of Dermatology, Clinical Institute of Inflammation and Immunology, Frontiers Science Center for Disease-related Molecular Network, West China Hospital of Sichuan University Chengdu CN
[2]Department of Dermatology, Nanbu County People's Hospital Nanchong CN
[3]West China School of Medicine, Sichuan University Chengdu CN
[*]these authors contributed equally

Corresponding Author:
Xian Jiang MD, PhD
Department of Dermatology, West China Hospital of Sichuan University
Laboratory of Dermatology, Clinical Institute of Inflammation and Immunology, Frontiers Science Center for Disease-related Molecular Network, West China Hospital of Sichuan University
West China Hospital, Sichuan University, 37 Guoxue Xiang, Wuhou District, Chengdu, Sichuan , China
Chengdu
CN

Abstract

Background: Publicly available, accessible, and user-friendly artificial intelligence is expected to serve in medical processes. Claude3, a newly introduced large-scale multimodal model, has demonstrated significantly superior image analysis capabilities compared to other models in official tests. However, there is no research reporting the potential of Claude3 in medical image analysis.

Objective: To explore the analysis capabilities and potential applications of Claude3 Opus on dermatologic images.

Methods: Dermoscopy and dermatopathology images from textbooks were collected, and question templates for different types of diseases were designed for Claude3 Opus. Three dermatologists used a structured scoring system with five modules to evaluate Claude3 Opus' analysis of images based on recognition, description, completeness, diagnosis, and clinical application, with each module scored out of 5 and a total score of 25.

Results: A total of 330 images were collected. Claude3 Opus scored highest in pigmented disorders dermoscopy images (18.65/25), followed by vascular disorders dermoscopy (15.97/25) and vascular disorders dermatopathology images (15.86/25). In pigmented disorders, its score in dermoscopy (18.65/25) was significantly higher than in dermatopathology images (14.54/25), but no such difference existed in vascular disorders. Among the five modules, Claude3 Opus' recognition score (3.65/5) was significantly higher than the other four modules. There was no difference between description score (3.14/5) and completeness score (3.22/5), but they were significantly higher than the diagnostic score (2.47/5). Claude3 Opus scored higher in malignant diseases than benign diseases, regardless of dermoscopy or dermatopathology images (all p-values <0.05), with no impact from different magnifications in dermatopathology images (p>0.05) and no impact from different number of evaluators.

Conclusions: Claude3 Opus exhibits strong recognition capabilities for dermatologic disease images, can accurately describe abnormalities in images completely, and shows high sensitivity to malignant diseases. Apart from medical assistance, Claude3 Opus could potentially be widely used in medical education and patient communication. Clinical Trial: no need

(JMIR Preprints 12/06/2024:63210)
DOI: https://doi.org/10.2196/preprints.63210

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✓ **Please make my preprint PDF available to anyone at any time (recommended).**
Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
Only make the preprint title and abstract visible.
No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**
Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v
Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

Preprint JMIR Publications

Original Manuscript

Preprint
JMIR Publications

Exploring the analysis capabilities and clinical application potential of the Claude 3 Opus in different dermatologic images: the development of a large-scale multimodal model to assist in dermatology clinical practice

Yuwei Huang,MD[1,2#], Xu Liu,MD[1,2#], Lu Zhang,MD[1,2#], Xinyu Zhou,MS[3], Min-kyu Kim,MS[1,2], Long Lin,BS[1,4], Xian Jiang,MD[1,2✉]

[1] Department of Dermatology, West China Hospital, Sichuan University, Chengdu 610041, China

[2] Laboratory of Dermatology, Clinical Institute of Inflammation and Immunology, Frontiers Science Center for Disease-related Molecular Network, West China Hospital, Sichuan University, Chengdu 610041, China

[3] Department of Dermatology, Nanbu County People's Hospital, Nanchong 637300, China

[4] West China School of Medicine, Sichuan University, Chengdu 610041, China

#These three authors contributed equally to this article.

✉Corresponding Author

Xian Jiang, Department of Dermatology, West China Hospital, Sichuan University, Chengdu 610041, China.

Email: **jiangxian@scu.edu.cn**

Key Words: Claude 3; large-scale multimodal model; dermatology; dermoscopy; dermatopathology

Abstract

Background: Publicly available, accessible, and user-friendly artificial intelligence is expected to serve in medical processes. Claude3, a newly introduced large-scale multimodal model, has demonstrated significantly superior image analysis capabilities compared to other models in official tests. However, there is no research reporting the potential of Claude3 in medical image analysis.

Objective: To explore the analysis capabilities and potential applications of Claude3 Opus on dermatologic images.

Methods: Dermoscopy and dermatopathology images from textbooks were collected, and question templates for different types of diseases were designed for Claude3 Opus. Three dermatologists used a structured scoring system with five modules to evaluate Claude3 Opus' analysis of images based on recognition, description, completeness, diagnosis, and clinical application, with each module scored out of 5 and a total score of 25.

Results: A total of 330 images were collected. Claude3 Opus scored highest in pigmented disorders dermoscopy images (18.65/25), followed by vascular disorders dermoscopy (15.97/25) and vascular disorders dermatopathology images (15.86/25). In pigmented disorders, its score in dermoscopy (18.65/25) was significantly higher than in dermatopathology images (14.54/25), but no such difference existed in vascular disorders. Among the five modules, Claude3 Opus' recognition score (3.65/5) was significantly higher than the other four modules. There was no difference between description score (3.14/5) and completeness score (3.22/5), but they were significantly higher than the diagnostic score (2.47/5). Claude3 Opus scored higher in malignant diseases than benign diseases, regardless of dermoscopy or dermatopathology images (all p-values <0.05), with no impact from different magnifications in dermatopathology images ($p>0.05$) and no impact from different number of evaluators.

Conclusion: Claude3 Opus exhibits strong recognition capabilities for dermatologic disease images, can accurately describe abnormalities in images completely, and shows high sensitivity to malignant diseases. Apart from medical assistance, Claude3 Opus could potentially be widely used in medical education and patient communication.

Preprint JMIR Publications

1. Introduction

Dermatologists often use examinations to identify complex cutaneous diseases beyond visual diagnosis[1]. The results of non-invasive examinations, such as dermoscopy images, and invasive examinations, such as dermatopathology images, are commonly used as evidence to support clinical diagnoses and are crucial for early disease detection[2,3]. Recognition, description, and diagnosis of skin images are crucial in clinical diagnosis and medical student education[4,5]. Due to the complexity and confusion of skin images, accurate assessment often requires experienced dermatologists and misdiagnosis is not uncommon[6–8]. It poses challenges in regions with limited medical and educational resources.

Artificial Intelligence (AI) shows great potential in clinical processes. With the introduction of open-source AI models like ChatGPT and Claude, which primarily function in a question-and-answer format, the widespread and practical application of AI has become achievable in medical practice. Despite the lack of specific training in the medical field, large-language models (LLMs) have demonstrated excellent performance in various medical support tasks[9–11] and diagnosis[12,13]. LLMs even passed the United States Medical Licensing Examination (USMLE)[14]. With the introduction of large AI models incorporating image processing capabilities, the potential of large multimodal models (LMMs) to recognise images and respond to the medical field is gaining attention. ChatGPT 4.0V has been explored for its capabilities in images of multiple diseases[15–19]. However, it does not excel in image recognition capabilities. In March 2024, Anthropic introduced the Claude 3, a novel LMM that surpasses other large AI models including ChatGPT 4.0V in both language and image analysis capabilities[20]. Given the less-than-ideal medical research results regarding ChatGPT 4.0V's image recognition, the potential of Claude 3 in image processing is worth our attention. In official tests and published data, Claude 3 has superior sophisticated vision capabilities and excellent detail-capturing abilities compared to other LMMs. It can handle various image formats, and when it comes to processing science diagrams, it far surpasses ChatGPT 4.0V[20].

As an open-source and user-friendly advanced AI, Claude 3, with its outstanding image processing capabilities, may perform well in the analysis of dermatological images, potentially having certain clinical and educational applications[21]. If Claude 3 can effectively analyze images of cutaneous diseases, especially a wide range of used dermoscopy and dermatopathology images, it could provide valuable assistance for clinical practice and medical education. In this study, we selected the most intelligent Claude 3 Opus model to preliminarily explore its analysis capabilities on dermatopathology images and dermoscopy images and evaluated its strengths, weaknesses, and potential applications.

2. Method

2.1 Image sources and inclusion/exclusion criteria

The dermatopathology images of all available skin diseases used were sourced from "Diagnostic dermatopathology: Nonneoplastic dermatopathology" (ISBN: 9780323798235) and "Diagnostic dermatopathology: Neoplastic dermatopathology" (ISBN: 9780323798266) published by Elsevier Health Sciences. The dermatoscopy images are sourced from "Dermoscopy: The Essentials" (ISBN: 9780702068829) published by Elsevier Health Sciences. The researchers submitted a request to use book images on the marketplace (https://marketplace.copyright.com/rs-ui-web/mp). We had collectively applied for the download and usage rights of 400 images, all of which have been directly extracted from purchased ePub format e-books.

We investigate the skin image analysis capability of Claude 3 Opus to three common clinical features of skin lesions: inflammation abnormalities, pigmentation abnormalities, and vascular abnormalities. The inclusion criteria for images are as follows: i) HE-stained pathological images and dermatoscopic images; ii) diagnosed and described in textbooks; iii) Presence of distinguishable inflammation, pigmentation, or vascular abnormalities in the images; iv) Image size does not exceed the maximum attachment size specified by Claude 3 Opus. Exclusion criteria: i) Images of atypical disease features or rare diseases; ii) Poor image quality; iii) Pathological images with excessively

high magnification leading to indistinguishable structures; iv) Dermatoscopic images alone cannot diagnose or diagnose skin lesions with disputes.

One researcher independently conducted the initial image capture. Subsequently, a second researcher evaluated the image quality, excluding all images with questionable quality.

2.2 Interact with Claude 3 Opus for image recognition

We interacted with Claude 3 Opus through the official website (https://claude.ai). Referring to previously published medical literature on testing the image interpretation ability of LMMs, we have developed specific question templates for different types of diseases in Claude 3 Opus input[17,22–24]. A new dialogue was created after each answer to proceed with the next question. In all cases, we ensure that the question template is always accompanied by the image for input. All question statements can be found in Supplementary Table 1.

2.3 Establishment of evaluation criteria

To quantify and provide an objective assessment, a structured scoring system consisting of 5 modules was employed. Five modules evaluated the recognition ability, description ability, diagnosis ability, completeness, and clinical application potential of Claude 3 Opus on dermatological images. Each module was scored out of 5, amounting to 25 points. The specific scoring criteria are detailed in Supplementary Table 2. The scoring criteria were established based on image descriptions from the selected textbook and a literature review on dermatopathology and dermoscopy images[25–37]. Two highly experienced dermatology experts with over 10 years of practice have reviewed and revised the scoring criteria.

2.4 Selection of evaluators and outcome evaluation

Three dermatologists were invited as evaluators to provide a structured evaluation of the descriptions of dermatological images, based on the reference material from the source of the pictures. The selection criteria for evaluator 1 and evaluator 2 include: i) holding a Master's degree (MS) or above in dermatology; ii) having at least 5 years of clinical work experience; iii) having at least 5 years of experience in dermatology clinics, being proficient in analysing dermoscopy images and dermatopathology images of common skin diseases. Evaluator 3 is a highly experienced dermatologist with a PhD degree and over 10 years of work experience. Before the evaluations, all three evaluators underwent scoring training and completed simulated training papers for structured evaluations.

Independent scores were obtained from evaluator 1 and evaluator 2, resulting in score 1 and score 2. If the difference between score 1 and score 2 in any evaluation module of any image was ≥2 points, a third dermatologist will conduct a reassessment, resulting in score 3. In any evaluation module, scores differing by ≥2 points from the score 3 will be discarded. The remaining scores will be averaged with the score 3 to determine the final score of a single image. If there is a difference of ≥2 points between the three scores in the module, the average of the three scores will be taken as the score for that imaging module. The total score for each module was calculated by summing the scores and dividing by the number of images to obtain the average score for 5 modules. The sum of the 5 module scores was then calculated as the total score for that type of image.

2.5 Data analyses

Use one-way analysis of variance to analyze quantitative variables between three or more groups, with homogeneity of variance testing conducted before the analysis of variance. For non-normally distributed data, use the Kruskal-Wallis test for comparison. Pairwise comparisons will be conducted after the analysis of variance or the Kruskal-Wallis test. All statistical analyses were performed using R version 4.3.0 (R Project for Statistical Computing), and all tests considered differences statistically significant at $P < 0.05$. See Figure 1 for the research process.

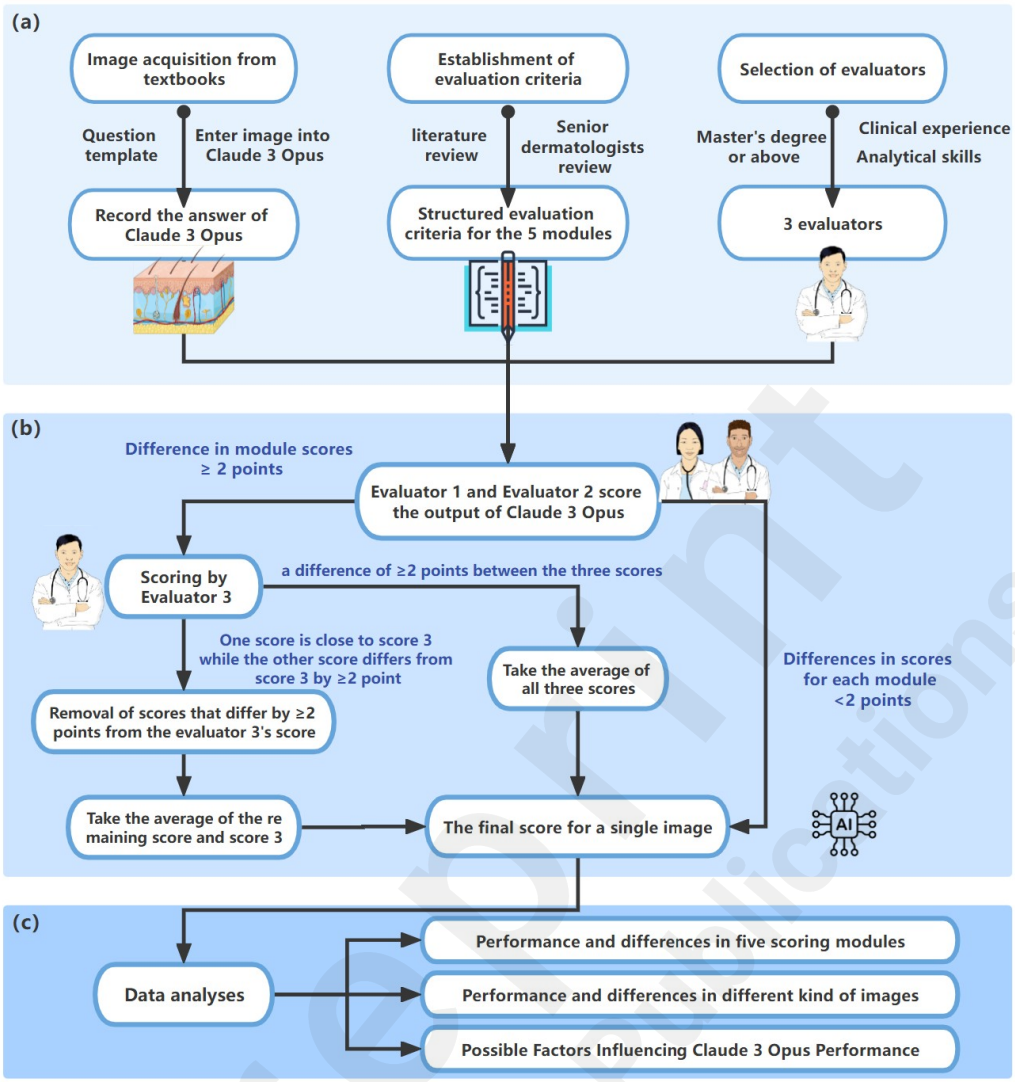

Figure 1 Flowchart of the research process. (a) Image acquisition, establishment of evaluation criteria, and selection of evaluators; (b) Scoring process and determination of the final score for each image; (c) Data analysis.

3. Results

3.1 Performance and differences of Claude 3 Opus in five scoring modules.

Based on the inclusion and exclusion criteria, we included a total of 330 clinical images of skin diseases, including 230 dermatopathology images (100 inflammatory skin disorders, 80 pigmentary skin disorders, 50 vascular skin disorders), and 100 dermatoscopy images (50 pigmentary skin disorders, 50 vascular skin disorders). All images were inputted into Claude 3 Opus along with the corresponding question template, and after receiving answers, they were evaluated by three selected dermatologists. Image collection results and responses of Claude 3 Opus to images can be found in Supplementary Table 3. In terms of scoring results, the independent scores of Evaluator 1 and Evaluator 2 were similar, and there were a total of 38 images where there was a difference of ≥2 points in one of the module scores that required rescoring by Evaluator 3. In all images rescored by Evaluator 3, there were no cases where all three scores differed by ≥2 points, requiring the average of all three to be taken as the final score.

To investigate the analysis capability of Claude 3 Opus on clinical skin images, we established five different modules to evaluate its performance in recognition, description, analysis completeness, diagnostic ability, and clinical application capability. Claude 3 Opus showed variations in performance across the five modules. Figure 2a displays the score distribution of 330 clinical skin

images in different modules. Among them, the average recognition score (3.65/5) was significantly higher than the scores of the other four modules, followed by the average description score (3.14/5) and the average completeness score (3.22/5), with no statistical difference between the two. The average diagnostic score (2.47/5) was the lowest among all modules (Figure 2b).

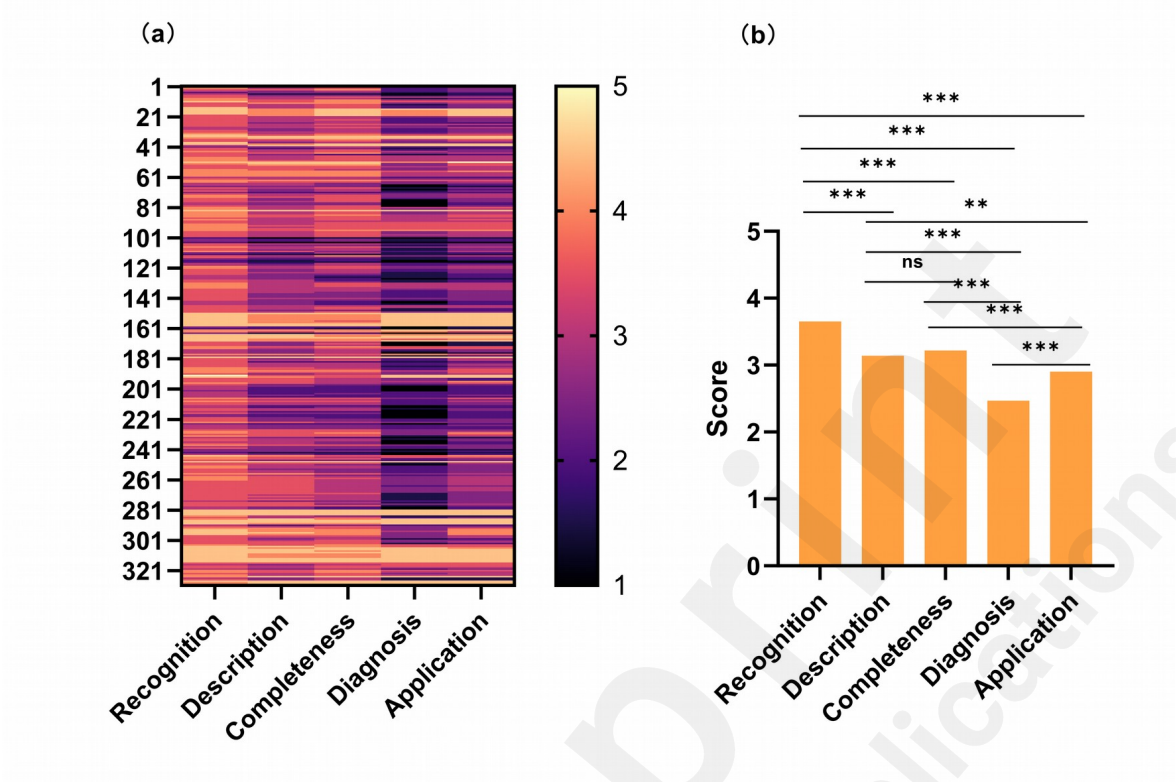

Figure 2 Compare the performance of Claude 3 Opus in five different modules. (a) Distribution of scores for Claude 3 Opus' image analysis ability in different modules(n=330). (b) Differences in scores for Claude 3 Opus across five different modules(n=330).
*ns: no statistical difference; *: p<0.05; **: p<0.01, ***: p<0.001.*

3.2 Scores of Claude 3 Opus in different skin disease images

The performance of Claude 3 opus in different types of diseases is shown in Table 1. Among dermatopathology images, Claude 3 opus performs best in vascular disorders (15.86/25), followed by pigmentary disorders (14.54/25), with the lowest score in inflammatory disorders (13.83/25) (*p<0.001*). Apart from the recognition part, the scores of the remaining four modules in vascular disorders are superior to those of inflammatory and pigmentary disorders. In dermoscopy images, Claude 3 Opus performs exceptionally well in images of pigmentary diseases (18.65/25), with all five modules scoring higher than in vascular skin diseases (15.97/25) (all *p-value <0.05*).

We compared the performance of Claude 3 Opus between dermoscopy images and dermatopathology images (Figure 3). In pigmentary diseases, Claude 3 Opus scored significantly higher in dermoscopy images across five different modules compared to dermatopathology images (all *p-values <0.001*). However, this difference was only observed in the recognition module in vascular diseases, with no significant statistical differences in scores in the other 4 modules.

Table 1 Score of performance of Claude 3 Opus in different types of skin disease images.

	Recognition Score (5)	Description Score (5)	Completeness score (5)	Diagnosis Score (5)	Application Score (5)	Overall score (25)
Dermatopathology Images						
Inflammatory skin disorders	3.76	2.98	2.82	1.61	2.62	13.83

(n=100)							
Pigmented skin (n=80)	disorders	3.81	2.9	3.01	2.39	2.73	14.54
Vascular skin (n=50)	disorders	3.88	3.22	3.44	2.54	3	15.86
p-value		0.343	*0.02*	*□0.001*	*□0.001*	*0.004*	*□0.001*
Average		3.8	3.01	3.02	2.08	2.74	14.52
Dermoscopy Images							
Pigmented skin (n=50)	disorders	4.02	3.75	3.86	3.41	3.61	18.65
Vascular skin disorders(n=50)		3.69	3.17	3.51	2.65	2.95	15.97
p-value		*0.003*	*□0.001*	*0.007*	*□0.001*	*□0.001*	*□0.001*
Average		3.86	3.46	3.69	3.03	3.28	17.31

Note: Statistically significant *p-values* are in *italics.*

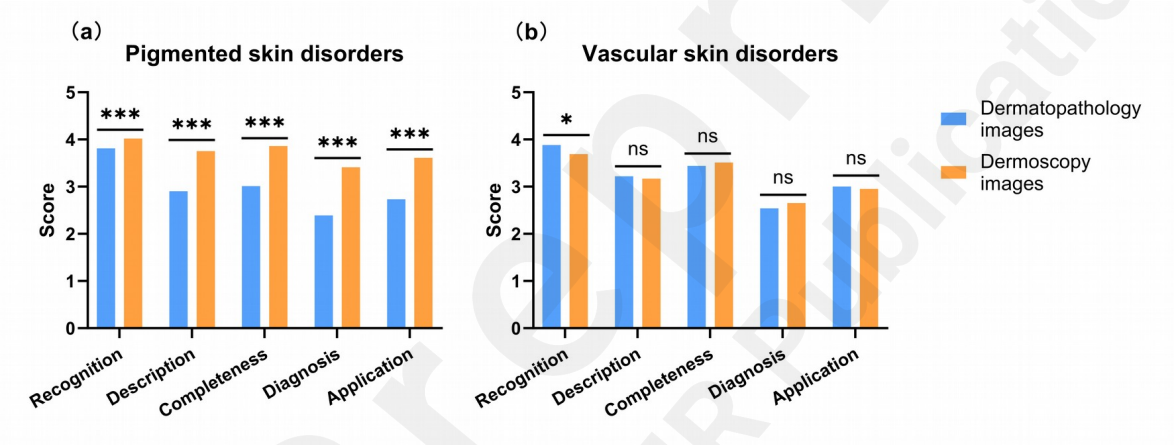

Figure 3 Comparison of the performance of Claude 3 Opus in dermoscopy images and dermatopathology images: (a) Differences in scores between dermoscopy images (n=50) and dermatopathology images (n=80) of pigmentary skin disorders. (b) Differences in scores between dermoscopy images (n=50) and dermatopathology images (n=50) of vascular skin disorders.
*ns: no statistical difference; *: p<0.05; **: p<0.01; ***: p<0.001.*

3.3 Possible Factors Influencing Claude 3 Opus Performance

To explore the potential factors influencing the performance of Claude 3 Opus, we differentiated between images diagnosed as malignant skin diseases (such as melanoma, skin squamous cell carcinoma, basal cell carcinoma, etc.) and benign skin diseases (such as nevus, seborrheic keratosis, etc.) in dermoscopy and dermatopathology images. The results indicate that Claude 3 Opus performs better in dermoscopy images of malignant diseases compared to benign diseases, showing significant differences across all five scoring modules (Figure 4a). Similarly, these differences were observed in dermatopathology images, with Claude 3 Opus scoring higher in all five modules for malignant diseases compared to benign diseases (Figure 4b).

We also differentiated between high-power and low-power images in the pathological images. The results showed no difference in the performance of Claude 3 Opus in images at different magnifications, with no statistically significant differences found in the 5 modules (Figure 4c). This indicates that its ability to recognize images is minimally affected by the image's perspective,

resulting in stable recognition capability. Similarly, we compared the scores of 38 images where a third evaluator was introduced to provide a score when the difference between the scores of evaluator 1 and evaluator 2 was two points or more in a certain module, with the image scores from two evaluators. The results indicate that there is no significant statistical difference in any module, suggesting the robustness of the scoring results (Figure 4d).

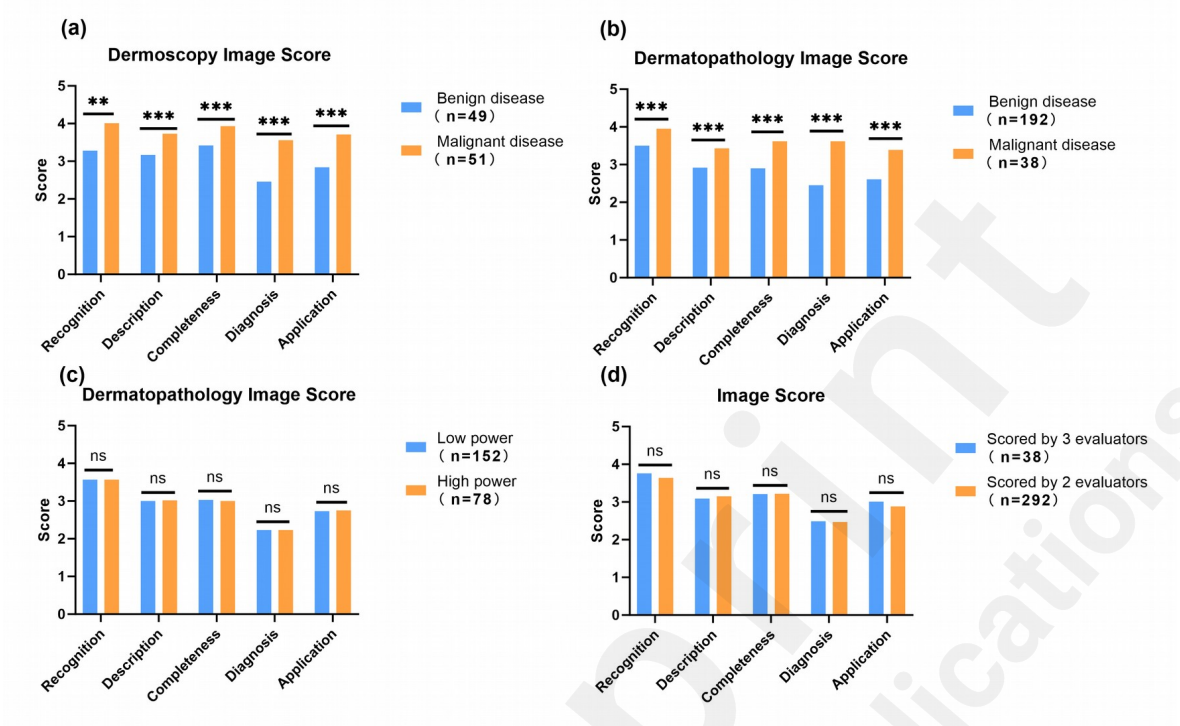

Figure 4 Factors that may affect the image analysis capability of Claude 3 Opus. (a) Score differences of Claude 3 Opus in dermoscopy images for benign and malignant diseases (b) Score differences of Claude 3 Opus in dermatopathology images for benign and malignant diseases (c) Score differences of Claude 3 Opus in dermatopathology images for low power and high power magnification views. (d) Score differences of Claude 3 Opus in images for scored by 3 evaluators and scored by 2 evaluators.
*ns: No statistical difference; *: p<0.05; **: p<0.01, ***: p<0.001.*

4. Discussion

In this study, we collected dermatopathology and dermoscopy images to investigate the analysis capabilities of Claude 3 Opus on clinically common skin disease images. Out of 330 dermatopathology or dermoscopy images, using 15 points (60%) as the "pass" standard out of a total of 25 points, Claude 3 Opus achieved an overall pass rate of 56.4%. The pass rate for dermoscopy images was 70%, while for dermatopathology images it was 50.4%. Claude 3 Opus exhibited different performances in various disease categories. In dermatopathology images, it excelled in vascular abnormalities, whereas in dermoscopy images, pigment abnormalities showed the best performance. Claude 3 Opus scored significantly higher in pigment skin disease dermoscopy images compared to other types of images in our study, especially in the analysis of malignant diseases, achieving high scores. The scores of Claude 3 Opus in different modules we set up indicate its good ability to recognize image types and contents, and it can describe abnormalities well, but there may still be cases of missing important features or incorrect descriptions. Despite scoring the lowest in diagnostic scoring among the five modules, even professional dermatologists find it challenging to make completely accurate diagnoses based solely on a single dermoscopy or dermatopathology image in the absence of clinical information about the patient. Therefore, our study in terms of diagnosis can only be considered an attempt, and we cannot conclude that Claude 3 Opus lacks

diagnostic capabilities for skin images. Building on the exploration in this study, our next phase aims to integrate patients' clinical information and medical history features using Claude 3 Opus. We will assess whether Claude 3 Opus enhances the diagnostic capability of skin disease images when clinical information is available.

This study is the first to focus on the application of Claude 3 in medical imaging. Previous research by Shifai et al[19]. evaluated the ability of ChatGPT 4.0V to recognize melanoma in dermoscopic images, with a correct diagnosis rate of 54.7% (ranking among the top three), a sensitivity of 46% for melanoma diagnosis (ranking first), or 78% (ranking among the top three). In this study, Claude 3 Opus achieved an average score of 3.41 in pigmented skin disease dermoscopy images. Out of 50 images, 27 were diagnosed correctly (54%), similar to Shifai's findings. However, in 21 melanoma images, Claude 3 Opus provided a definitive diagnosis of melanoma in 18 images (85.7%), with suspected or considered melanoma diagnosis in the remaining 3 images. This suggests that Claude 3 Opus may be more sensitive in analysing melanoma, as it consistently scored higher for malignant skin diseases in both dermatopathology and dermoscopy images. Claude 3 Opus appears to be more sensitive to malignant diseases and tends to interpret diseases as malignant when in doubt, possibly due to a wider range of malignant disease data sources. In another unpublished study within our team, we compared the diagnostic differences of ChatGPT-4V and Claude 3 Opus in melanoma dermoscopy images, showing that Claude 3 Opus indeed exhibits a stronger superiority in melanoma diagnosis.

The ability of LMM to handle the analysis of clinical images has been a concern for many clinical professionals recently. Previous studies on ChatGPT 4.0V have shown that it may not be suitable for clinical work. For example, Zhang et al[18]. found that in liver CT images, ChatGPT-4V can hardly recognize the correct anatomical structures; Haver et al[17]. discovered that ChatGPT-4V had an incorrect rate of up to 68.9% in describing mammography image Features; Miao et al[38]. observed that ChatGPT 4.0V only achieved a 29% accuracy in directly asking questions about kidney dermatopathology images. Given that Claude 3 Opus has demonstrated significantly higher image analysis capabilities than other models in official tests, we look forward to more research exploring the clinical application potential of Claude 3 Opus in various clinical fields. In this study, we found some undeniable advantages of Claude 3 Opus in interpreting clinical skin images. For instance, Claude 3 Opus can accurately identify whether an image is a dermatopathology image or a dermoscopy image, recognizing the epidermis, dermis, and basal layer in skin dermatopathology images, and identifying epidermal keratinization. It also provided comprehensive descriptions of image abnormalities, describing the locations of abnormalities in most images and using common terms such as "blue-white structures" and "dot-like vasculature pattern". Although Claude 3 Opus's diagnostic performance in our results was mediocre, based on the scores for recognition, description, and completeness, we believe it can identify and describe skin pathological diseases and dermoscopy diseases.

Interpreting clinical images and their diagnostic implications requires specific domain knowledge, an area where LMMs not have been trained. AI models specially trained have shown remarkable performance in image analysis. Convolutional neural networks (CNN) in deep learning demonstrate skin cancer classification abilities comparable to clinical professionals[39,40]. Algorithms developed by the International Skin Imaging Collaboration (ISIC) can match the expertise of professional dermatologists in simulated tests[41,42]. However, data biases in AI training models can affect the universality across different ethnicities and socio-economic populations, leading to algorithmic biases and health disparities to some extent[43–46]. For example, the diagnostic accuracy of ISIC does not include individuals with Fitzpatrick phototype III or higher[47]. Additionally, limited understanding of training algorithms by patients or the public can hinder the trust and usage of artificial intelligence. Claude 3 Opus, due to its accessibility to the public, can partially alleviate patient distrust towards AI and its extensive data sources contribute to a more balanced performance in health equity[48–50]. Based on our research findings, we believe Claude 3 Opus holds value in assisting

physicians in clinical practice, and due to its accessibility, it may also have broader potential applications in medical education and patient communication.

Our study evaluated the analysis potential of Claude 3 Opus in dermatological images. The strength of our study lies in the early assessment of Claude 3 Opus' ability to interpret dermatological images across multiple dimensions and a comprehensive evaluation of its strengths and limitations. However, as a pioneering study, we only tested Claude 3 Opus' performance on a limited number of images and did not apply it to clinical practice. Furthermore, in the selection of diseases, we only roughly classified skin diseases, which is insufficient in disciplines with a wide variety of symptoms and disease types. In conclusion, this pilot study found Claude 3 Opus has the potential to identify dermatological images, particularly performing well in dermoscopy images of pigmentary skin disorders. Claude 3 Opus has advantages in the identification, description, and diagnosis of malignant skin diseases, and its image recognition and description capabilities have the potential to be applied in clinical assistance, medical education, and patient communication.

Acknowledgment

We would like to express our gratitude to all the authors and image providers of "Diagnostic dermatopathology: Nonneoplastic dermatopathology" (ISBN: 9780323798235), "Diagnostic dermatopathology: Neoplastic dermatopathology" (ISBN: 9780323798266), and "Dermoscopy: The Essentials" (ISBN: 9780702068829).

Conflict of Interest Statement

The authors have no conflict of interest to declare.

Funding Sources

This work was supported by the National Natural Science Foundation of China (Grant No. 82273559 and 82304052), and the Department of Science and Technology of Sichuan Province (Grant No. 2022YFQ0054 and 24NSFSC6130).

Data Availability statement

The article's data will be shared on reasonable request to the corresponding author.

Author contributions statement

Acquisition of data: Yuwei Huang and Xu Liu; Analysis and interpretation of data: All authors; Drafting of the manuscript: Yuwei Huang and Xu Liu; Critical revision of the manuscript for important intellectual content: Lu Zhang; Obtained funding: Lu Zhang and Xian Jiang; Study supervision: Xian Jiang.

References

1. Sonthalia S, Yumeen S, Kaliyadan F. Dermoscopy Overview and Extradiagnostic Applications. In: *StatPearls*. StatPearls Publishing; 2024. Accessed April 26, 2024. http://www.ncbi.nlm.nih.gov/books/NBK537131/

2. Micali G, Verzì AE, Lacarrubba F. Alternative uses of dermoscopy in daily clinical practice: An update. *J Am Acad Dermatol*. 2018;79(6):1117-1132.e1. doi:10.1016/j.jaad.2018.06.021

3. Lehman JS, Erickson LA. Emerging concepts in dermatopathology: a special issue of neoplastic, inflammatory, and special-site dermatopathology and important practice considerations. *Hum Pathol*. 2023;140:1-4. doi:10.1016/j.humpath.2023.07.012

4. Yélamos O, Braun RP, Liopyris K, et al. Dermoscopy and dermatopathology correlates of

cutaneous neoplasms. *J Am Acad Dermatol*. 2019;80(2):341-363. doi:10.1016/j.jaad.2018.07.073

5. Boespflug A, Guerra J, Dalle S, Thomas L. Enhancement of Customary Dermoscopy Education With Spaced Education e-Learning: A Prospective Controlled Trial. *JAMA Dermatol*. 2015;151(8):847-853. doi:10.1001/jamadermatol.2015.0214

6. Janowska A, Oranges T, Iannone M, Romanelli M, Dini V. Seborrheic keratosis-like melanoma: a diagnostic challenge. *Melanoma Res*. 2021;31(5):407-412. doi:10.1097/CMR.0000000000000756

7. Papageorgiou V, Apalla Z, Sotiriou E, et al. The limitations of dermoscopy: false-positive and false-negative tumours. *J Eur Acad Dermatol Venereol JEADV*. 2018;32(6):879-888. doi:10.1111/jdv.14782

8. Elder DE, Eguchi MM, Barnhill RL, et al. Diagnostic error, uncertainty, and overdiagnosis in melanoma. *Pathology (Phila)*. 2023;55(2):206-213. doi:10.1016/j.pathol.2022.12.345

9. Trager MH, Queen D, Bordone LA, Geskin LJ, Samie FH. Assessing ChatGPT responses to common patient queries regarding basal cell carcinoma. *Arch Dermatol Res*. 2023;315(10):2979-2981. doi:10.1007/s00403-023-02705-3

10. Yan S, Du D, Liu X, et al. Assessment of the Reliability and Clinical Applicability of ChatGPT's Responses to Patients' Common Queries About Rosacea. *Patient Prefer Adherence*. 2024;18:249-253. doi:10.2147/PPA.S444928

11. Szczesniewski JJ, Tellez Fouz C, Ramos Alba A, Diaz Goizueta FJ, García Tello A, Llanes González L. ChatGPT and most frequent urological diseases: analysing the quality of information and potential risks for patients. *World J Urol*. 2023;41(11):3149-3153. doi:10.1007/s00345-023-04563-0

12. Brancaccio G, Balato A, Malvehy J, Puig S, Argenziano G, Kittler H. Artificial Intelligence in Skin Cancer Diagnosis: A Reality Check. *J Invest Dermatol*. 2024;144(3):492-499. doi:10.1016/j.jid.2023.10.004

13. Fonseca Â, Ferreira A, Ribeiro L, Moreira S, Duque C. Embracing the future-is artificial intelligence already better? A comparative study of artificial intelligence performance in diagnostic accuracy and decision-making. *Eur J Neurol*. 2024;31(4):e16195. doi:10.1111/ene.16195

14. Mihalache A, Huang RS, Popovic MM, Muni RH. ChatGPT-4: An assessment of an upgraded artificial intelligence chatbot in the United States Medical Licensing Examination. *Med Teach*. 2024;46(3):366-372. doi:10.1080/0142159X.2023.2249588

15. Laohawetwanit T, Namboonlue C, Apornvirat S. Accuracy of GPT-4 in histopathological image detection and classification of colorectal adenomas. *J Clin Pathol*. Published online January 10, 2024:jcp-2023-209304. doi:10.1136/jcp-2023-209304

16. Zhu L, Mou W, Lai Y, et al. Step into the era of large multimodal models: A pilot study on ChatGPT-4V(ision)'s ability to interpret radiological images. *Int J Surg Lond Engl*. Published online March 18, 2024. doi:10.1097/JS9.0000000000001359

17. Haver HL, Bahl M, Doo FX, et al. Evaluation of Multimodal ChatGPT (GPT-4V) in Describing Mammography Image Features. *Can Assoc Radiol J J Assoc Can Radiol*. Published online April 6, 2024:8465371241247043. doi:10.1177/08465371241247043

18. Zhang Y, Wu L, Wang Y, et al. Unexpectedly low accuracy of GPT-4 in identifying common liver diseases from CT scan images. *Dig Liver Dis Off J Ital Soc Gastroenterol Ital Assoc Study Liver*. 2024;56(4):718-720. doi:10.1016/j.dld.2024.01.191

19. Shifai N, van Doorn R, Malvehy J, Sangers TE. Can ChatGPT vision diagnose melanoma? An exploratory diagnostic accuracy study. *J Am Acad Dermatol*. Published online January 19, 2024:S0190-9622(24)00076-8. doi:10.1016/j.jaad.2023.12.062

20. Introducing the next generation of Claude. Accessed April 8, 2024. https://www.anthropic.com/news/claude-3-family

21. Ngo TB, Niu W, Fang Z, Gold L. Dermatology residents' perspectives on virtual dermatopathology education. *J Cutan Pathol*. Published online March 28, 2024. doi:10.1111/cup.14618

22. Jmm R, Hv J, G M, et al. Keeping Up With ChatGPT: Evaluating Its Recognition and Interpretation of Nuclear Medicine Images. *Clin Nucl Med*. Published online January 4, 2024. doi:10.1097/RLU.0000000000005207

23. Zhu L, Lai Y, Ta N, Cheng L, Chen R. Multimodal Approach in the Diagnosis of Urologic Malignancies: Critical Assessment of ChatGPT-4V's Image-Reading Capabilities. *JCO Clin Cancer Inform*. 2024;8:e2300275. doi:10.1200/CCI.23.00275

24. Busch F, Han T, Makowski MR, Truhn D, Bressem KK, Adams L. Integrating Text and Image Analysis: Exploring GPT-4V's Capabilities in Advanced Radiological Applications Across Subspecialties. *J Med Internet Res*. 2024;26:e54948. doi:10.2196/54948

25. Argenziano G, Soyer HP. Dermoscopy of pigmented skin lesions--a valuable tool for early diagnosis of melanoma. *Lancet Oncol*. 2001;2(7):443-449. doi:10.1016/s1470-2045(00)00422-8

26. Aung PP, Mutyambizi KK, Danialan R, Ivan D, Prieto VG. Differential diagnosis of heavily pigmented melanocytic lesions: challenges and diagnostic approach. *J Clin Pathol*. 2015;68(12):963-970. doi:10.1136/jclinpath-2015-202887

27. Carlson JA. The histological assessment of cutaneous vasculitis. *Histopathology*. 2010;56(1):3-23. doi:10.1111/j.1365-2559.2009.03443.x

28. Elston DM, Stratman EJ, Miller SJ. Skin biopsy: Biopsy issues in specific diseases. *J Am Acad Dermatol*. 2016;74(1):1-16; quiz 17-18. doi:10.1016/j.jaad.2015.06.033

29. González SB. Histopathologic diagnosis of pigmented lesions of the skin. *Pathol Res Pract*. 1991;187(4):387-431. doi:10.1016/S0344-0338(11)80002-6

30. J L, A von K, J S. Dermatopathology 101: Part 1 - Inflammatory skin diseases. *J Dtsch Dermatol Ges J Ger Soc Dermatol JDDG*. 2017;15(1). doi:10.1111/ddg.13176

31. Lallas A, Giacomel J, Argenziano G, et al. Dermoscopy in general dermatology: practical tips for the clinician. *Br J Dermatol*. 2014;170(3):514-526. doi:10.1111/bjd.12685

32. Uzundere C, Akay BN, Heper AO, Clark SP, Tschandl P, Kittler H. Integrated diagnosis of nevi with severely atypical features and impact of second opinions. *J Eur Acad Dermatol Venereol JEADV*. 2023;37(6):1184-1189. doi:10.1111/jdv.18996

33. Wozniak-Rito A, Zalaudek I, Rudnicka L. Dermoscopy of basal cell carcinoma. *Clin Exp Dermatol*. 2018;43(3):241-247. doi:10.1111/ced.13387

34. Xie F, Fan H, Li Y, Jiang Z, Meng R, Bovik A. Melanoma Classification on Dermoscopy Images Using a Neural Network Ensemble Model. *IEEE Trans Med Imaging*. 2017;36(3):849-858. doi:10.1109/TMI.2016.2633551

35. Fu X, Bi L, Kumar A, Fulham M, Kim J. Graph-Based Intercategory and Intermodality Network for Multilabel Classification and Melanoma Diagnosis of Skin Lesions in Dermoscopy and Clinical Images. *IEEE Trans Med Imaging*. 2022;41(11):3266-3277. doi:10.1109/TMI.2022.3181694

36. Brodsky V, Levine L, Solans EP, Dola S, Chervony L, Polak S. Performance of Automated Classification of Diagnostic Entities in Dermatopathology Validated on Multisite Data Representing the Real-World Variability of Pathology Workload. *Arch Pathol Lab Med*. 2023;147(9):1093-1098. doi:10.5858/arpa.2021-0550-OA

37. Koch EAT, Erdmann M, Berking C, et al. Standardized Computer-Assisted Analysis of PRAME Immunoreactivity in Dysplastic Nevi and Superficial Spreading Melanomas. *Int J Mol Sci*. 2023;24(7):6388. doi:10.3390/ijms24076388

38. Miao J, Thongprayoon C, Cheungpasitporn W, Cornell LD. Performance of GPT-4 Vision on kidney pathology exam questions. *Am J Clin Pathol*. Published online April 3, 2024:aqae030. doi:10.1093/ajcp/aqae030

39. Haggenmüller S, Maron RC, Hekler A, et al. Skin cancer classification via convolutional neural networks: systematic review of studies involving human experts. *Eur J Cancer Oxf Engl 1990*. 2021;156:202-216. doi:10.1016/j.ejca.2021.06.049

40. Goessinger EV, Cerminara SE, Mueller AM, et al. Consistency of convolutional neural networks in dermoscopic melanoma recognition: A prospective real-world study about the pitfalls of augmented intelligence. *J Eur Acad Dermatol Venereol JEADV*. 2024;38(5):945-953. doi:10.1111/jdv.19777

41. Menzies SW, Sinz C, Menzies M, et al. Comparison of humans versus mobile phone-powered artificial intelligence for the diagnosis and management of pigmented skin cancer in secondary care: a multicentre, prospective, diagnostic, clinical trial. *Lancet Digit Health*. 2023;5(10):e679-e691. doi:10.1016/S2589-7500(23)00130-9

42. Mehta PP, Sun M, Betz-Stablein B, et al. Improving Artificial Intelligence-Based Diagnosis on Pediatric Skin Lesions. *J Invest Dermatol*. 2023;143(8):1423-1429.e1. doi:10.1016/j.jid.2022.08.058

43. Chen RJ, Wang JJ, Williamson DFK, et al. Algorithmic fairness in artificial intelligence for medicine and healthcare. *Nat Biomed Eng*. 2023;7(6):719-742. doi:10.1038/s41551-023-01056-8

44. Carter SM, Aquino YSJ, Carolan L, et al. How should artificial intelligence be used in Australian

health care? Recommendations from a citizens' jury. *Med J Aust*. 2024;220(8):409-416. doi:10.5694/mja2.52283

45. Tejani AS, Ng YS, Xi Y, Rayan JC. Understanding and Mitigating Bias in Imaging Artificial Intelligence. *Radiogr Rev Publ Radiol Soc N Am Inc*. 2024;44(5):e230067. doi:10.1148/rg.230067

46. Schaekermann M, Spitz T, Pyles M, et al. Health equity assessment of machine learning performance (HEAL): a framework and dermatology AI model case study. *EClinicalMedicine*. 2024;70:102479. doi:10.1016/j.eclinm.2024.102479

47. MacLellan AN, Price EL, Publicover-Brouwer P, et al. The use of noninvasive imaging techniques in the diagnosis of melanoma: a prospective diagnostic accuracy study. *J Am Acad Dermatol*. 2021;85(2):353-359. doi:10.1016/j.jaad.2020.04.019

48. Goldust M. Utilizing AI to address skin disorders and healthcare disparities among undocumented immigrants. *Int J Dermatol*. Published online April 22, 2024. doi:10.1111/ijd.17208

49. Elani HW, Giannobile WV. Harnessing Artificial Intelligence to Address Oral Health Disparities. *JAMA Health Forum*. 2024;5(4):e240642. doi:10.1001/jamahealthforum.2024.0642

50. Chen F, Wang L, Hong J, Jiang J, Zhou L. Unmasking bias in artificial intelligence: a systematic review of bias detection and mitigation strategies in electronic health record-based models. *J Am Med Inform Assoc JAMIA*. 2024;31(5):1172-1183. doi:10.1093/jamia/ocae060

Supplementary Table 1: Question Templates for Claude 3 Opus based on Different Types of Images
Supplementary Table 2：Structured Evaluation Criteria for Image Analysis Results Generated by Claude 3 Opus
Supplementary Table 3: Results of Image Acquisition and Responses from Claude 3 Opus to the Images

Supplementary Files

Preprint JMIR Publications