Original Article Evaluating ChatGPT's diagnostic accuracy in skin diseases based on images

Rizwan Ali^{1,2}, Haiyan Cui^{1,2}

¹Department of Plastic and Cosmetic Surgery, Tongji Hospital, School of Medicine, Tongji University, Shanghai 200092, China; ²Institute of Aesthetic Plastic Surgery and Medicine, School of Medicine, Tongji University, Shanghai 200092, China

Received December 23, 2024; Accepted June 12, 2025; Epub July 15, 2025; Published July 30, 2025

Abstract: Objectives: This study aims to evaluate the performance of Dr. FerManda, a custom ChatGPT Al-based model, in diagnosing 42 common skin diseases. The focus is on assessing its diagnostic accuracy and the potential for Al-assisted dermatology without relying on detailed patient information like age, sex, or symptoms. Methods: The Dr. FerManda model was trained using publicly available image datasets and clinical literature related to dermatological conditions. Its diagnostic accuracy was tested across various skin diseases and compared against evaluations by expert dermatologists. Additionally, the model provided descriptions of symptoms, causes, and treatment options for each diagnosed condition. Results: The model achieved 100% accuracy across test cases, although it initially misdiagnosed two diseases; these errors were corrected following further guidance. It also delivered detailed and accurate information on each condition, aligning closely with expert dermatologists' assessments regarding symptoms, causes, and treatment recommendations. Conclusions: These findings indicate that Al, particularly custom ChatGPT models like Dr. FerManda, holds great promise for improving dermatological diagnostics. With its high accuracy and rapid response times, Al could significantly enhance diagnostic support in dermatology, paving the way for broader applications and future research to expand its capabilities.

Keywords: Skin diseases, custom ChatGPT model, Dr. FerManda AI model, image-based diagnosis, future research

Introduction

ChatGPT is a natural language processing (NLP) artificial intelligence (AI) chatbot that generates responses based on both text and image inputs. It functions as an accessible online interface, analyzing data from the internet to provide relevant, tailored answers to user inquiries. ChatGPT has shown significant potential across various fields, including clinical practice [1]. An increasing number of studies have explored its applications in the medical domain [2, 3], with many concluding that ChatGPT can serve as an effective complementary tool for clinicians [4].

Skin diseases present a significant global health challenge, ranking as the fourth most common cause of illness worldwide [5]. Although these conditions are more prevalent among the elderly and men, they can affect

individuals of all ages [6]. Neglecting skin diseases is not an option; proactive management is essential [7]. According to the World Health Organization (WHO), over one billion people in developing countries suffer from one or more skin diseases [8]. Therefore, early identification of skin conditions is critical to prevent their progression [7, 9]. Clinical diagnosis typically involves multiple healthcare professionals and can be time-consuming and costly due to expensive laboratory fees. It is also prone to human error, which complicates reproducibility [10, 11]. Leveraging current technologies for early detection of skin diseases may help facilitate more cost-effective diagnostics [7]. By integrating real-time image analysis with both objective and subjective metrics, ChatGPT can provide more personalized and precise evaluations [12]. Timely detection is crucial, as many skin conditions can worsen rapidly, leading to potentially life-threatening complications. For

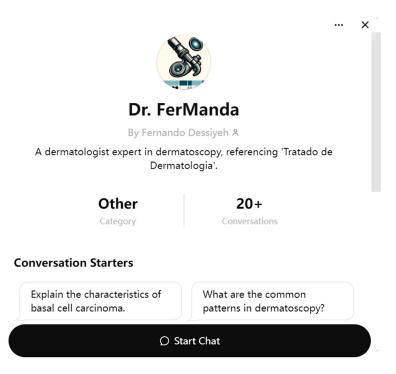


Figure 1. Overview of the custom ChatGPT Dr. FerManda Al model interface. The model diagnoses skin diseases by analyzing image-based inputs and comparing visual data against a comprehensive dataset of dermatological references. This figure displays the homepage of the Dr. FerManda model, where users can upload images for diagnostic purposes.

example, skin cancers such as melanoma can progress quickly if not diagnosed and treated early, increasing the risk of metastasis and death. Similarly, infections like cellulitis can spread fast, potentially causing sepsis - a severe systemic response that can be fatal. Untreated chronic skin conditions may also result in serious health problems, affecting vital organs and overall well-being. Early recognition of symptoms and prompt medical attention can significantly reduce the risk of severe outcomes and improve the chances of effective treatment and recovery. With advances in Al technology within the medical field, there is growing demand for systems capable of diagnosing common skin diseases from images. However, research in this area remains limited, with many studies still in the early stages of developing diagnostic tools for skin conditions.

Dr. FerManda is a custom-built artificial intelligence (AI) model based on the ChatGPT architecture, specifically designed to assist in dermatological diagnostics. The model leverages the natural language processing (NLP) capabilities of ChatGPT to analyze and interpret der-

matological conditions using both clinical literature and image datasets. Unlike traditional AI systems that require detailed patient-specific information (e.g., age, sex, symptoms), Dr. FerManda primarily relies on visual data and clinical knowledge, making it versatile for use without extensive patient profiles. It processes images of skin conditions and generates diagnostic results by matching visual cues against a vast database of medical references, correlating patterns and symptoms with a wide range of known dermatological diseases. Additionally, Dr. FerManda provides detailed explanations for each condition, including common symptoms, potential causes, and recommended treatment options. It does this by drawing on its pretrained language model to produce informative, coher-

ent text, similar to how a dermatology expert might explain a diagnosis to a patient.

This study aims to evaluate the performance of Dr. FerManda in diagnosing 42 skin diseases without any additional patient information. Furthermore, we examine two case studies where the model initially made incorrect diagnoses without guidance, followed by the application of guidance to achieve accurate results.

Methods

This study uses publicly available DermNet skin disease image datasets to train the custom ChatGPT model, referred to as the Dr. FerManda Al-based model, with the goal of enabling it to recognize and accurately diagnose 42 skin diseases. The model is built on a comprehensive dataset sourced from DermNet, a globally recognized repository of dermatological images and information. This dataset includes a wide variety of high-quality clinical images, case studies, and professional literature, making it an ideal resource for training an Al model intended for medical applications.



Figure 2. This figure displays images from the dataset representing a wide range of skin diseases used for diagnostic testing by Dr. FerManda's Al model. All images were presented to the model without any additional patient information or clinical guidance. The figure highlights the visual diversity of the skin conditions included in the study.

The training process focuses on equipping Dr. FerManda with a deep understanding of the visual and clinical characteristics associated with each skin disease. The model is trained on images representing 42 different skin conditions, ranging from common ailments such as eczema, psoriasis, and acne vulgaris to more serious diseases like melanoma, squamous cell carcinoma, and pemphigoid vulgaris. Some cases require additional information due to similarities in disease presentation. Specifically, two cases - Darier's disease and pemphigoid vulgaris - were tested both without and with supplementary clinical information. This diversity is essential for improving the model's diagnostic accuracy in real-world scenarios, as skin diseases can manifest differently depending on factors such as ethnicity, environmental influences, and underlying health conditions. Each test case is manually verified to ensure the quality and accuracy of the model's output. The link to the publicly available DermNet skin disease dataset is provided below (https://dermnetnz.org/).

Figure 1 provides an overview of the homepage of the custom ChatGPT Dr. FerManda Al-based model.

Figure 2 displays images from the skin disease datasets used in this study.

Table 1 lists the 42 skin diseases from the dataset that were diagnosed without any additional patient information.

Table 1. All 42 disease images used for diagnostic purposes in the custom ChatGPT Dr. FerManda model are sourced from the DermNet skin disease image repository

No.	Skin Disease	Repository
1	Acne vulgaris	DermNet
2	Vitiligo	DermNet
3	Eczema (Atopic Dermatitis)	DermNet
4	Impetigo	DermNet
5	Urticaria	DermNet
6	Alopecia areata	DermNet
7	Dermatitis herpetiformis	DermNet
8	Tinea pedis	DermNet
9	Tinea cruris	DermNet
10	Tinea capitis	DermNet
11	Tinea unguium (onychomycosis)	DermNet
12	Tinea corporis	DermNet
13	Pityriasis rosea	DermNet
14	Pityriasis versicolor	DermNet
15	Psoriasis	DermNet
16	Acanthosis nigricans	DermNet
17	Angular cheilitis	DermNet
18	Cutaneous horn	DermNet
19	Herpes labialis	DermNet
20	Herpes zoster (shingles)	DermNet
21	Melasma	DermNet
22	Lichen planus	DermNet
23	Hirsutism	DermNet
24	Erythema multiform	DermNet
25	Dyshidrotic eczema (Pompholyx)	DermNet
26	Hidradenitis suppurative (HS)	DermNet
27	Keloid	DermNet
28	Hypertrophic scar	DermNet
29	Acne scars	DermNet
30	Rhytides (wrinkles)	DermNet
31	Infantile hemangioma	DermNet
32	Molluscum contagiosum	DermNet
33	Telangiectasia	DermNet
34	Basal cell carcinoma (BCC)	DermNet
35	Squamous cell carcinoma (SCC)	DermNet
36.	Melanoma	DermNet
37	Rhinophyma	DermNet
38	Cellulitis	DermNet
39	Napkin dermatitis	DermNet
40	Striae distensae (stretch marks)	DermNet
41	Ichthyosis vulgaris	DermNet
42	Skin tags	DermNet

The inclusion and exclusion criteria for diseases were based on the quality of images in the

repository and the similarity of their clinical presentations. The custom ChatGPT Dr. FerManda model successfully diagnosed all 42 diseases without requiring any additional information. However, certain diseases - such as pemphigoid vulgaris, Darier's disease, measles, chickenpox, pityriasis rubra pilaris, pityriasis alba, furuncles, and carbuncles - required extra information due to their resemblance to other conditions, like vitiligo and pityriasis alba.

In this study, we demonstrated the diagnosis of 42 skin diseases without supplementary data and highlighted two cases where the model initially misdiagnosed the condition but achieved accurate results after receiving additional information.

The steps for using this model are outlined below: (1) Open ChatGPT on OpenAI. (2) Click on "Explore GPTs" on the left side of the page. (3) Search for Dr. FerManda. (4) Open the Dr. FerManda model, upload the skin disease image in the input bar, and type: "Please diagnose the disease and break down its symptoms, causes, and management". (5) No additional information - such as symptoms, age, gender, affected area, or family history - was provided to Dr. FerManda to ensure diagnosis based solely on images. (6) For some diseases, additional information (e.g., symptoms, age, gender, affected area, family history) is necessary for accurate diagnosis.

Results

The results of this study highlight the impressive performance of the custom ChatGPT Dr. FerManda Al-based model, which achieved a perfect 100% accuracy rate in diagnosing 42 skin diseases without any additional information. This high level of precision was consistently maintained across all test cases, regardless of the complexity or variability of the conditions. In every instance, Dr. FerManda not only accurately identified the specific skin disease but also provided detailed descriptions of the associated symptoms. Additionally, the model analyzed the underlying causes of each condition and recommended appropriate treatment plans, demonstrating its effectiveness as a comprehensive diagnostic tool. These findings indicate that the model can efficiently handle a wide range of dermatological cases. Images of two of the 42 diagnosed cases without additional information are shown in Figures 3 and 4

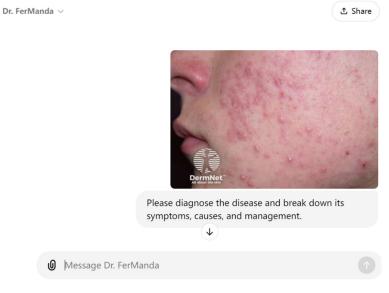


Figure 3. Image of Case No. 1 submitted to Dr. FerManda's Al model for diagnosis, symptom analysis, and management recommendations. The model accurately diagnosed the condition without any attending tips such as age, gender, or clinical history. This figure demonstrates the model's diagnostic capability using image-only input.

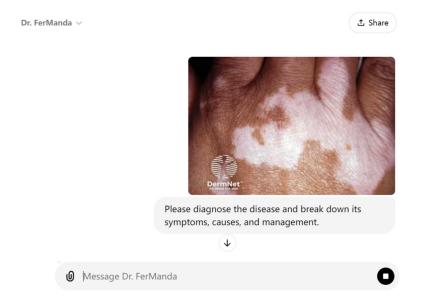


Figure 4. Image of Case No. 2 submitted to Dr. FerManda's Al model for diagnostic analysis. The model delivered a comprehensive diagnosis, including symptoms, causes, and management recommendations, without relying on attending tips. This figure highlights the model's efficiency in working with image-only inputs.

(Case Nos. 1 and 2), followed by **Figures 5A-F**, which presents a comprehensive diagnosis of the disease, including symptoms, causes, pathophysiology, management, and prognosis. **Figures 6** and **7** show Case Nos. 3 and 4.

Cases 1 and 2 were diagnosed by the custom ChatGPT Dr. FerManda Al-based model without any additional information, using the prompt: "Please diagnose the disease and break down its symptoms, causes, and management". However, some diseases - such as pemphigoid vulgaris, Darier's disease, measles, chickenpox, pityriasis rubra pilaris, pityriasis alba, furuncles, and carbuncles require additional information to ensure accurate diagnosis due to similarities in their presentations (e.g., vitiligo and pityriasis alba). Two initially misdiagnosed diseases, pemphigoid vulgaris and Darier's disease, were reevaluated with supplementary clinical details - including symptoms, age, gender, affected area, and family history - leading to more accurate diagnoses. Cases 3 and 4 served as examples where additional information was provided to achieve precise diagnoses. The results for Case 3 are shown in Figures 6 and 7, while Figures 8 and 9 present the results for Case 4. Specifically, Figure 9 displays the results for Case 4, where supplementary information helped the model reach an accurate diagnosis.

The custom ChatGPT Dr. Fer-Manda model demonstrated state-of-the-art performance, achieving 100% accuracy in diagnosing 42 skin diseases without requiring additional information. Although some diseases required further clinical details due to overlapping

features, the model remained highly effective. It makes accurate diagnoses in under 10 seconds and provides comprehensive information on disease symptoms, causes, and management plans. The model is time-efficient, cost-

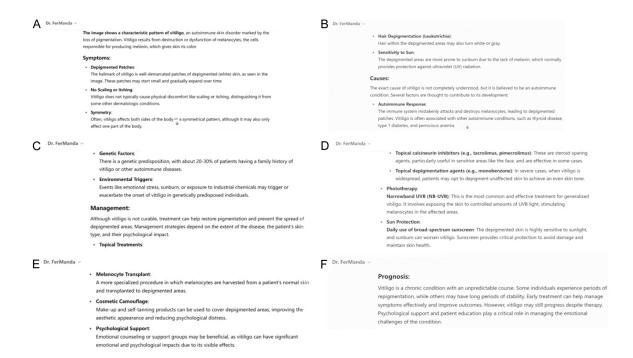


Figure 5. Detailed diagnostic results for Case No. 1 and Case No. 2, demonstrating the model's ability to provide comprehensive information on the disease, including symptoms, causes, treatment management, and prognosis. The outputs are presented across (A-F).



Figure 6. Case No. 3 was submitted to Dr. FerManda's model without attending tips, and the model incorrectly diagnosed the disease as intertrigo.

free, and user-friendly. Dermatologists and skin specialists can rely on this tool to diagnose a wide range of skin diseases and guide precise treatment. Artificial intelligence acts as a bridge between skin specialists and patients, facilitating accurate diagnoses and timely interventions. Furthermore, this model can serve as a

public education tool, enabling individuals in remote areas to perform initial self-assessments and then consult skin specialists via telemedicine platforms to confirm diagnoses and develop personalized treatment plans based on the disease's severity and location.

Discussion

This study demonstrated that the custom ChatGPT model, Dr. FerManda, achieved an exceptional accuracy rate of 100% in diagnosing 42 skin diseases based solely on image inputs. This impressive performance highlights the

model's significant potential for integrating Al into dermatological practice. A key strength of Dr. FerManda is its ability to diagnose conditions without requiring detailed patient information, emphasizing its innovative capability to identify diseases based purely on visual data. This feature is especially valuable in remote or



Figure 7. Output for Case No. 3 after adding attending tips. The model correctly identifies Darier's disease and provides detailed information on its symptoms, causes, and management. This demonstrates how attending tips can enhance the model's diagnostic accuracy.



Figure 8. Case No. 3 was submitted to Dr. FerManda's model without attending tips, and the model incorrectly diagnosed the disease as herpangina.

underserved areas, where access to experienced dermatologists is limited, hindering early detection of skin diseases.

However, the misdiagnoses of conditions such as pemphigoid vulgaris and Darier's disease - when relying solely on images - illustrate the limitations of image-based diagnosis. These diseases share visual similarities with other conditions, necessitating a more nuanced anal-

ysis that incorporates clinical context. The initial errors were corrected by adding clinical details such as symptoms, age, gender, affected area, and family history, indicating that for certain diseases, combining image recognition with patient-specific information is essential for accurate diagnosis.

Dr. FerManda's rapid diagnosis time of under 10 seconds represents a significant improvement over traditional methods, offering efficiency gains in clinical settings and enhancing patient self-assessment scenarios. This speed, combined with high diagnostic accuracy, makes the model a valuable tool for busy dermatology clinics and telemedicine platforms, where reducing patient wait times and improving diagnostic throughput are critical. By streamlining workflows and minimizing human error, Dr. FerManda can serve as an effective adjunct to dermatologists, enhancing initial diagnostic suggestions and improving patient care.

Overall, these findings suggest a promising future for AI in dermatology, enabling innovative approaches to skin disease diagnosis and management. However, important limitations and considerations remain. Dr. FerManda

cannot replace dermatologists because it lacks comprehensive clinical judgment. While it draws from image data and pre-existing knowledge, it cannot evaluate a patient's full medical history, symptoms, or conduct a physical examination. The model may also face challenges with rare or complex cases that require expert clinical insight. Additionally, Dr. FerManda cannot interact with patients or address nuances in their concerns, nor can it make broader clini-

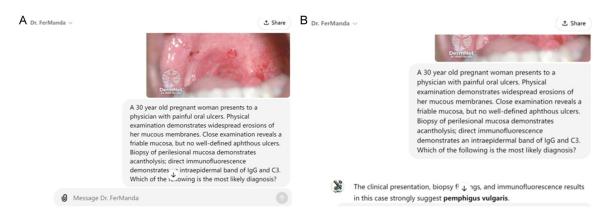


Figure 9. Case No. 4 inputted into Dr. FerManda's model with attending tips, including detailed patient information such as symptoms and family history. The model initially misdiagnosed the case as herpangina but, after the attending tips were provided, it correctly identified the condition as pemphigoid vulgaris. This demonstrates the model's improved accuracy with added clinical context.

cal decisions such as treatment planning. Although it is a valuable diagnostic support tool, it cannot replicate the experience, ethical judgment, and patient-centered care provided by dermatologists.

Furthermore, expanding the model's training with a larger and more diverse dataset - including both common and rare skin diseases - is necessary to improve accuracy. We believe that incorporating high-resolution images along with contextual information such as symptoms, age, gender, disease location, and family history could further enhance diagnostic precision. Nevertheless, it is essential to emphasize that management of skin diseases should always be overseen by qualified dermatologists. Patients should avoid self-prescribing medications, as improper treatment can lead to complications or worsen their condition.

Conclusion

In conclusion, the custom ChatGPT model, Dr. FerManda, demonstrated exceptional performance in diagnosing skin diseases, achieving an impressive 100% accuracy rate across a diverse set of 42 conditions. This study highlights the model's ability to accurately identify diseases using image-based inputs alone, positioning it as a revolutionary tool in dermatology - particularly in remote or underserved areas where access to experienced dermatologists is limited. While the model excels in diagnosing many conditions without additional informa-

tion, certain diseases with similar presentations may require contextual data to ensure precise diagnosis. Dr. FerManda's capability to provide comprehensive insights into symptoms, causes, and management plans further enhances its value as a diagnostic aid.

Overall, this research underscores the promising integration of artificial intelligence in dermatology, paving the way for improved patient outcomes through earlier detection and treatment of skin diseases. However, it is important to recognize that this model cannot replace dermatologists, as it lacks clinical judgment and the ability to assess a patient's full medical history, symptoms, or conduct physical examinations. It may face challenges with complex or rare cases and cannot make treatment decisions or interact with patients. While a valuable diagnostic support tool, it cannot replicate the expertise, ethical judgment, or patient-centered care provided by dermatologists. Future studies should focus on expanding the dataset and refining the model to enhance its diagnostic capabilities across an even broader spectrum of dermatological conditions.

Disclosure of conflict of interest

None.

Address correspondence to: Haiyan Cui, Department of Plastic and Cosmetic Surgery, Tongji Hospital, School of Medicine, Tongji University, Shanghai 200092, China. E-mail: u2beauty1@sina.com

References

- [1] Manoharan P and Surapaneni KM. Assessing the diagnostic capability of ChatGPT through clinical case scenarios in dermatology. Indian J Dermatol Venereol Leprol 2025; 91: S27-S28.
- [2] Sallam M. ChatGPT utility in healthcare education, research, and practice: systematic review on the promising perspectives and valid concerns. Healthcare (Basel) 2023; 11: 887.
- [3] Ayub I, Hamann D, Hamann CR and Davis MJ. Exploring the potential and limitations of chat generative pre-trained transformer (ChatGPT) in generating board-style dermatology questions: a qualitative analysis. Cureus 2023; 15: e43717.
- [4] Munoz-Zuluaga C, Zhao Z, Wang F, Greenblatt MB and Yang HS. Assessing the accuracy and clinical utility of ChatGPT in laboratory medicine. Clin Chem 2023; 69: 939-940.
- [5] Flohr C and Hay R. Putting the burden of skin diseases on the global map. Br J Dermatol 2021; 184: 189-190.
- [6] Tizek L, Schielein MC, Seifert F, Biedermann T, Böhner A and Zink A. Skin diseases are more common than we think: screening results of an unreferred population at the Munich Oktoberfest. J Eur Acad Dermatol Venereol 2019; 33: 1421-1428.

- [7] Hajgude J, Bhavsar A, Achara H and Khubchandani N. Skin disease detection using image processing with data mining and deep learning. Int Res J Eng Technol 2019; 6: 4363-4366
- [8] AL-Khikani FHO. Dermatophytosis a worldwide contiguous fungal infection: growing challenge and few solutions. Biomed Biotechnol Res J 2020; 4: 117-122.
- [9] Yasir R, Rahman MA and Ahmed N. Dermatological disease detection using image processing and artificial neural network. In: Proceedings of the 8th International Conference on Electrical and Computer Engineering; 2014. pp. 687-690.
- [10] Kushimo O, Salau A, Adeleke O and Olaoye D. Deep learning model to improve melanoma detection in people of color. Arab J Basic Appl Sci 2023; 30: 92-102.
- [11] Bajwa MN, Muta K, Malik MI, Siddiqui SA, Braun SA, Homey B, Dengel A and Ahmed S. Computer-aided diagnosis of skin diseases using deep neural networks. Appl Sci 2020; 10: 2488.
- [12] Ali R and Cui H. Leveraging ChatGPT for enhanced aesthetic evaluations in minimally invasive facial procedures. Aesthetic Plast Surg 2025; 49: 950-961.