Pre-trained Multimodal Large Language Model Enhances Dermatological Diagnosis using SkinGPT-4

Juexiao Zhou[1,2,#], Xiaonan He[3,#,*], Liyuan Sun[4,#], Jiannan Xu[4], Xiuying Chen[1,2], Yuetan Chu[1,2], Longxi Zhou[1,2], Xingyu Liao[1,2], Bin Zhang[1,2], Xin Gao[1,2,*]

Abstract—Large language models (LLMs) are seen to have tremendous potential in advancing medical diagnosis recently. However, it is important to note that most current LLMs are limited to text interaction alone. Meanwhile, the development of multimodal large language models for medical diagnosis is still in its early stages, particularly considering the prevalence of image-based data in the field of medical diagnosis, among which dermatological diagnosis is a very important task as skin and subcutaneous diseases rank high among the leading contributors to the global burden of nonfatal diseases. Inspired by current state-of-the-art multimodal large language models, we present SkinGPT-4, which is the world's first interactive dermatology diagnostic system based on multimodal large language models. To implement SkinGPT-4, we have designed a new framework that aligned a pre-trained vision transformer with a large language model named Falcon-40B-Instruct, which is based on Falcon. To train SkinGPT-4, we have collected an extensive collection of skin disease images (comprising 52,929 publicly available and proprietary images) along with clinical concepts and doctors' notes and designed a two-step training strategy. To demonstrate the robustness of SkinGPT-4, we have conducted quantitative evaluations on 150 real-life cases, which were independently reviewed by certified dermatologists. With SkinGPT-4, users could upload their own skin photos for diagnosis, and the system could autonomously evaluate the images, identifies the characteristics and categories of the skin conditions, performs in-depth analysis, and provides interactive treatment recommendations. Meanwhile, SkinGPT-4's local deployment capability and commitment to user privacy also render it an appealing choice for patients. Though SkinGPT-4 is not a substitute for doctors, it could enhance users' comprehension of their medical conditions, facilitate improve communication between patients and doctors, expedite the diagnostic process for dermatologists, facilitate triage, and potentially promote human-centred care and healthcare equity in underdeveloped areas. In summary, SkinGPT-4 represents a significant leap forward in the field of dermatology diagnosis in the era of large language models and a valuable exploration of multimodal large language models in medical diagnosis.

Index Terms—Dermatology, Deep learning, Large language model

1 Introduction

Skin and subcutaneous diseases rank as the fourth major cause of nonfatal disease burden worldwide, affecting a considerable proportion of individuals, with a prevalence ranging from 30% to 70% across all ages and regions [1]. However, dermatologists are consistently in short supply, particularly in rural areas, and consultation costs are on the rise [2], [3], [4]. As a result, the responsibility of diagnosis often falls on non-specialists such as primary care physicians, nurse practitioners, and physician assistants, which may have limited knowledge and training [5] and low accuracy on diagnosis [6], [7]. The use of store-and-forward teledermatology has become dramatically popular in order to expand the range of services available to medical professionals [8], which involves transmitting digital images of the affected skin area (usually taken using a digital camera or smartphone) [9] and other relevant medical information from users to dermatologists. Then, the dermatologist reviews the case remotely and advises on diagnosis, workup, treatment, and follow-up recommendations [10], [11]. Nonetheless, the field of dermatology diagnosis faces three significant hurdles [12]. Firstly, there is a shortage of dermatologists accessible to diagnose patients, particularly in rural regions. Secondly, accurately interpreting skin disease images poses a considerable challenge. Lastly,

[1] *Computer Science Program, Computer, Electrical and Mathematical Sciences and Engineering Division, King Abdullah University of Science and Technology (KAUST), Thuwal 23955-6900, Kingdom of Saudi Arabia*
[2] *Computational Bioscience Research Center, King Abdullah University of Science and Technology (KAUST), Thuwal 23955-6900, Kingdom of Saudi Arabia*
[3] *Emergency Critical Care Center, Beijing AnZhen Hospital, Affiliated to Capital Medical University, Beijing 100029, China*
[4] *Department of Dermatology, Beijing AnZhen Hospital, Affiliated to Capital Medical University, Beijing 100029, China*
[#] *These authors contributed equally.*
[*] *Corresponding author. e-mail: xin.gao@kaust.edu.sa*

generating patient-friendly diagnostic reports is usually a time-consuming and labor-intensive task for dermatologists [4], [13].

Advancements in technology have led to the development of various tools and techniques to aid dermatologists in their diagnosis [13], [14], [15]. For example, the development of artificial intelligence tools to aid in the diagnosis of skin disorders from images has been made possible by recent advancements in deep learning [16], [17], such as skin cancer classification [18], [19], [20], [21], [22], [23], [24], [25], [26], [27], dermatopathology [28], [29], [30], predicting novel risk factors or epidemiology [31], [32], identifying onychomycosis [33], quantifying alopecia areata [34], classify skin lesions from mpox virus infection [35], and so on [4]. Among these, most studies have predominantly concentrated on identifying skin lesions through dermoscopic images [36], [37], [38]. However, dermatoscopy is often not readily available outside of dermatology clinics. Some studies have explored the use of clinical photographs of skin cancer [18], onychomycosis [33], and skin lesions on educational websites [39]. Nevertheless, those methods are tailored for particular diagnostic objectives as classification tasks and their approach still requires further analysis by dermatologists to issue reports and make clinical decisions. Those methods are unable to automatically generate detailed reports in natural language and allow interactive dialogues with patients. At present, there are no such diagnostic systems available for users to self-diagnose skin conditions by submitting images that can automatically and interactively analyze and generate easy-to-understand text reports.

Over the past few months, the field of large language models (LLMs) has seen significant advancements [40], [41], offering remarkable language comprehension abilities and the potential to perform complex linguistic tasks. One of the most anticipated models is GPT-4 [42], which is a large-scale multimodal model that has demonstrated exceptional capabilities, such as generating accurate and detailed image descriptions, providing explanations for atypical visual occurrences, constructing websites based on handwritten textual descriptions, and even acting as family doctors [43]. Despite these remarkable advancements, some features of GPT-4 are still not accessible to the public and are closed-source. Users need to pay and use some features through API. As an accessible alternative, ChatGPT, which is also developed by OpenAI, has demonstrated the potential to assist in disease diagnosis through conversation with patients [44], [45], [46], [46], [47], [48], [49]. By leveraging its advanced natural language processing capabilities, ChatGPT could interpret symptoms and medical history provided by patients and make suggestions for potential diagnoses or referrals to appropriate dermatological specialists [50]. However, it is important to note that most LLMs are limited to text interaction alone currently. Nevertheless, the development of multimodal large language models for medical diagnosis is still in its early stages [51], particularly considering the prevalence of image-based data in the field of medical diagnosis, among which, dermatological diagnosis is a very important task but lacks relevant research on enhanced diagnosis with multimodal large language models.

The idea of providing skin images directly for automatic dermatological diagnosis and generating text reports is exciting because it could greatly help solve the three aforementioned challenges in the field of dermatology diagnosis. However, there exists no method to accomplish this at present. But in related areas, ChatCAD [52] is one of the most advanced approaches that designed various networks to take X-rays, CT scans, and MRIs images to generate diverse outputs, which are then transformed into text descriptions. These descriptions are combined as inputs to ChatGPT to generate a condensed report and offer interactive explanations and medical recommendations based on the given image. However, their proposed vision-text models were limited to certain tasks. Meanwhile, for ChatCAD, users need to use ChatGPT's API to upload text descriptions, which could raise data privacy issues [41], [53], [54] as both medical images and text descriptions contain a lot of patients' private information [55], [56], [57], [58]. To address those issues, MiniGPT-4 [59] is the first open-source method that allows users to deploy locally to interface images with state-of-the-art LLMs and interact using natural language without the need to fine-tune both pre-trained large models but only a small alignment layer. MiniGPT-4 aims to combine the power of a large language model with visual information obtained from a pre-trained vision encoder. To achieve this, the model uses Vicuna [60] as its language decoder, which is built on top of LLaMA [61] and is capable of performing complex linguistic tasks. To process visual information, the same visual encoder used in BLIP-2 [62] is employed, which consists of a ViT [63] backbone combined with a pre-trained Q-Former. Both the language and vision models are open-source. To bridge the gap between the visual encoder and the language model, MiniGPT-4 utilizes a linear projection layer. However, MiniGPT-4 is trained on the combined dataset of Conceptual Caption [64], SBU [65], and LAION [66], which are irrelevant to medical images, especially dermatological images. Therefore, it is still challenging to directly apply MiniGPT-4 to specific domains such as formal dermatology diagnosis. Meanwhile, due to the limitations of Vicuna, MiniGPT-4 could not support commercial use, which could also be further improved by incorporating other state-of-the-art large language models.

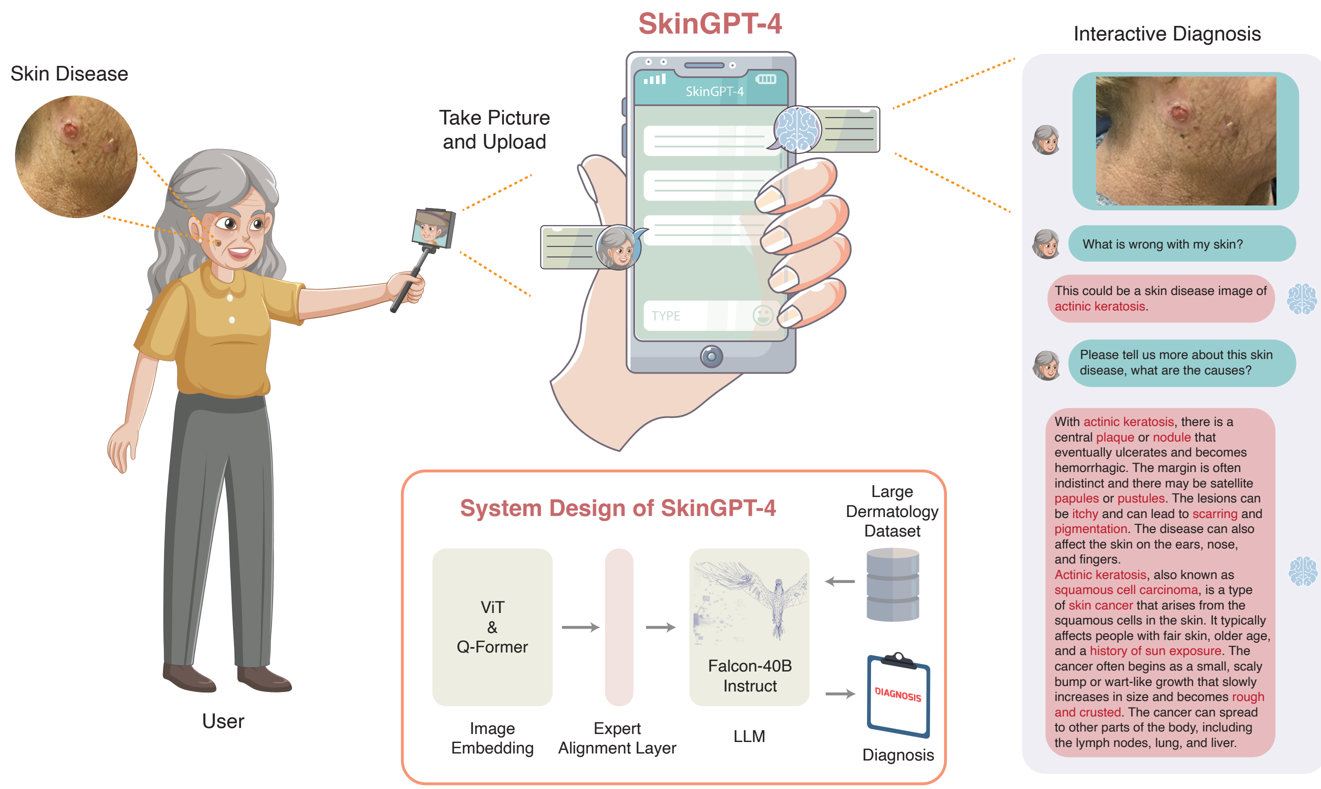

Fig. 1. Illustration of SkinGPT-4. SkinGPT-4 is the world's first interactive dermatology diagnostic system based on multimodal large language models. To implement SkinGPT-4, we have designed a new framework that aligned a pre-trained vision transformer with a large language model named Falcon-40B-Instruct. SkinGPT-4 was trained on a vast collection (52,929) of both public and in-house skin disease images, accompanied by clinical concepts and doctors' notes. With SkinGPT-4, users could upload their own skin photos for diagnosis, and SkinGPT-4 could autonomously determine the characteristics and categories of skin conditions, perform analysis, provide treatment recommendations, and allow interactive diagnosis. On the right is an example of interactive diagnosis.

Inspired by current state-of-the-art multimodal large language models, we present SkinGPT-4, which is the world's first interactive dermatology diagnostic system based on multimodal large language models. (**Figure** 1). SkinGPT-4 brings innovation on two fronts. Firstly, SkinGPT-4 is the first multimodal large language model aligned with the Falcon-40B-Instruct. Secondly, SkinGPT-4 is the first multimodal large language model applied to dermatologic diagnosis. To implement SkinGPT-4, we have designed a new framework that aligned a pre-trained vision transformer with a pre-trained large language model named Falcon-40B-Instruct, which is based on Falcon [67]. To train SkinGPT-4, we have collected an extensive collection of skin disease images (comprising 52,929 publicly available and proprietary images) along with clinical concepts and doctors' notes. We designed a two-step training process to develop SkinGPT-4 as shown in **Figure** 2. In the initial step, SkinGPT-4 aligns visual and textual clinical concepts, enabling it to recognize medical features within skin disease images and express those medical features with natural language. In the subsequent step, SkinGPT-4 learns to accurately diagnoses the specific types of skin diseases. This comprehensive training methodology ensures the system's proficiency in analyzing and classifying various skin conditions. With SkinGPT-4, users have the ability to upload their own skin photos for diagnosis. The system autonomously evaluates the images, identifies the characteristics and categories of the skin conditions, performs in-depth analysis, and provides interactive treatment recommendations (**Figure** 3). Meanwhile, SkinGPT-4's local deployment capability and commitment to user privacy also render it an appealing choice for patients in search of a dependable and precise diagnosis of their skin ailments. To demonstrate the robustness of SkinGPT-4, we conducted quantitative evaluations on 150 real-life cases, which were independently reviewed by certified dermatologists (**Figure** 4 and **Supplementary information**). The results showed that SkinGPT-4 consistently provided accurate diagnoses of skin diseases. Though SkinGPT-4 is not a substitute for doctors, it greatly enhances users' understanding of their medical conditions, facilitates improved communication between patients and doctors, expedites the diagnostic process for dermatologists, facilitates

Fig. 2. Illustration of our datasets for two-step training of SkinGPT-4. The notes below each image indicate clinical concepts and types of skin diseases. In addition, we have detailed descriptions from the certified dermatologists for images in the step 2 dataset. To avoid causing discomfort, we used a translucent grey box to obscure the displayed skin disease images.

triage, and has the potential to advance human-centred care and healthcare equity, particularly in underdeveloped regions [68]. In summary, SkinGPT-4 represents a significant leap forward in the field of dermatology diagnosis in the era of large language models and a valuable exploration of multimodal large language models in medical diagnosis.

2 Results

2.1 The Overall Design of SkinGPT-4

SkinGPT-4 is an interactive system designed to provide a natural language-based diagnosis of skin disease images as shown in Figure 1. The process commences when the user uploads a skin image, which undergoes encoding by the Vision Transformer (ViT) and Q-Former models to comprehend its contents. The ViT model partitions the image into smaller patches and extracts vital features like edges, textures, and shapes. After that, the Q-Former model generates an embedding of the image based on the features identified by the ViT model, which is done by using a transformer-based architecture that allows the model to consider the context of the image. The alignment layer facilitates the synchronization of visual information and natural language, and the large language model named Falcon-40B-Instruct generates the text-based diagnosis. SkinGPT-4 was trained using large skin disease images along with clinical concepts and doctors' notes to allow for interactive dermatological diagnosis. The system could provide an interactive and user-friendly way to help users self-diagnose skin diseases.

2.2 Interactive, Informative and Understandable Dermatology Diagnosis of SkinGPT-4

SkinGPT-4 brings forth a multitude of advantages for both patients and dermatologists. One notable benefit lies in its utilization of comprehensive and trustworthy

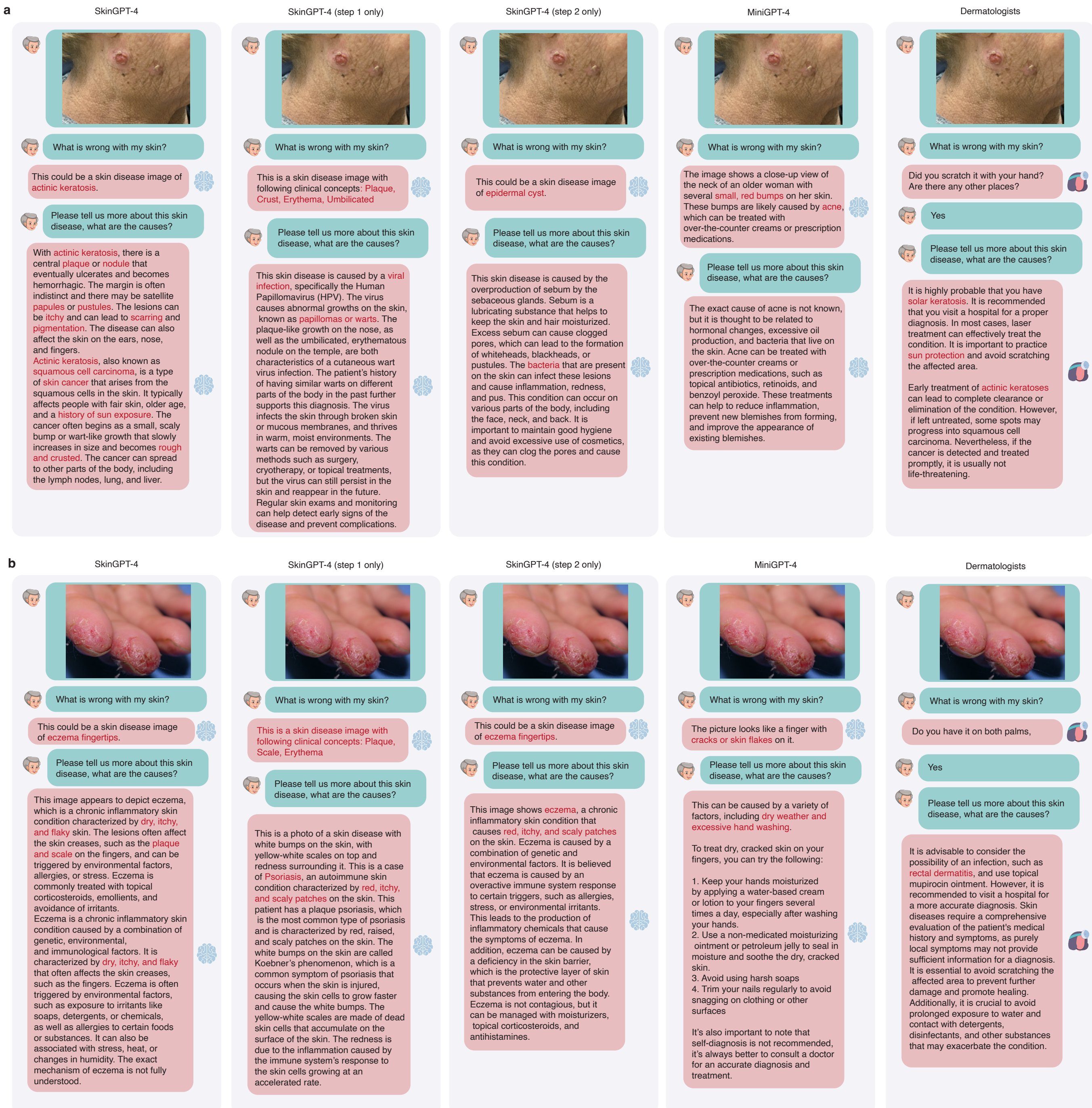

Fig. 3. Diagnosis generated by SkinGPT-4, SkinGPT-4 (step 1 only), SkinGPT-4 (step 2 only), MiniGPT-4 and Dermatologists. **a**. A case of actinic keratosis. **b**. A case of eczema fingertips.

medical knowledge specifically tailored to skin diseases. This empowers SkinGPT-4 to deliver interactive diagnoses, explanations, and recommendations for skin diseases (**Supplementary Video**), which presents a challenge for MiniGPT-4. Unlike MiniGPT-4, which lacks training with pertinent medical knowledge and domain-specific adaptation, SkinGPT-4 overcomes this limitation, enhancing its proficiency in the dermatological domain. To demonstrate the advantage of SkinGPT-4 over MiniGPT-4, we presented two real-life examples of interactive diagnosis as shown in Figure 3. In Figure 3a, an image is presented of an elderly with actinic keratosis on her face. In Figure 3b, an image is provided of a patient with eczema fingertips.

For the actinic keratosis case (Figure 3a), MiniGPT-4 identified features like small and red bumps, and incorrectly diagnosed the skin disease as acne, while SkinGPT-4 identified features like plaque, nodules, pustules, and scarring, and diagnosed the skin disease as actinic keratosis, which is

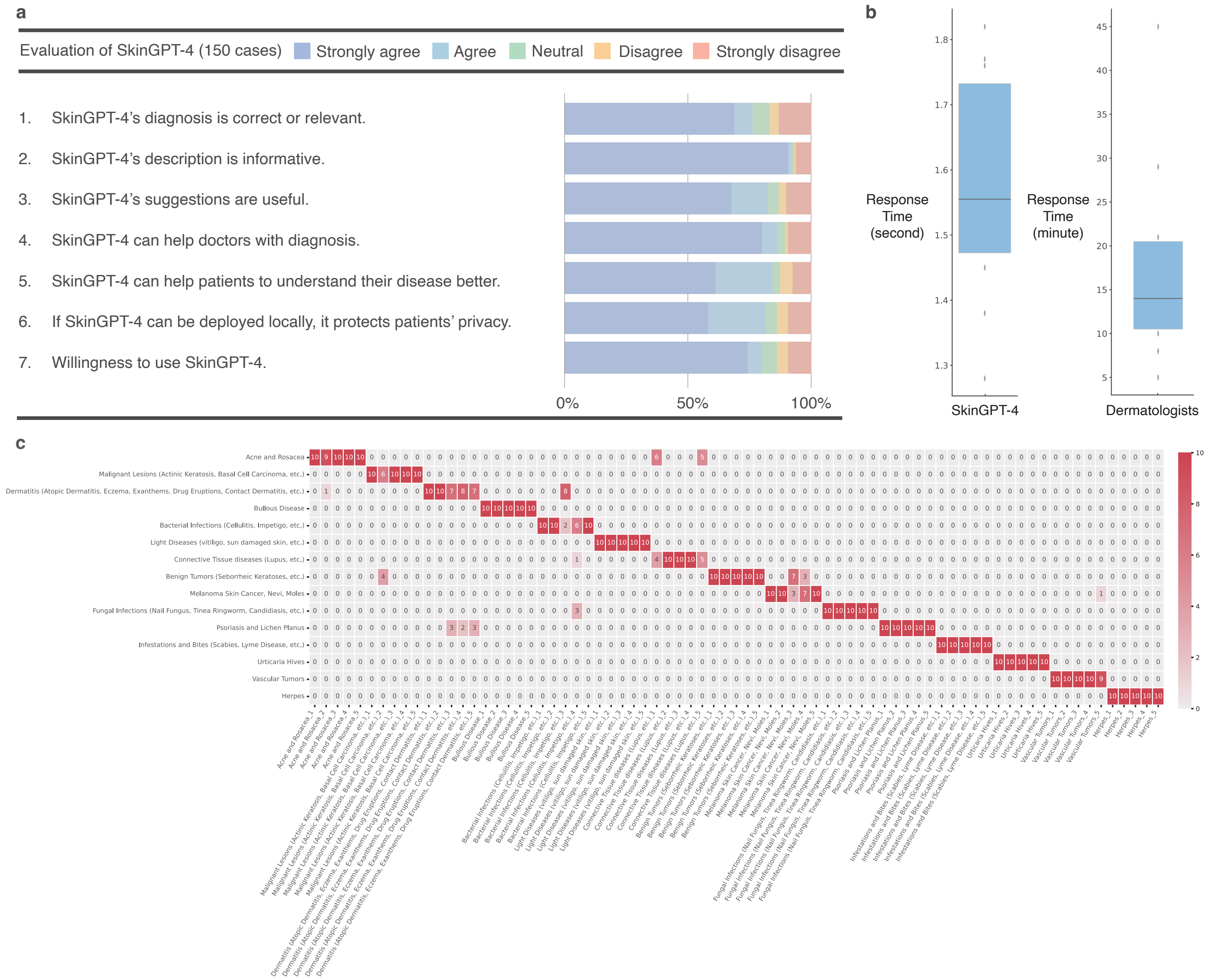

Fig. 4. Clinical evaluation of SkinGPT-4 by certified offline and online dermatologists. **a**. Questionnaire-based assessment of SkinGPT-4 by offline dermatologists. **b**. Response time of SkinGPT-4 compared to consulting dermatologists online. **c**. Consistency test of SkinGPT-4's responses. The x-axis indicates test samples, and the y-axis indicates the diagnostic results.

a common skin condition caused by prolonged exposure to the sun's ultraviolet (UV) rays [69]. During the interactive dialogue, SkinGPT-4 also suggested the cause of the skin disease to be sun exposure, which was also verified as correct by the certified dermatologist. For the example of eczema fingertips case (Figure 3b), MiniGPT-4 identified some features like cracks and skin flakes, missed the type of the skin disease, and diagnosed the cause of the skin disease to be dry weather and excessive hand washing. In comparison, SkinGPT-4 identified either the features of the skin disease as dry itchy and flaky skin, and diagnosed the type of the skin disease to be eczema fingertips, which was also verified by certified dermatologists.

In summary, the absence of dermatological knowledge and domain-specific adaptation poses a significant challenge for MiniGPT-4 in achieving accurate dermatological diagnoses. Contrastingly, SkinGPT-4 successfully and accurately identified the characteristics of the skin diseases displayed in the images. It not only suggested potential disease types but also provided recommendations for potential treatments. This further highlights that domain-specific adaption is crucial for SkinGPT-4 to work for the dermatological diagnosis.

2.3 SkinGPT-4 Masters Medical Features to Improve Diagnosis with the Two-step Training

To further illustrate the capability of SkinGPT-4 in enhancing dermatological diagnosis through learning medical features in skin disease images, we conducted ablation studies, as depicted in Figure 3 by training SkinGPT-4 using either solely the step 1 dataset or solely the step 2 dataset. As specified in **Method** and illustrated in Figure 2, we designed a two-step training process for SkinGPT-4. Initially, we utilized the step 1 dataset to familiarize SkinGPT-4 with the medical features present in dermatological images and allow SkinGPT-4 to express medical features in skin disease images with natural language. Subsequently, we employed the step 2 dataset to train SkinGPT-4 to achieve a more precise diagnosis of disease types.

In the instance of actinic keratosis (Figure 3a), which is a hard case, SkinGPT-4 trained solely on the step 1 dataset demonstrated its proficiency in identifying pertinent medical features such as plaque, crust, erythema, and umbilicated. These precise and comprehensive morphological descriptions accurately captured the characteristics of the skin disease depicted in the image. However, when SkinGPT-4 was exclusively trained on the step 1 dataset, it erroneously diagnosed the skin condition as a viral infection, indicating the importance of incorporating the step 2 dataset for more accurate disease identification. In contrast, when trained solely on the step 2 dataset, SkinGPT-4 failed to capture the accurate morphological descriptions of the skin diseases and instead incorrectly diagnosed it as the result of excessive sebum production. It highlights the necessity of incorporating the step 1 dataset to effectively recognize and comprehend the specific medical features essential for precise dermatological diagnoses. In comparison, SkinGPT-4 with our two-step training simultaneously identified the medical features, such as plaque, nodules, pustules and scarring, and diagnosed the skin disease as actinic keratosis. For simple cases such as the eczema fingertips shown in Figure 3b, SkinGPT-4 could also provide more detailed descriptions of the skin disease image, encompass the medical features and accurately identify the type of skin disease. In conclusion, the two-step training process we have implemented allows SkinGPT-4 to effectively comprehend and master medical features in dermatological images, thereby significantly enhancing the accuracy of diagnoses, which is particularly crucial for hard cases where precise identification of medical features is paramount to accurately determining the type of disease.

2.4 Clinical Evaluation of SkinGPT-4 by Certified Dermatologists

To evaluate the reliability and robustness of SkinGPT-4, we conducted a comprehensive study involving a large number of real-life cases (150) and compared its diagnoses with those of certified dermatologists. The results, presented in Table 2 and Supplementary information, demonstrated that SkinGPT-4 consistently provided accurate diagnoses that were in agreement with those of the certified dermatologists as shown in Figure 4, as well as in all cases detailed in the Supplementary information.

Among the 150 cases, a significant percentage of SkinGPT-4's diagnoses (80.00%) were evaluated as correct or relevant by certified dermatologists. This evaluation encompassed both strongly agree (74.38%) and agree (5.62%). Additionally, SkinGPT-4's responses regarding the causes of the disease and potential treatments were considered informative (81.25%) and useful (84.38%) by the doctors. Furthermore, SkinGPT-4 proved to be a valuable tool for doctors in the diagnosis process (86.25%) and for patients in gaining a better understanding of their diseases (82.50%). The capability of SkinGPT-4 to support local deployment, ensuring user privacy, garnered high agreement (91.88%), further enhancing the willingness to utilize SkinGPT-4 (76.25%).

Overall, the study demonstrated that SkinGPT-4 delivers reliable diagnoses, aids doctors in the diagnostic process, facilitates patient understanding, and prioritizes user privacy, making it a valuable asset in the field of dermatology.

2.5 SkinGPT-4 Acts as a 24/7 On-call Family Doctor

In comparison to online consultations with dermatologists, which often entail waiting minutes for a response, SkinGPT-4 offers several advantages. Firstly, it is available 24/7, ensuring constant access to medical advice. Additionally, SkinGPT-4 provides faster response times, typically within seconds, as depicted in Figure 4b, which makes it a swift and convenient option for patients requiring immediate diagnoses outside of regular office hours.

Moreover, SkinGPT-4's ability to offer preliminary diagnoses empowers patients to make informed decisions about seeking in-person medical attention. This feature can help reduce unnecessary visits to the doctor's office, saving patients both time and money. The potential to improve healthcare access is particularly significant in rural areas or regions experiencing a scarcity of dermatologists. In such areas, patients often face lengthy waiting times or must travel considerable distances to see a dermatologist [70]. By leveraging SkinGPT-4, patients can swiftly and conveniently receive preliminary diagnoses, potentially diminishing the need for in-person visits and alleviating the strain on healthcare systems in these underserved regions.

2.6 Consistency of SkinGPT-4's Diagnosis

GPT tends to generate results in various formats according to probability and thus the risks and consistency associated with AI-generated content must be carefully considered [71], especially in medical diagnosis. To demonstrate the consistency of the results from SkinGPT-4, we randomly selected 45 samples (5 from each class as depicted in Table 2). For each sample, we conducted 10 independent diagnoses. As shown in Figure 4c, the diagnoses made on the same graph were consistent with a consistency ratio of 93.73%. For inconsistent cases, features of multiple possible skin types could be observed by certified dermatologists, such as the benign tumour could be easily confused with melanoma skin cancer. Overall, the diagnoses of SkinGPT-4 are consistent and reliable.

3 METHODS

3.1 Dataset

Our datasets include two public datasets and our private in-house dataset, where the first public dataset was used for the step 1 training, and the second public dataset and our in-house dataset were used for the step 2 training.

The first public dataset named SKINCON [72] is the first medical dataset densely annotated by domain experts to provide annotations useful across multiple disease processes. SKINCON is a skin disease dataset densely annotated by dermatologists and it includes 3230 images from the Fitzpatrick 17k skin disease dataset densely annotated with 48 clinical concepts as shown in **Table** 1, 22 of which have at least 50 images representing the concept, and 656 skin disease images from the Diverse Dermatology Images dataset. The 48 clinical concepts proposed by SKINCON include Vesicle, Papule, Macule, Plaque, Abscess, Pustule, Bulla, Patch, Nodule, Ulcer, Crust, Erosion, Excoriation, Atrophy, Exudate, Purpura/Petechiae, Fissure, Induration, Xerosis, Telangiectasia, Scale, Scar, Friable, Sclerosis, Pedunculated, Exophytic/Fungating, Warty/Papillomatous, Dome-shaped, Flat-topped, Brown (Hyperpigmentation), Translucent, White (Hypopigmentation), Purple, Yellow, Black, Erythema, Comedo, Lichenification, Blue, Umbilicated, Poikiloderma, Salmon, Wheal, Acuminate, Burrow, Gray, Pigmented, and Cyst.

The second public dataset named the Dermnet contains 18,856 images, which are further classified into 15 classes by our board-certified dermatologists, including Acne and Rosacea, Malignant Lesions (Actinic Keratosis, Basal Cell Carcinoma, etc.), Dermatitis (Atopic Dermatitis, Eczema, Exanthems, Drug Eruptions, Contact Dermatitis, etc.), Bullous Disease, Bacterial Infections (Cellulitis, Impetigo, etc.), Light Diseases (vitiligo, sun damaged skin, etc.), Connective Tissue diseases (Lupus, etc.), Benign Tumors (Seborrheic Keratoses, etc.), Melanoma Skin Cancer (Nevi, Moles, etc.), Fungal Infections (Nail Fungus, Tinea Ringworm, Candidiasis, etc.), Psoriasis and Lichen Planus, Infestations and Bites (Scabies, Lyme Disease, etc.), Urticaria Hives, Vascular Tumors, Herpes, and others.

Our private in-house dataset contains 30,187 pairs of skin disease images and corresponding doctors' descriptions. The complete dataset for step 2 training comprises in total of 49,043 pairs of images and textual descriptions as shown in **Table** 2.

TABLE 1
Characteristics of Step 1 Dataset. It is possible for a single image to have multiple medical concepts at the same time. The total number of samples is 3886.

Clinical Concepts	Number of Samples
Erythema	2139
Plaque	1966
Papule	1169
Brown(Hyperpigmentation)	759
Scale	686
Crust	497
White(Hypopigmentation)	257
Yellow	245
Erosion	200
Nodule	189
Ulcer	154
Friable	153
Patch	149
Dome-shaped	146
Exudate	144
Scar	123
Pustule	103
Telangiectasia	100
Black	90
Purple	85
Atrophy	69
Bulla	64
Umbilicated	49
Vesicle	46
Warty/Papillomatous	46
Excoriation	46
Exophytic/Fungating	42
Xerosis	35
Induration	33
Fissure	32
Sclerosis	27
Pedunculated	26
Lichenification	25
Comedo	24
Wheal	21
Flat topped	18
Translucent	16
Macule	13
Salmon	10
Purpura/Petechiae	10
Acuminate	8
Cyst	6
Blue	5
Abscess	5
Poikiloderma	5
Burrow	5
Gray	5
Pigmented	5

TABLE 2
Characteristics of Step 2 Dataset and Clinical Evaluation Dataset.

Major Classes of Skin Disease	Number of Samples in Step 2 Dataset	Number of Samples in Clinical Evaluation Dataset
Acne and Rosacea	840	10
Malignant Lesions (Actinic Keratosis, Basal Cell Carcinoma, etc.)	8166	10
Dermatitis (Atopic Dermatitis, Eczema, Exanthems, Drug Eruptions, Contact Dermatitis, etc.)	5262	10
Bullous Disease	448	10
Bacterial Infections (Cellulitis, Impetigo, etc.)	228	10
Light Diseases (vitiligo, sun damaged skin, etc.)	568	10
Connective Tissue diseases (Lupus, etc.)	420	10
Benign Tumors (Seborrheic Keratoses, etc.)	1916	10
Melanoma Skin Cancer, Nevi, Moles	23373	10
Fungal Infections (Nail Fungus, Tinea Ringworm, Candidiasis, etc.)	2340	10
Psoriasis and Lichen Planus	3460	10
Infestations and Bites (Scabies, Lyme Disease, etc.)	431	10
Urticaria Hives	212	10
Vascular Tumors	735	10
Herpes	405	10
Others	239	/
Total	49043	150

3.2 Details of the Model Structure of SkinGPT-4

SkinGPT-4 is composed of several components, including a frozen image encoder called ViT, a frozen Q-Former, a trainable linear alignment layer, and a frozen large language model known as Falcon-40B-Instruct.

When a patient uploads an image, denoted as $\mathbf{x} \in \mathbb{R}^{H\times W\times C}$, it undergoes a reshaping process to form a sequence of flattened 2D patches, represented as $\mathbf{x_p} \in \mathbb{R}^{N\times(P^2\cdot C)}$. Here, (H, W) denotes the resolution of the original image, C represents the number of channels, (P, P) signifies the resolution of each image patch, and $N = HW/P^2$ represents the total number of patches. In the case of SkinGPT-4, the values of H and W are set to 224, C is 3, and P is 14. These patches are then flattened and projected to a D-dimensional space using a pre-trained linear projection within ViT [73]. Additionally, position embeddings denoted as $\mathbf{E_{pos}}$ are added to the patch embeddings to preserve positional information, following Equation 1. Subsequently, a transformer encoder [74] is applied, which consists of alternating layers of multiheaded self-attention (MSA) and MLP blocks. Layer normalization (LN) is applied before each block, and residual connections are employed after each block, as illustrated in Equations 2 and 3. The pre-trained ViT utilized in SkinGPT-4 possesses the following parameters: an embedding dimension of 1408, a depth of 39, and a number of heads set to 16. These values contribute to the effectiveness and efficiency of the image encoding process.

$$\begin{gathered}\mathbf{z_0} = [\mathbf{x_{class}}; \mathbf{x_p^1 E}; \mathbf{x_p^2 E}; ...; \mathbf{x_p^N E}] + \mathbf{E_{pos}} \\ where\ \mathbf{E} \in \mathbb{R}^{(P^2\cdot C)\times D}, \mathbf{E_{pos}} \in \mathbb{R}^{(N+1)\times D}\end{gathered} \tag{1}$$

$$\mathbf{z'_l} = MSA(LN(\mathbf{z_{l-1}})) + \mathbf{z_{l-1}}, \quad l = 1...L \tag{2}$$

$$\mathbf{z_l} = MLP(LN(\mathbf{z'_l})) + \mathbf{z'_l}, \quad l = 1...L \tag{3}$$

Each output image representation z, generated by ViT, is aligned with the text representation t that is produced by the text transformer and represents the output embedding of the [CLS] token with the pre-trained Q-Former [62]. Subsequently, the last hidden layer of Q-Former is passed through the linear alignment layer, which has an input size equivalent to the hidden size of Q-Former and an output size matching the hidden size of Falcon-40B-Instruct.

A specific prompt format is employed to enable Falcon-40B-Instruct to generate desired text corresponding to the uploaded image. The prompt is structured as follows: "### Instruction: <Img><Image></Img> Could you describe the skin disease in this image for me? ### Response:". The first section of the prompt "### Instruction: <Img>" and the last section of the prompt "</Img> Could you describe the skin disease in this image for me? ### Response:" are tokenized and embedded by Falcon-40B-Instruct. The middle section "<Image>" is replaced with the output obtained from the trainable linear alignment layer. All the embeddings, including the prompt sections, are concatenated and fed into the encoder of Falcon-40B-Instruct to generate the desired text output.

3.3 The two-step Training of SkinGPT-4

SkinGPT-4 was trained using a vast of skin disease images along with clinical concepts and doctors' notes (Figure 1). In the first step, we trained SkinGPT-4 using the step 1 training dataset. This dataset consists of paired skin disease images along with corresponding descriptions of clinical concepts. By training SkinGPT-4 on this dataset, we enabled the model to grasp the nuances of clinical concepts specific to skin diseases.

In the second step, we further refined the model by fine-tuning it using the step 2 dataset, which comprises additional skin images and refined doctors' notes. This iterative training process facilitated the accurate diagnosis of various skin diseases, as SkinGPT-4 incorporated the refined medical insights from the doctors' notes.

By following this two-step fine-tuning approach, SkinGPT-4 attained an enhanced understanding of clinical concepts related to skin diseases and acquired the proficiency to generate accurate diagnoses.

3.4 Hyperparameters and Resources for Model Training

During the training of both steps, the max number of epochs was fixed to 20, the iteration of each epoch was set to 5000, the warmup step was set to 5000, the learning rate was set to 1e-4, and the max text length was set to 160. The entire fine-tuning process required approximately 24 hours to complete and utilized eight NVIDIA A100 (80GB) GPUs. To host the model in a single A100 GPU, we load Falcon-40B-Instrut in 8-bit. During inference, only one NVIDIA A100 (80GB) GPU was necessary. SkinGPT-4 was developed using Python 3.7, PyTorch 1.9.1, and CUDA 11.4. For a comprehensive list of dependencies, please refer to our code availability documentation.

3.5 Clinical Evaluation of SkinGPT-4

To assess the reliability and effectiveness of SkinGPT-4, we assembled a dataset comprising 150 real-life cases of various skin diseases as shown in Table 2. Interactive diagnosis sessions were conducted with SkinGPT-4, utilizing four specific prompts:

1. Could you describe the skin disease in this image for me?

2. Please provide a paragraph listing additional features you observed in the image.

3. Based on the previous information, please provide a detailed explanation of the cause of this skin disease.

4. What treatment and medication should be recommended for this case?

To conduct the clinical evaluation, certified dermatologists were provided with the same set of four questions and were required to make diagnoses based on the given skin disease images. Meanwhile, the dermatologists also evaluated the reports generated by SkinGPT-4 and assigned scores (strongly agree, agree, neutral, disagree, and strongly disagree) to each item in the evaluation form (Figure 4a), including the following questions:

1. SkinGPT-4's diagnosis is correct or relevant.

2. SkinGPT-4's description is informative.

3. SkinGPT-4's suggestions are useful.

4. SkinGPT-4 can help doctors with diagnosis.

5. SkinGPT-4 can help patients to understand their disease better.

6. If SkinGPT-4 can be deployed locally, it protects patients' privacy.

7. Willingness to use SkinGPT-4.

In particular, for questions 3 and 5, we further collected the opinions of users of SkinGPT-4, who usually do not have strong background knowledge in dermatology, to show that SkinGPT-4 is friendly to the general users. Those results allowed for a comprehensive evaluation of SkinGPT-4's performance in relation to certified dermatologists and patients.

4 Conclusion and Discussion

Our study showcases the promising potential of utilizing visual inputs in LLMs to enhance dermatological diagnosis. With the upcoming release of more advanced LLMs like GPT-4, the accuracy and quality of diagnoses could be further improved. However, it is essential to address potential privacy concerns associated with using LLMs like ChatGPT and GPT-4 as an API, as it requires users to upload their private data. In contrast, SkinGPT-4 offers a solution to this privacy issue. By allowing users to deploy the model locally, the concerns regarding data privacy are effectively resolved. Users have the autonomy to use SkinGPT-4 within the confines of their own system, ensuring the security and confidentiality of their personal information.

During the course of a patient's consultation with a dermatologist, the doctor often asks additional questions to gather crucial information that aids in arriving at a precise diagnosis. In contrast, SkinGPT-4 relies on the information provided by users to assist in the diagnostic process. Additionally, doctors often engage in empathetic interactions with patients, as the emotional connection could contribute to the diagnostic process. Due to these factors, it remains challenging for SkinGPT-4 to fully replace dermatologists at present. However, SkinGPT-4 still holds significant value as a tool for both patients and dermatologists. It can greatly expedite the diagnostic process and enhance the overall service delivery. By leveraging its capabilities, SkinGPT-4 empowers patients to obtain preliminary insights into their skin conditions and aids dermatologists in providing more efficient care. While it may not fully substitute for the expertise and empathetic nature of dermatologists, SkinGPT-4 serves as a valuable complementary resource in the field of dermatological diagnosis.

As LLMs-based applications like SkinGPT-4 continue to evolve and improve with the acquisition of even more reliable medical training data, the potential for significant advancements in online medical services is enormous. SkinGPT-4 could play a critical role in improving access to healthcare and enhancing the quality of medical services for patients worldwide. We will continue our research in this field to further develop and refine this technology.

5 Acknowledgements

Special thanks: Thanks to Jun Chen, the author of MiniGPT-4 for the discussion of this work.

Funding: Juexiao Zhou, Xiuying Chen, Yuetan Chu, Longxi Zhou, Xingyu Liao, Bin Zhang, and Xin Gao were supported in part by grants from the Office of Research Administration (ORA) at King Abdullah University of Science and Technology (KAUST) under award number FCC/1/1976-44-01, FCC/1/1976-45-01, REI/1/5202-01-01, REI/1/5234-01-01, REI/1/4940-01-01, RGC/3/4816-01-01, and REI/1/0018-01-01. Xiaonan He was supported by the foundation of the National Natural Science Foundation of China (No. 62272327).

Competing Interests: The authors have declared no competing interests.

Author Contribution Statements: J.Z. and X.G. conceived of the presented idea. J.Z. designed the computational framework and analysed the data. J.Z, L.S., J.X., X.C., Y.C., L.Z., X.L., B.Z. and X.H. conducted the clinical evaluation. X.G. supervised the findings of this work. J.Z., X.H., L.S., J.X. and X.G. took the lead in writing the manuscript and supplementary information. All authors discussed the results and contributed to the final manuscript.

Data availability: The data that support the findings of this study are divided into two groups: shared data and restricted data. Shared data include the SKINCON dataset and the Dermnet dataset. The SKINCON dataset can be accessed at https://skincon-dataset.github.io/. The Dermnet dataset can be accessed at https://www.kaggle.com/datasets/shubhamgoel27/dermnet. The restricted in-house skin disease images used in this study are not publicly available due to restrictions in the data-sharing agreement.

Code availability: To promote academic exchanges, under the framework of data and privacy security, the code proposed by SkinGPT-4 is publicly available at https://github.com/JoshuaChou2018/SkinGPT-4. In the case of non-commercial use, researchers can sign the license provided in the above link and contact J.Z. or X.G. to access the latest non-commercial trained model weights.

References

[1] R. J. Hay, N. E. Johns, H. C. Williams, I. W. Bolliger, R. P. Dellavalle, D. J. Margolis, R. Marks, L. Naldi, M. A. Weinstock, S. K. Wulf *et al.*, "The global burden of skin disease in 2010: an analysis of the prevalence and impact of skin conditions," *Journal of Investigative Dermatology*, vol. 134, no. 6, pp. 1527–1534, 2014.

[2] H. Feng, J. Berk-Krauss, P. W. Feng, and J. A. Stein, "Comparison of dermatologist density between urban and rural counties in the united states," *JAMA dermatology*, vol. 154, no. 11, pp. 1265–1271, 2018.

[3] J. Resneck Jr and A. B. Kimball, "The dermatology workforce shortage," *Journal of the American Academy of Dermatology*, vol. 50, no. 1, pp. 50–54, 2004.

[4] Y. Liu, A. Jain, C. Eng, D. H. Way, K. Lee, P. Bui, K. Kanada, G. de Oliveira Marinho, J. Gallegos, S. Gabriele *et al.*, "A deep learning system for differential diagnosis of skin diseases," *Nature medicine*, vol. 26, no. 6, pp. 900–908, 2020.

[5] D. Seth, K. Cheldize, D. Brown, and E. E. Freeman, "Global burden of skin disease: inequities and innovations," *Current dermatology reports*, vol. 6, pp. 204–210, 2017.

[6] D. G. Federman, J. Concato, and R. S. Kirsner, "Comparison of dermatologic diagnoses by primary care practitioners and dermatologists: a review of the literature," *Archives of family medicine*, vol. 8, no. 2, p. 170, 1999.

[7] G. Moreno, H. Tran, A. L. Chia, A. Lim, and S. Shumack, "Prospective study to assess general practitioners' dermatological diagnostic skills in a referral setting," *Australasian journal of dermatology*, vol. 48, no. 2, pp. 77–82, 2007.

[8] K. M. Yim, A. G. Florek, D. H. Oh, K. McKoy, and A. W. Armstrong, "Teledermatology in the united states: an update in a dynamic era," *Telemedicine and e-Health*, vol. 24, no. 9, pp. 691–697, 2018.

[9] P. R. Kshirsagar, H. Manoharan, S. Shitharth, A. M. Alshareef, N. Albishry, and P. K. Balachandran, "Deep learning approaches for prognosis of automated skin disease," *Life*, vol. 12, no. 3, p. 426, 2022.

[10] F. Martora, A. Ruggiero, G. Fabbrocini, and A. Villani, "Patient satisfaction with remote dermatology consultations during the covid-19 pandemic. comment on 'a qualitative assessment of patient satisfaction with remote dermatology consultations used during the uk's first wave of the covid-19 pandemic in a single, secondary-care dermatology department'," *Clinical and Experimental Dermatology*, vol. 47, no. 11, pp. 2037–2038, 2022.

[11] R. López-Liria, M. Á. Valverde-Martínez, A. López-Villegas, R. J. Bautista-Mesa, F. A. Vega-Ramírez, S. Peiró, and C. Leal-Costa, "Teledermatology versus face-to-face dermatology: An analysis of cost-effectiveness from eight studies from europe and the united states," *International Journal of Environmental Research and Public Health*, vol. 19, no. 5, p. 2534, 2022.

[12] N. Lakdawala, C. Gronbeck, and H. Feng, "Workforce characteristics of nonphysician clinicians in dermatology in the united states," *Journal of the American Academy of Dermatology*, vol. 87, no. 5, pp. 1108–1110, 2022.

[13] I. K. Pious and R. Srinivasan, "A review on early diagnosis of skin cancer detection using deep learning techniques," in *2022 International Conference on Computer, Power and Communications (ICCPC)*. IEEE, 2022, pp. 247–253.

[14] P. Puri, N. Comfere, L. A. Drage, H. Shamim, S. A. Bezalel, M. R. Pittelkow, M. D. Davis, M. Wang, A. R. Mangold, M. M. Tollefson *et al.*, "Deep learning for dermatologists: Part ii. current applications," *Journal of the American Academy of Dermatology*, vol. 87, no. 6, pp. 1352–1360, 2022.

[15] S. Reshma and S. Reeja, "A review of computer assistance in dermatology," in *2023 International Conference on Intelligent and Innovative Technologies in Computing, Electrical and Electronics (IITCEE)*. IEEE, 2023, pp. 66–71.

[16] S. S. Han, I. Park, S. E. Chang, W. Lim, M. S. Kim, G. H. Park, J. B. Chae, C. H. Huh, and J.-I. Na, "Augmented intelligence dermatology: deep neural networks empower medical professionals in diagnosing skin cancer and predicting treatment options for 134 skin disorders," *Journal of Investigative Dermatology*, vol. 140, no. 9, pp. 1753–1761, 2020.

[17] D. Popescu, M. El-Khatib, H. El-Khatib, and L. Ichim, "New trends in melanoma detection using neural networks: a systematic review," *Sensors*, vol. 22, no. 2, p. 496, 2022.

[18] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun, "Dermatologist-level classification of skin cancer with deep neural networks," *nature*, vol. 542, no. 7639, pp. 115–118, 2017.

[19] S. S. Han, M. S. Kim, W. Lim, G. H. Park, I. Park, and S. E. Chang, "Classification of the clinical images for benign and malignant cutaneous tumors using a deep learning algorithm," *Journal of Investigative Dermatology*, vol. 138, no. 7, pp. 1529–1538, 2018.

[20] H. A. Haenssle, C. Fink, R. Schneiderbauer, F. Toberer, T. Buhl, A. Blum, A. Kalloo, A. B. H. Hassen, L. Thomas, A. Enk *et al.*, "Man against machine: diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists," *Annals of oncology*, vol. 29, no. 8, pp. 1836–1842, 2018.

[21] M. A. Marchetti, N. C. Codella, S. W. Dusza, D. A. Gutman, B. Helba, A. Kalloo, N. Mishra, C. Carrera, M. E. Celebi, J. L. DeFazio *et al.*, "Results of the 2016 international skin imaging collaboration international symposium on biomedical imaging challenge: Comparison of the accuracy of computer algorithms to dermatologists for the diagnosis of melanoma from dermoscopic images," *Journal of the American Academy of Dermatology*, vol. 78, no. 2, pp. 270–277, 2018.

[22] T. J. Brinker, A. Hekler, A. Hauschild, C. Berking, B. Schilling, A. H. Enk, S. Haferkamp, A. Karoglan, C. von Kalle, M. Weichenthal *et al.*, "Comparing artificial intelligence algorithms to 157 german dermatologists: the melanoma classification benchmark," *European Journal of Cancer*, vol. 111, pp. 30–37, 2019.

[23] J. Yap, W. Yolland, and P. Tschandl, "Multimodal skin lesion classification using deep learning," *Experimental dermatology*, vol. 27, no. 11, pp. 1261–1267, 2018.

[24] s. L. P. Aggarwal, "Data augmentation in dermatology image recognition using machine learning," *Skin Research and Technology*, vol. 25, no. 6, pp. 815–820, 2019.

[25] P. Tschandl, C. Rosendahl, B. N. Akay, G. Argenziano, A. Blum, R. P. Braun, H. Cabo, J.-Y. Gourhant, J. Kreusch, A. Lallas *et al.*, "Expert-level diagnosis of nonpigmented skin cancer by combined convolutional neural networks," *JAMA dermatology*, vol. 155, no. 1, pp. 58–65, 2019.

[26] S. S. Han, I. J. Moon, W. Lim, I. S. Suh, S. Y. Lee, J.-I. Na, S. H. Kim, and S. E. Chang, "Keratinocytic skin cancer detection on the face using region-based convolutional neural network," *JAMA dermatology*, vol. 156, no. 1, pp. 29–37, 2020.

[27] O. Jones, R. Matin, M. van der Schaar, K. P. Bhayankaram, C. Ranmuthu, M. Islam, D. Behiyat, R. Boscott, N. Calanzani, J. Emery *et al.*, "Artificial intelligence and machine learning algorithms for early detection of skin cancer in community and primary care settings: a systematic review," *The Lancet Digital Health*, vol. 4, no. 6, pp. e466–e476, 2022.

[28] A. Hekler, J. S. Utikal, A. H. Enk, C. Berking, J. Klode, D. Schadendorf, P. Jansen, C. Franklin, T. Holland-Letz, D. Krahl *et al.*, "Pathologist-level classification of histopathological melanoma images with deep neural networks," *European Journal of Cancer*, vol. 115, pp. 79–83, 2019.

[29] Y. Jiang, J. Xiong, H. Li, X. Yang, W. Yu, M. Gao, X. Zhao, Y. Ma, W. Zhang, Y. Guan *et al.*, "Recognizing basal cell carcinoma on smartphone-captured digital histopathology images with a deep neural network," *British Journal of Dermatology*, vol. 182, no. 3, pp. 754–762, 2020.

[30] A. Hekler, J. S. Utikal, A. H. Enk, W. Solass, M. Schmitt, J. Klode, D. Schadendorf, W. Sondermann, C. Franklin, F. Bestvater *et al.*, "Deep learning outperformed 11 pathologists in the classification of histopathological melanoma images," *European Journal of Cancer*, vol. 118, pp. 91–96, 2019.

[31] D. Roffman, G. Hart, M. Girardi, C. J. Ko, and J. Deng, "Predicting non-melanoma skin cancer via a multi-parameterized artificial neural network," *Scientific reports*, vol. 8, no. 1, p. 1701, 2018.

[32] J. P. Lott, D. M. Boudreau, R. L. Barnhill, M. A. Weinstock, E. Knopp, M. W. Piepkorn, D. E. Elder, S. R. Knezevich, A. Baer, A. N. Tosteson *et al.*, "Population-based analysis of histologically confirmed melanocytic proliferations using natural language processing," *JAMA dermatology*, vol. 154, no. 1, pp. 24–29, 2018.

[33] S. S. Han, G. H. Park, W. Lim, M. S. Kim, J. I. Na, I. Park, and S. E. Chang, "Deep neural networks show an equivalent and often superior performance to dermatologists in onychomycosis diagnosis: Automatic construction of onychomycosis datasets by region-based convolutional deep neural network," *PloS one*, vol. 13, no. 1, p. e0191493, 2018.

[34] E. Bernardis and L. Castelo-Soccio, "Quantifying alopecia areata via texture analysis to automate the salt score computation," in *Journal of Investigative Dermatology Symposium Proceedings*, vol. 19, no. 1. Elsevier, 2018, pp. S34–S40.

[35] A. H. Thieme, Y. Zheng, G. Machiraju, C. Sadee, M. Mittermaier, M. Gertler, J. L. Salinas, K. Srinivasan, P. Gyawali, F. Carrillo-Perez *et al.*, "A deep-learning algorithm to classify skin lesions from mpox virus infection," *Nature Medicine*, vol. 29, no. 3, pp. 738–747, 2023.

[36] A. A. Cruz-Roa, J. E. Arevalo Ovalle, A. Madabhushi, and F. A. González Osorio, "A deep learning architecture for image representation, visual interpretability and automated basal-cell carcinoma cancer detection," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2013: 16th International Conference, Nagoya, Japan, September 22-26, 2013, Proceedings, Part II 16*. Springer, 2013, pp. 403–410.

[37] Y. Yuan, M. Chao, and Y.-C. Lo, "Automatic skin lesion segmentation using deep fully convolutional networks with jaccard distance," *IEEE transactions on medical imaging*, vol. 36, no. 9, pp. 1876–1886, 2017.

[38] P. Tschandl, N. Codella, B. N. Akay, G. Argenziano, R. P. Braun, H. Cabo, D. Gutman, A. Halpern, B. Helba, R. Hofmann-Wellenhof *et al.*, "Comparison of the accuracy of human readers versus machine-learning algorithms for pigmented skin lesion classification: an open, web-based, international, diagnostic study," *The lancet oncology*, vol. 20, no. 7, pp. 938–947, 2019.

[39] X. Sun, J. Yang, M. Sun, and K. Wang, "A benchmark for automatic visual classification of clinical skin disease images," in *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part VI 14*. Springer, 2016, pp. 206–222.

[40] T. H. Kung, M. Cheatham, A. Medenilla, C. Sillos, L. De Leon, C. Elepaño, M. Madriaga, R. Aggabao, G. Diaz-Candido, J. Maningo *et al.*, "Performance of chatgpt on usmle: Potential for ai-assisted medical education using large language models," *PLoS digital health*, vol. 2, no. 2, p. e0000198, 2023.

[41] M. Sallam, N. Salim, M. Barakat, and A. Al-Tammemi, "Chatgpt applications in medical, dental, pharmacy, and public health education: A descriptive study highlighting the advantages and limitations," *Narra J*, vol. 3, no. 1, pp. e103–e103, 2023.

[42] S. Bubeck, V. Chandrasekaran, R. Eldan, J. Gehrke, E. Horvitz, E. Kamar, P. Lee, Y. T. Lee, Y. Li, S. Lundberg *et al.*, "Sparks of artificial general intelligence: Early experiments with gpt-4," *arXiv preprint arXiv:2303.12712*, 2023.

[43] P. Lee, S. Bubeck, and J. Petro, "Benefits, limits, and risks of gpt-4 as an ai chatbot for medicine," *New England Journal of Medicine*, vol. 388, no. 13, pp. 1233–1239, 2023.

[44] M. Balas and E. B. Ing, "Conversational ai models for ophthalmic diagnosis: Comparison of chatgpt and the isabel pro differential diagnosis generator," *JFO Open Ophthalmology*, vol. 1, p. 100005, 2023.

[45] M. Mijwil, M. Aljanabi, and A. H. Ali, "Chatgpt: Exploring the role of cybersecurity in the protection of medical information," *Mesopotamian journal of cybersecurity*, vol. 2023, pp. 18–21, 2023.

[46] R. K. Sinha, A. D. Roy, N. Kumar, H. Mondal, and R. Sinha, "Applicability of chatgpt in assisting to solve higher order problems in pathology," *Cureus*, vol. 15, no. 2, 2023.

[47] F. Ufuk, "The role and limitations of large language models such as chatgpt in clinical settings and medical journalism," *Radiology*, p. 230276, 2023.

[48] M. Hu, S. Pan, Y. Li, and X. Yang, "Advancing medical imaging with language models: A journey from n-grams to chatgpt," *arXiv preprint arXiv:2304.04920*, 2023.

[49] R. Vaishya, A. Misra, and A. Vaish, "Chatgpt: Is this version good for healthcare and research?" *Diabetes & Metabolic Syndrome: Clinical Research & Reviews*, vol. 17, no. 4, p. 102744, 2023.

[50] E. J. Beltrami and J. M. Grant-Kels, "Consulting chatgpt: Ethical dilemmas in language model artificial intelligence," *Journal of the American Academy of Dermatology*, 2023.

[51] C. Li, C. Wong, S. Zhang, N. Usuyama, H. Liu, J. Yang, T. Naumann, H. Poon, and J. Gao, "Llava-med: Training a large language-and-vision assistant for biomedicine in one day," *arXiv preprint arXiv:2306.00890*, 2023.

[52] S. Wang, Z. Zhao, X. Ouyang, Q. Wang, and D. Shen, "Chatcad: Interactive computer-aided diagnosis on medical image using large language models," *arXiv preprint arXiv:2302.07257*, 2023.

[53] H. Li, D. Guo, W. Fan, M. Xu, and Y. Song, "Multi-step jailbreaking privacy attacks on chatgpt," *arXiv preprint arXiv:2304.05197*, 2023.

[54] B. Lund and D. Agbaji, "Information literacy, data literacy, privacy literacy, and chatgpt: Technology literacies align with perspectives on emerging technology adoption within communities," *Data Literacy, Privacy Literacy, and ChatGPT: Technology Literacies Align with Perspectives on Emerging Technology Adoption within Communities (January 14, 2023)*, 2023.

[55] P. Rajpurkar, E. Chen, O. Banerjee, and E. J. Topol, "Ai in health and medicine," *Nature medicine*, vol. 28, no. 1, pp. 31–38, 2022.

[56] J. Zhou, S. Chen, Y. Wu, H. Li, B. Zhang, L. Zhou, Y. Hu, Z. Xiang, Z. Li, N. Chen *et al.*, "Ppml-omics: a privacy-preserving federated machine learning system protects patients' privacy from omic data," *bioRxiv*, pp. 2022–03, 2022.

[57] J. Zhou, L. Zhou, D. Wang, X. Xu, H. Li, Y. Chu, W. Han, and X. Gao, "Personalized and privacy-preserving federated heterogeneous medical image analysis with pppml-hmi," *medRxiv*, pp. 2023–02, 2023.

[58] J. Zhou, H. Li, X. Liao, B. Zhang, W. He, Z. Li, L. Zhou, and X. Gao, "Audit to forget: A unified method to revoke patients' private data in intelligent healthcare," *bioRxiv*, pp. 2023–02, 2023.

[59] D. Zhu, J. Chen, X. Shen, X. Li, and M. Elhoseiny, "Minigpt-4: Enhancing vision-language understanding with advanced large language models," *arXiv preprint arXiv:2304.10592*, 2023.

[60] W.-L. Chiang, Z. Li, Z. Lin, Y. Sheng, Z. Wu, H. Zhang, L. Zheng, S. Zhuang, Y. Zhuang, J. E. Gonzalez, I. Stoica, , and E. P. Xing., "Vicuna: An open-source chatbot impressing gpt-4 with 90%* chatgpt quality, march 2023."

[61] H. Touvron, T. Lavril, G. Izacard, X. Martinet, M.-A. Lachaux, T. Lacroix, B. Rozière, N. Goyal, E. Hambro, F. Azhar *et al.*, "Llama: Open and efficient foundation language models," *arXiv preprint arXiv:2302.13971*, 2023.

[62] J. Li, D. Li, S. Savarese, and S. Hoi, "Blip-2: Bootstrapping language-image pre-training with frozen image encoders and large language models," *arXiv preprint arXiv:2301.12597*, 2023.

[63] Y. Fang, W. Wang, B. Xie, Q. Sun, L. Wu, X. Wang, T. Huang, X. Wang, and Y. Cao, "Eva: Exploring the limits of masked visual representation learning at scale," *arXiv preprint arXiv:2211.07636*, 2022.

[64] P. Sharma, N. Ding, S. Goodman, , and R. Soricut, "Conceptual captions: A cleaned, hypernymed, image alt-text dataset for automatic image captioning," *In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pages 2556–2565, 2018*.

[65] V. Ordonez, G. Kulkarni, and T. Berg, "Im2text: Describing images using 1 million captioned photographs," *Advances in neural information processing systems*, vol. 24, 2011.

[66] C. Schuhmann, R. Vencu, R. Beaumont, R. Kaczmarczyk, C. Mullis, A. Katta, T. Coombes, J. Jitsev, and A. Komatsuzaki, "Laion-400m: Open dataset of clip-filtered 400 million image-text pairs," *arXiv preprint arXiv:2111.02114*, 2021.

[67] E. Almazrouei, H. Alobeidli, A. Alshamsi, A. Cappelli, R. Cojocaru, M. Debbah, E. Goffinet, D. Heslow, J. Launay, Q. Malartic, B. Noune, B. Pannier, and G. Penedo, "Falcon-40B: an open large language model with state-of-the-art performance," 2023.

[68] C. Preiksaitis, C. A. Sinsky, and C. Rose, "Chatgpt is not the solution to physicians' documentation burden," *Nature Medicine*, pp. 1–2, 2023.

[69] W. Fu and C. J. Cockerell, "The actinic (solar) keratosis: a 21st-century perspective," *Archives of dermatology*, vol. 139, no. 1, pp. 66–70, 2003.

[70] G. R. Kanthraj, "The twenty factors that made teledermatology consultation a matured application: A systematic review," *Clinical Dermatology Review*, vol. 7, no. 1, pp. 10–15, 2023.

[71] W. Liang, M. Yuksekgonul, Y. Mao, E. Wu, and J. Zou, "Gpt detectors are biased against non-native english writers," 2023.

[72] R. Daneshjou, M. Yuksekgonul, Z. R. Cai, R. Novoa, and J. Y. Zou, "Skincon: A skin disease dataset densely annotated by domain experts for fine-grained debugging and analysis," *Advances in Neural Information Processing Systems*, vol. 35, pp. 18 157–18 167, 2022.

[73] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly *et al.*, "An image is worth 16x16 words: Transformers for image recognition at scale," *arXiv preprint arXiv:2010.11929*, 2020.

[74] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," *Advances in neural information processing systems*, vol. 30, 2017.