

Assessing the diagnostic capability of ChatGPT through clinical case scenarios in dermatology

Dear Editor,

ChatGPT is a natural language processing artificial intelligence (AI)-based chatbot that uses input via text and images to generate responses. It is not an app but rather an online user-friendly interface that analyses the data on the internet to provide the users with responses based on their input. ChatGPT has demonstrated its high capabilities in various fields including clinical practice.¹ An increasing number of studies have been conducted to explore the capability of ChatGPT in various medical fields.^{2,3} Most of these studies have resulted in the favourable conclusion that ChatGPT could be used as an effective complementary tool to assist clinicians.⁴

Delving into the realm of dermatological practice, the most critical aspect of good patient care is an accurate diagnosis of the disease. However, diverse clinical presentations and overlapping signs and symptoms in dermatological diseases often pose a challenge to clinicians in accurately diagnosing the condition. In this regard, artificial intelligence holds immense potential, particularly in diagnosing and generating personalised treatment plans. Hence, this letter seeks to further explore the diagnostic capabilities of ChatGPT using clinical case scenarios.

For this study, ChatGPT versions 3.5 and 4.0 were used. A total of ten clinical case scenarios pertaining to different dermatological conditions from a publicly available website were used.⁵ The clinical cases included the patient's presentation with detailed signs and symptoms. Also, a description of the skin abnormality was explained. At the end of the case scenario, a lead-in question about the diagnosis was asked with four multiple choices. The case scenarios were based on: widespread reticulate erythema on the abdomen, erythematous edematous plaques on the face and extremities, papule on a papillomatous brown plaque, widespread peeling reticulate erythema, firm round papule on the foot, hair loss after COVID-19, Swelling on knee, recurring reticulate erythema on back, yellow papules on

neck and arms and skin thickening around fingers. All the questions were typed on the input board of ChatGPT 3.5 and 4.0, and answers were generated. Each answer was generated twice to ensure that the algorithm does not directly generate answers from the websites due to the black-box effect. The black box effect refers to situations where the inner workings of a system, particularly in complex algorithms or machine learning models, are not transparent or understandable to users. The responses of ChatGPT 3.5 and 4.0 were cross-checked with the answer provided. The sample questions for this evaluation are provided below.

Case 1: Widespread rash on the abdomen

A 15-year-old adolescent with ulcerative colitis (UC) is evaluated for an unusual rash on the abdomen. The patient was admitted to the hospital for the management of pain related to UC. The patient reports that she noticed the rash developing slowly over the last few weeks. Over time, it has darkened and started to develop open sores. On examination, hyperpigmented and erythematous reticulated patches with scattered erosions are present on the central and lower abdomen. Upon further questioning, the patient reports that for several months she has regularly applied a heating pad to her abdomen to alleviate pain. What is your diagnosis?

- A. Cutis Marmorata
- B. Livedo reticularis
- C. Erythema ab igne
- D. Cutaneous COVID-19

Case 2: Erythematous edematous plaques on the face and extremities

A 38-year-old black woman presents with a history of relapsing rash. The patient reports experiencing 2–3 flares of the rash per year, generally in the summertime, over the past 20 years. The rash is extremely pruritic and involves the face and extremities but tends to spare the trunk. The rash

How to cite this article: Manoharan P, Surapaneni KM. Assessing the diagnostic capability of ChatGPT through clinical case scenarios in dermatology. *Indian J Dermatol Venereol Leprol*. doi: 10.25259/IJDVL_1267_2023

Received: November, 2023 **Accepted:** January, 2024 **Epub Ahead of Print:** May, 2024

DOI: 10.25259/IJDVL_1267_2023 **Supplementary are available on:** https://doi.org/10.25259/IJDVL_1267_2023

This is an open-access article distributed under the terms of the Creative Commons Attribution-Non Commercial-Share Alike 4.0 License, which allows others to remix, tweak, and build upon the work non-commercially, as long as the author is credited and the new creations are licensed under the identical terms.

resolves with the administration of oral steroids. Extensive rheumatologic serologic workup evaluating for systemic lupus erythematosus and dermatomyositis has been negative. Physical examination reveals annular erythematous plaques on the face and extremities with each plaque studded with a single central small flaccid bulla. What is the most likely diagnosis?

- a. Polymorphous light eruption
- b. Contact dermatitis
- c. Solar urticaria
- d. Erythropoietic porphyria

Out of the ten clinical cases, both ChatGPT 3.5 [Supplementary Material 1] and 4.0 [Supplementary Material 2] have managed to generate correct responses to nine questions. ChatGPT 3.5 has generated the wrong answer for the case scenario on Firm Round Papule on Foot. The answer generated by ChatGPT was 'Verruca vulgaris' whereas the answer was 'Poroma' in the key. But ChatGPT 4.0 has rightly diagnosed this case. The explanations of ChatGPT 4.0 were more accurate describing the characteristic 'smooth and dome-shaped' appearance of poroma. However, in the case of a 'swelling on the knee', ChatGPT 3.5 rightly diagnosed the condition as a ganglionic cyst, whereas ChatGPT 4.0 diagnosed it to be a lipoma. The main discrepancy in the explanations generated was based on the typical features of lipoma and ganglionic cysts. While ChatGPT 3.5 has considered the location of joints to be the most common site for ganglionic cysts, ChatGPT 4.0 has given the diagnosis based on size and skin over the bump. Apart from these discrepancies, the explanations given for each of the other responses were matching with the key. Also, ChatGPT provided reasons for why other options cannot be the answer.

The positive performances of ChatGPT versions 3.5 and 4.0 in the majority of cases, indicate its utility as a valuable complementary tool for clinicians. The proficiency of both models in providing accurate responses could potentially enhance diagnostic speed, enabling quicker assessments and preliminary insights into diverse dermatological conditions where treatment could be initiated rapidly. This efficiency could prove particularly advantageous in managing a high volume of cases or streamlining the initial phases of the diagnostic process. Moreover, these findings also suggest educational applications for ChatGPT. Its ability to process information and generate differential diagnoses could be harnessed for training purposes, aiding healthcare professionals in expanding their diagnostic skills and familiarising themselves with a broad spectrum of dermatological cases.

Despite these positive implications, the study emphasises the need for exercising caution in relying solely on ChatGPT

for diagnostic decisions. ChatGPT-4 has been trained on a larger and more diverse dataset compared to ChatGPT-3.5, and it incorporates improvements in its training algorithms. These enhancements enable it to better understand context, handle subtleties in language, and generate more coherent and contextually appropriate responses. ChatGPT 3.5 was more accurate in diagnosing the 'Swelling on the knee' case, while ChatGPT 4.0 showed better performance in identifying the 'Firm Round Papule on the Foot'. This highlights the variability in AI model performance and suggests that neither version is consistently superior. The models' fallibility, as evidenced by the misclassification in the cases mentioned, underscores the need for human oversight and validation in clinical settings. Furthermore, ethical considerations such as patient privacy, data security, and responsible AI deployment are paramount. The study underscores the continuous need for improvement, ongoing validation efforts, and clear regulatory frameworks to ensure the ethical and effective integration of Artificial Intelligence models, including ChatGPT, into the intricate landscape of dermatological practice. However, only ten clinical cases were used for this study. To generalise the findings, more studies that delve deeper into the diagnosis and treatment aspects should be conducted.

Declaration of patient consent

Patient's consent not required as there are no patients in this study.

Financial support and sponsorship

Nil.

Conflicts of interest

There are no conflicts of interest.

Use of artificial intelligence (AI)-assisted technology for manuscript preparation

The authors confirm that there was no use of artificial intelligence (AI)-assisted technology for assisting in the writing or editing of the manuscript and no images were manipulated using AI.

*Perumal Manoharan, Krishna Mohan Surapaneni*¹

Departments of Dermatology, Venereology and Leprosy, ¹Biochemistry, Medical Education, Research, Panimalar Medical College Hospital & Research Institute, Varadharajapuram, Poonamallee, Chennai, Tamil Nadu, India.

Corresponding author:

Dr. Krishna Mohan Surapaneni,
Departments of Biochemistry, Medical Education, Panimalar Medical College Hospital & Research Institute, Varadharajapuram, Poonamallee, Chennai, Tamil Nadu, India.
krishnamohan.surapaneni@gmail.com

References

1. ChatGPT. Available from: <https://chat.openai.com/auth/login>. [Accessed 24 November 2023].
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, 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, Yang HS. Assessing the accuracy and clinical utility of ChatGPT in Laboratory Medicine. *Clin Chem* 2023;69:939–40.
5. Clinical Advisor. Dermatology Clinic. Available from: <https://www.clinicaladvisor.com/home/dermatology-clinic/>. [Accessed 24 November 2023].

Net Letter 1

Assessing the diagnostic capability of ChatGPT through clinical case scenarios in dermatology

Perumal Manoharan, Krishna Mohan Surapaneni¹

Departments of Dermatology, Venereology and Leprosy, ¹Biochemistry, Medical Education, Research, Panimalar Medical College Hospital & Research Institute, Varadharajapuram, Poonamallee, Chennai, India.

DOI: 10.25259/IJDVL_1267_2023 PMID: ***

How to cite this article: Manoharan P, Surapaneni KM. Assessing the diagnostic capability of ChatGPT through clinical case scenarios in dermatology. Indian J Dermatol Venereol Leprol doi: 10.25259/IJDVL_1267_2023

Summary

ChatGPT is an advanced AI-based chatbot designed to understand and generate human-like text, making it a versatile tool for a wide range of applications including health. This study evaluates the diagnostic capabilities of ChatGPT versions 3.5 and 4.0 using ten dermatological case scenarios. Results showed both versions performed well, correctly diagnosing nine out of ten cases, indicating potential as valuable tools for clinicians. However, discrepancies in certain diagnoses highlight the need for human oversight. The study suggests ChatGPT's utility in enhancing diagnostic speed and educational applications, while emphasising the importance of ethical considerations and continuous improvement.