

Generative Artificial Intelligence in Medicine: Has ChatGPT the Potential in Assisting Dermatologists?

Adela-Vasilica GUDIU* and Lăcrămioara STOICU-TIVADAR

Faculty of Automation and Computer Science, Politehnica University Timișoara, Vasile Pârvan Blvd., no. 2, 300223 Timișoara, Romania.

E-mails: adela.gudiu@student.upt.ro; lacramioara.stoicu-tivadar@upt.ro;

* Author to whom correspondence should be addressed;

Abstract

Although Generative Artificial Intelligence has been around for quite some time, the recent surge in interest in this subject was caused by the launch of OpenAI's chatbot. ChatGPT is an artificial intelligence model that processes natural language and outputs information according to the input provided by the user. The responses are contingent upon the model used for the inquiry and the user's desired output. The launch of the 4o model created a frenzy among technology enthusiasts, as it bridged some of the gaps documented in the research literature for the previous models (i.e. acceptance of image and audio input/output). Considering the improved capabilities of the 4o version, there was room to ask whether this tool could be reliable in dermoscopic image diagnosis and, also, how the variety in possible diagnoses of input data might affect the accuracy of its responses. To test the accuracy of the 4o model two datasets were used, both containing dermoscopic images randomly chosen from the ISIC archive gallery (<https://gallery.isic-archive.com/>), with diagnoses confirmed through histopathology. The first dataset consisted of 60 dermoscopic images of multiple skin diseases, whereas the other included 60 images of nevus, carcinoma, and melanoma images (20 images per label). Two chat sessions were initiated on 10th of February using a ChatGPT Plus subscription, each corresponding to one of the datasets. In the first session, each inquiry asked for the most likely diagnosis, whereas in the second one a list with possible diagnoses was first provided, and only afterwards were the images input. The results show improved accuracy in the second chat, where approximately 53% of the images were labeled correctly, as opposed to the first one, which had a success rate of only 20%. The preliminary findings indicate that the 4o model shows promise as an assisted tool in diagnosis of dermoscopic images. However, further research is necessary to establish its reliability as a standalone diagnostic instrument.

Keywords: ChatGPT; Medical Education; Artificial Intelligence (AI); Chatbot; Dermatology.