SOHO ANNUAL MEETING 2024

Evaluating the effectiveness of ChatGPT in diagnosing hematologic malignancies

Ibrahim Khamees[1], Xiangni Wu[1], Hana Qasim[1], Haoran Li [2]

(1) University of Missouri-Kansas City, Kansas City, MO, (2) University of Kansas Cancer Center, Westwood, KS

UMKC School of Medicine

INTRODUCTION

- ChatGPT demonstrates tremendous potential across various industries, including medicine.
- The aim of this study is to assess the effectiveness of ChatGPT 4.0 in diagnosing hematologic malignancies using pathology images and descriptive texts.

METHODS

•Reference cases were collected from the American Society of Hematology (ASH) Image bank.

•Combined images and descriptive clinical data for each case were pasted along our prompts into ChatGPT 4.0 in two steps. The first step included minimal clinical data along with the case's images.

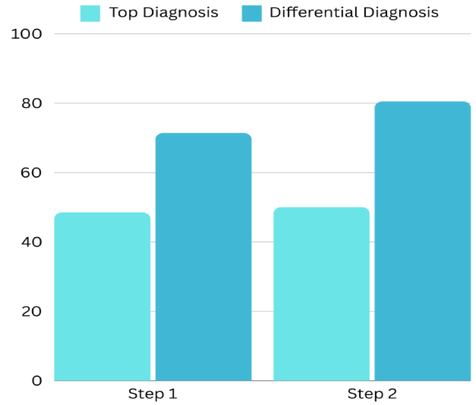
•Step 2 was only utilized if ChatGPT 4.0 failed to identify the correct diagnosis in step 1 and it included all the clinical data along with the combined images.

•ChatGPT 4.0 was asked to provide a top diagnosis and a list of differential diagnoses for each step based on probability.

•Percentages of providing the correct diagnosis or including the correct diagnosis in the list of differential diagnoses were calculated.

RESULTS

- 70 cases of hematologic malignancies were included in the analysis.
- In step 1, ChatGPT 4.0 provided the correct diagnosis as the top diagnosis in 48.5% of the cases and the correct diagnosis in the differential diagnoses list in 71% of the cases.
- Step 2 was run for the remaining 36 cases undiagnosed as the top diagnosis by step 1. ChatGPT 4.0 provided the correct diagnosis as the top diagnosis in 50% of these cases and the correct diagnosis in the differential list in 80.5% of the cases.
- Overall, the correct diagnosis was mentioned in the list of differential diagnoses in 91.4% of the cases.
- When it was not mentioned as the top diagnosis, the correct diagnosis was ranked 3.5 in step 1 and 2.1 in step 2.



Graph 1: Effectiveness of ChatGPT 4.0 in diagnosing hematologic malignancies provided by images and minimal clinical data (Step 1), and when provided by images and full clinical data (Step 2).

CONCLUSIONS

- ChatGPT 4.0 showed a moderate ability to diagnose hematologic malignancies based on clinical data and images.
- It can potentially help hematologists reach the correct diagnosis and narrow the differential diagnoses list. However, more studies are needed to evaluate its reliability for clinical application.

REFERENCES

1. https://imagebank.hematology.org/reference-cases/list
2. Dave T et al., ChatGPT in medicine: an overview of its applications, advantages, limitations, future prospects, and ethical considerations. Frontiers in artificial intelligence. 2023 May 4;6:1169595.
3. Apornvirat S et al., Comparing customized ChatGPT and pathology residents in histopathologic description and diagnosis of common diseases. Annals of Diagnostic Pathology. 2024 Dec 1;73:152359.