LETTER TO THE EDITOR
DOI: 10.4274/tjh.galenos.2025.2025.0225

Can ChatGPT Diagnose Hematological Diseases from Peripheral Blood Smear Images?

Serkan Güven[1], Deniz Esin Tekcan Şanlı[2]
[1]Mehmet Akif Ersoy State Hospital, Clinic of Hematology, Çanakkale, Türkiye
[2]Gaziantep University Faculty of Medicine, Department of Radiology, Gaziantep, Türkiye

Serkan Güven, M.D., Mehmet Akif Ersoy State Hospital, Clinic of Hematology, Çanakkale, Türkiye
merkmanuel@hotmail.com

June 10, 2025
July 17, 2025

Dear Editor,

Recent advancements in artificial intelligence (AI), especially in natural language processing, have led to the growing interest in integrating large language models (LLMs) into clinical workflows. ChatGPT, a widely used LLM, has demonstrated potential in generating coherent responses across a wide array of domains. However, its applicability in medical image interpretation, particularly hematological image analysis, remains a matter of scientific inquiry.

In this pilot study, we assessed ChatGPT's ability to analyze 42 peripheral blood smear (PBS) images representing common hematological conditions, including both malignant (n=34) and benign (n=8) cases. Images were selected from Anderson's Atlas of Hematology [1]. For each image, ChatGPT (GPT-4, OpenAI, April 2024 version) was prompted three times with identical instructions: Identify if there are any pathological cells, specify the cell type (RBC, WBC, PLT), suggest a preliminary diagnosis, and recommend further confirmatory tests.

ChatGPT consistently identified abnormal cells and correctly classified the cell type in 97.6% of cases. However, its diagnostic accuracy was limited: the correct diagnosis was achieved in 23.8% of cases overall, with better performance in acute leukemias (33.3%) and benign conditions (25%). Notably, ChatGPT failed to correctly diagnose any of the chronic leukemia cases. The model performed better in recommending appropriate confirmatory tests (e.g., bone marrow biopsy, flow cytometry) in 69% of cases.

We acknowledge several methodological limitations: (i) the use of standardized textbook images may introduce bias if similar data were part of ChatGPT's pretraining set, (ii) real-world clinical smear images may differ significantly in resolution, staining quality, and morphological complexity, and (iii) ChatGPT's output can vary due to prompt randomness and temperature parameters if not fixed. Moreover, the version used (GPT-4) is now surpassed by GPT-4o (May 2024), which may exhibit different capabilities [2].

Although our findings suggest that ChatGPT may assist in preliminary morphological recognition, its accuracy in disease diagnosis remains insufficient for clinical use. This is not unexpected, considering that hematological diagnoses, particularly for chronic leukemias, often rely on integrated morphological, immunophenotypic, cytogenetic, and molecular data [3,4]. Diagnostic accuracy is further hindered in the presence of low-resolution or poorly illuminated images, where subtle cytological clues are difficult to discern limitations inherent to LLMs without dedicated vision processing [5,6].

Several recent studies have evaluated LLMs in this domain. For example, Cai et al. demonstrated GPT-4o's capacity to identify abnormal cells with moderate accuracy in simulated blood morphology settings [5]. Negrini et al. and Pighi et al. further explored ChatGPT's application in classifying urinary and hematological particles, with mixed results and significant variability [6,7].

Our work aligns with these findings and underscores the need for controlled validation in real-world clinical datasets. To clarify further, our study focused primarily on morphological identification and did not incorporate molecular or immunophenotypic data into the diagnostic process.

A detailed breakdown of evaluated cases has been added to the supplementary table, including AML-M1, AML-M2, AML-M4, reactive lymphocytosis, iron deficiency anemia, and inflammatory thrombocytosis.

In conclusion, while ChatGPT holds promise as an adjunctive tool for hematology education or initial triage, it cannot replace expert interpretation of peripheral blood smears. Future directions should focus on hybrid models that integrate image processing and contextual language understanding, alongside rigorous validation on institution-specific datasets.

References

1. Anderson's Atlas of Hematology. 3rd Edition. Wolters Kluwer; 2021. ISBN: 9781975118259.
2. Wang C, Wei XL, Li CX, et al. Efficient and Highly Accurate Diagnosis of Malignant Hematological Diseases Based on Whole-Slide Images Using Deep Learning. Front Oncol. 2022;12:879308.
3. Barth D. Approach to peripheral blood film assessment for pathologists. Semin Diagn Pathol. 2012;29(1):31–48.
4. Bain BJ. Diagnosis from the Blood Smear. N Engl J Med. 2005;353:498–507.
5. Cai X, Zhan L, Lin Y. Assessing the accuracy and clinical utility of GPT-4O in abnormal blood cell morphology recognition. Digital Health. 2024;10:20552076241298503.
6. Pighi L, Negrini D, Tosi M, Lippi G. Performance assessment of ChatGPT in the classification of urine particles. Scand J Clin Lab Invest. 2025;85(3):223-227.
7. Negrini D, Pighi L, Tosi M, Lippi G. Evaluating the accuracy of ChatGPT in classifying normal and abnormal blood cell morphology. Clin Chem Lab Med (CCLM). 2025.

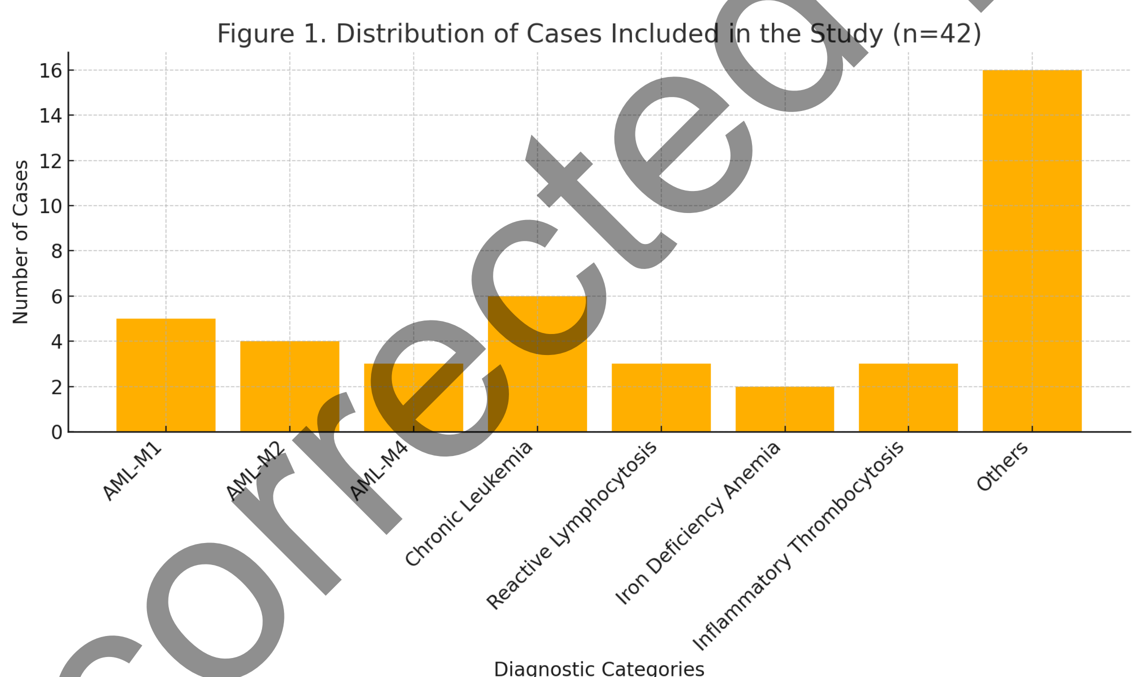

Figure 1. The figure below illustrates the diagnostic distribution of the 42 peripheral blood smear cases evaluated in this study.