marrow samples were processed for oncohematological karyotyping as part of the routine diagnostic workup for suspected neoplasms. To clarify the origin of certain findings, a second karyotype was performed using phytohemagglutinin stimulation, aimed at constitutional genetic analysis. Comparison of the two tests enabled a better understanding of the nature of the identified genetic aberration. **Results:** In five patients, cytogenetic abnormalities were found involving the sex chromosomes (X and Y). One patient presented a structural alteration consistent with a Robertsonian translocation. Additionally, 14 individuals were referred to the laboratory with previously identified constitutional abnormalities, with trisomy 21 being the most frequent and well-documented finding in the literature. In total, 0.4% of the individuals referred for oncohematological karyotyping exhibited genetic aberrations unrelated to hematologic malignancies. **Discussion and conclusion:** Clinical investigation of suspected hematologic disorders requires an integrated analysis of bone marrow samples, including cell morphology, immunophenotyping, and karyotyping. It is the combination of these tests that allows for a more accurate and reliable diagnostic approach in identifying hematologic neoplasms. In the cases analyzed, all genetic alterations were confirmed to be germline, exhibiting a non-clonal profile, often representing incidental findings. Proper interpretation of these variants — supported by consultation of public databases describing known aneuploidies and classical chromosomal abnormalities — is essential for correct clinical classification. The trained assessment by a qualified professional is equally important. Accurate characterization of these findings was crucial for guiding clinical decision-making and avoiding unnecessary treatments aimed at non-neoplastic genetic alterations.

https://doi.org/10.1016/j.htct.2025.104217

ID - 376

COMPARATIVE EffECTIVENESS OF HEMATOLOGISTS WITH LARGE LANGUAGE MODELS IN HEMATOLOGIC DIAGNOSIS

A Bunn Moreno, M Morgado Loureiro, R Doyle Portugal

Universidade Federal do Rio de Janeiro (UFRJ), Rio de Janeiro, RJ, Brazil

Introduction: Artificial intelligence (AI) use is rising and may benefit healthcare. Text-trained Large language model (LLM) seem to interpret images, but proof of accurate medical-image analysis is still scarce. **Objectives:** Primary objective: compare the diagnostic accuracy of a human expert vs four LLMs. **Material and methods:** We selected 3 - 10 images of microscopy fields to test 3 clinical situations. All images were derived from peripheral blood (PB) or bone marrow (BM) films, stained with WG and obtained with a cell phone camera. Ten test questions were developed based on the images. The images and questions were sent to a Hematologist and the following LLMs: ChatGPT 4.0 (CGPT4) and o3 (CGPTo3), Gemini Flash 2.5 (GF), and Claude Sonnet 4.0 (CS). Question type 1 sought to measure the machines' ability to evaluate images. Question type 2 provided a biased clinical context to test for confirmatory bias. Question type 3 provided a correct clinical context, aiming to obtain a correct diagnosis. The responses were compared to the standard response using a Likert scale as follows: 0 (no or incorrect answer); 1 point (correct answer with incorrect suggestions); 2 points (correct answer with or without suggestions). Considering the 10 questions, the score could vary from 0 to 20 points. **Results:** Images from 3 clinical situations (ClinSit) were tested, with the third clinical situation having two parts. ClinSit 1: pics of 10 fields of PB images of B-ALL at 1000x magnification. ClinSit 2: pics of 6 fields of BM images of multiple myeloma at 1000x magnification. ClinSit 3: BM images of acute myeloid leukemia at diagnosis and after treatment. At diagnosis, 3 pics of BM images at 1000x magnification were used, whereas after treatment, 9 images were used (magnification 100 × to 1000 ×). The overall score (maximum 20 points) was 18 for the Hematologist, 8 for CGPT4, 11 for CGPTo3, 16 for GF and 12 for CS. Regarding the image-specific questions: Hematologist scored 8/8, CGPT4 and o3 3/8, GF and CS 6/8. Confirmation bias questions: Hematologist scored 6/6, CGPT4 3/6, CGPTo3 and CS 2/6, and GF 4/6. Diagnostic questions: Hematologist scored 4/6, CGPT4 2/6, CS 4/6, and the CGPTo3 and GF models 6/6. **Discussion and conclusion:** We present a proof of concept: LLM AI as a potential hematology image analyst. We observed that machines could interpret these images, achieving a score of 8 to 16 (maximum 20), lower than the experienced hematologist (18/ 20). To date, there is no robust data that LLM AI can provide reliable answers based on images. For descriptive image analysis, the two CGPT models (4.0 and o3) performed worse than GF and CS, suggesting that the latter are more reliable for image interpretation. Carefully designed prompts can improve the accuracy of AI generated content. When using a partially correct context but a biased question, the different machine models were unable to maintain response coherence. This is a known flaw in these systems, which can have undesirable consequences. The physician's only incorrect answer was in scenario 3, question 3, possibly due to image quality. The performance of GF and CGPTo3 in type 3 questions was surprising. Limitations include small sample/physician numbers, expert-selected images, and ongoing model advances that may now surpass our results. Despite their text focus, LLMs may aid image analysis with open-ended prompts, but decision-making errors are still possible.

https://doi.org/10.1016/j.htct.2025.104218

ID – 2419

COMPARATIVO DOS PARÂMETROS PRÉ E PÓS TRANSFUSIONAIS PARA ANÁLISE DA RESPOSTA À TRANSFUSÃO.

MDS Arruda, LTD Marqui, CMS Assato, DCGP Magalhães, PDSPD Lima

HEMOVIDA - Hematologia e Hemoterapia de Bauru Ltda, Bauru, SP, Brasil