

IA NO DIAGNÓSTICO DE LEISHMANIOSE E PARACOCCIDIOIDOMICOSE: COMPARANDO GPT, GEMINI E CLAUDE - UM RELATO DE EXPERIÊNCIA EM AVANÇO TECNOLÓGICO

Ana Gabriella de Souza Mendes, Ana Júlia Ribeiro Gomes, Anne Yasmin de Oliveira Alves Monteiro, Felipe Camilo Santiago Veloso, Giovanni Monteiro Ribeiro, Maria Cecília Valadares Ribeiro, Yasmin Osama Abdel Wadud, Yasser Wadud Issler

Introdução: Diferenciar leishmaniose tegumentar (LT) e paracoccidiodomicose (PCM) é desafiador pela sobreposição de sinais clínicos e padrões morfológicos, podendo exigir exames invasivos. Modelos de Inteligência Artificial Large Language Model (IA) vem sendo testados por profissionais e alunos para o diagnóstico diferencial, mas essa prática carece de indicações

Objetivo: Comparar o desempenho de três modelos de IA no diagnóstico diferencial entre LT e PCM em etapas progressivas de informação **Método:** Trata-se de um relato de experiência conduzido entre 08/2025 a 09/2025, através de busca na base PubMed com os descritores: leishmaniasis; paracoccidiodomycosis; case report. Foram incluídos casos que apresentavam anamnese e imagem macroscópica da lesão, excluindo informações que revelassem diretamente o diagnóstico. Para evitar vies, testaram-se GPT-4, Claude 3 e Gemini 1.5 em sessões independentes, e dentro delas os prompts eram aplicados em sequência: Prompt 1 (P1) apenas imagem, Prompt 2 (P2) imagem+anamnese e Prompt 3 (P3) decisão binária (LT ou PCM). Além disso, aplicou-se um teste isolado de decisão binária apenas com imagens, para descartar a hipótese nula de acertos ao acaso

Resultados: Foram selecionados 37 casos (17 LT; 20 PCM), totalizando 148 interações com cada IA. O desempenho por prompt mostrou que, à medida que mais informações eram fornecidas, as IAs melhoraram seus acertos: GPT P1 24,3%, P2 51,4% e P3 81,1%; Claude P1 7,9%, P2 47,1% e P3 83,7%; Gemini P1 26,2%, P2 47,4% e P3 79,6%. Esses resultados indicam que o GPT apresentou aprendizado mais consistente, enquanto Claude e Gemini mostraram maior sensibilidade à quantidade e qualidade de informações fornecidas. No teste complementar com imagens, todas apresentaram desempenho acima do esperado ao acaso (GPT 71,9%, Gemini 62,2% e Claude 61,5%), resultado que indica acertos não aleatórios. Curiosamente, o Gemini destacou-se nas fases iniciais, sugerindo eficiência em reconhecimento visual puro, enquanto o Claude parecia se beneficiar mais da integração de dados clínicos. Esses padrões evidenciam estratégias diferentes de processamento entre as IA, oferecendo insights sobre como combinar abordagens de visão e linguagem para melhorar diagnósticos automatizados **Conclusão:** O uso de IA Large Language Model aparenta ser capaz de diferenciar quadros semelhantes apenas quando as opções são limitadas, mostrando que seu uso indiscriminado no diagnóstico ainda não deve ser implementado

Palavras-chave: Leishmaniose tegumentar, Paracoccidiodomicose, IA, Diagnóstico diferencial, Large Language Model