

Avaliação da acurácia diagnóstica do ChatGPT na emissão de laudos de tomografias de coerência óptica

Assessment of diagnostic accuracy of ChatGPT in optical coherence tomography reports

Alice Gomes de Almeida¹ , Vitor Barbosa Cerqueira¹ 

¹ Cirurgia Ocular de São Cristóvão, Rio de Janeiro, RJ, Brasil.

Almeida AG, Cerqueira VB. Avaliação da acurácia diagnóstica do ChatGPT na emissão de laudos de tomografias de coerência óptica. Rev Bras Oftalmol. 2025;84:e0089.

doi:

<https://doi.org/10.37039/1982.8551.20250089>

Descritores:

Tomografia de coerência óptica; Inteligência artificial; Diagnóstico por imagem; Acurácia diagnóstica; Edema macular; Membrana epirretiniana; Coriorretinopatia serosa central; Macula lútea

Keywords:

Optical coherence tomography; Artificial intelligence; Image diagnosis; Diagnostic accuracy; Macular edema; Epiretinal membrane; Central serous chorioretinopathy; Macula lutea

Recebido:
29/7/2025

Aceito:
25/8/2025

Autor correspondente:

Alice Gomes de Almeida
Rua Visconde de Pirajá, 414.Sl 421-422.
Ipanema
CEP: 22410-002 – Rio de Janeiro, RJ, Brasil
E-mail: alicegda@hotmail.com

Instituição de realização do trabalho:
Cirurgia Ocular de São Cristóvão, Rio de Janeiro, RJ, Brasil.

Fonte de auxílio à pesquisa:
não financiado.

Conflitos de interesse:
não há conflitos de interesses.

Declaração de Disponibilidade de Dados:
Os conjuntos de dados gerados e/ou analisados durante o estudo atual estão incluídos no manuscrito.

Editor Associado:

Ricardo Augusto Paletta Guedes
Universidade Federal de Juiz de Fora, Juiz de Fora, MG, Brasil
<https://orcid.org/0000-0002-9451-738X>



Copyright ©2025

RESUMO

Objetivo: Avaliar a acurácia diagnóstica do ChatGPT na análise de imagens de tomografia de coerência óptica.

Métodos: Foram analisadas 300 imagens de tomografia de coerência óptica previamente laudadas por especialistas, avaliadas pelo ChatGPT em três cenários: apenas com a imagem; com dados epidemiológicos adicionais; e com dados clínicos detalhados. Sensibilidade, especificidade e taxa de acurácia foram comparadas entre os cenários, utilizando análises estatísticas apropriadas ($p < 0,05$).

Resultados: A acurácia geral foi de 66,33% no cenário sem dados adicionais, 82,33% com dados epidemiológicos e 92,67% com dados clínicos. A sensibilidade para edema macular aumentou de 83,9% (cenário 1) para 100% (cenário 3). Especificidade foi de 100% em todos os cenários. Alterações como membrana epirretiniana e buraco macular apresentaram as menores acurácias, mesmo com dados adicionais (91,5% e 95,4%, respectivamente).

Conclusão: Os resultados destacam o impacto positivo da integração de dados contextuais no desempenho do ChatGPT, corroborando estudos prévios sobre Inteligência Artificial no diagnóstico médico. Embora promissor, o modelo apresenta limitações na análise de alterações mais complexas e dependência da qualidade dos dados fornecidos. Estudos adicionais são necessários para validar sua aplicação clínica em larga escala.

ABSTRACT

Objective: To assess the diagnostic accuracy of ChatGPT in analyzing OCT images.

Methods: A total of 300 OCT images, previously interpreted by specialists, were analyzed by ChatGPT under three scenarios: image-only input; input with additional epidemiological data; and input with detailed clinical data. Sensitivity, specificity, and accuracy rates were compared across scenarios using appropriate statistical analyses ($p < 0.05$).

Results: Overall accuracy was 66.33% in the image-only scenario, 82.33% with epidemiological data, and 92.67% with clinical data. Sensitivity for macular edema increased from 83.9% (scenario 1) to 100% (scenario 3). Specificity was 100% across all scenarios. Alterations such as epiretinal membrane and macular hole showed lower accuracy rates, even with additional data (91.5% and 95.4%, respectively).

Conclusion: The results highlight the positive impact of contextual data integration on ChatGPT's performance, corroborating previous studies on AI in medical diagnosis. Although the model is promising, it has limitations in analyzing more complex alterations and relies heavily on the quality of the data provided. Further studies are needed to validate its large-scale clinical application.

INTRODUÇÃO

As tecnologias de imagem multimodal têm um papel essencial no manejo de doenças oftalmológicas, particularmente no diagnóstico, tratamento e prognóstico de patologias retinianas.⁽¹⁾ A tomografia de coerência óptica (OCT) é uma técnica de imagem não invasiva que permite a visualização detalhada das microestruturas da retina. Observa-se que diversas características detectadas nas imagens estão relacionadas à função visual, possibilitando inclusive a estimativa da acuidade visual por meio da análise de microestruturas maculares.⁽¹⁾

Estudos recentes têm demonstrado o potencial da Inteligência Artificial (IA) para aprimorar esse processo. Modelos de IA podem identificar e mapear fluidos retinianos em condições como oclusões venosas da retina (OVR), degeneração macular relacionada à idade (DMRI) e edema macular diabético (EMD). Além disso, a IA não só é capaz de identificar a DMRI exsudativa, mas também de prever sua progressão.⁽²⁾

A doença oftalmológica pode ter um impacto profundo na qualidade de vida, e o diagnóstico e o tratamento precoces são essenciais para preservar a visão e prevenir complicações graves.⁽³⁾ Ferramentas eletrônicas de suporte ao diagnóstico diferencial têm sido usadas há muitos anos, auxiliando profissionais de saúde no diagnóstico de diversas condições médicas. Entre as mais conhecidas e eficazes está o *Isabel Pro Differential Diagnosis Generator*, que oferece suporte para diagnósticos de condições adultas e pediátricas desde 2006.⁽⁴⁾ Embora o Isabel e outras ferramentas possam melhorar a precisão e a eficiência diagnóstica, elas também apresentam limitações, como a necessidade de dados estruturados e a falta de processamento de linguagem contextualizado.

Modelos de linguagem de IA conversacional, como o ChatGPT, apresentam uma alternativa promissora para o suporte ao diagnóstico diferencial. Treinados em vastos conjuntos de dados textuais, esses modelos conseguem interpretar entradas em linguagem natural e fornecer respostas em tempo real, simulando interações humanas. Ao inserir descrições de casos nesses modelos, médicos podem obter uma lista de diagnósticos potenciais e orientações sobre opções de tratamento.^(2,5)

Este estudo busca avaliar a acurácia diagnóstica do ChatGPT na análise de imagens de OCT. Compreender as forças e as limitações dessas ferramentas pode apoiar o desenvolvimento de sistemas de diagnóstico diferencial mais eficientes e melhorar o atendimento a pacientes com doenças oftalmológicas.

MÉTODOS

Desenho do estudo

A presente pesquisa consiste em um estudo de acurácia diagnóstica.

Coleta de dados

Inicialmente, foi realizada a coleta de um grande conjunto de dados de imagens de OCT de pacientes com diferentes doenças oculares. Essas imagens foram obtidas a partir de um banco de imagens desidentificadas, garantindo a confidencialidade e o anonimato dos pacientes. O banco de imagens utilizado foi do Serviço de Cirurgia Ocular São Cristóvão (COSC), no Rio de Janeiro. Os critérios de inclusão das imagens foram exames de OCT de mácula de janeiro a outubro de 2024.

Ferramenta de Inteligência artificial

Cada uma das imagens foi copiada diretamente como entrada em um grande modelo de IA de linguagem chamado *ChatGPT* (<https://chat.openai.com/>), seguido pelo comando: “forneça o laudo dessa OCT”. O ChatGPT é uma implementação publicamente disponível e de livre acesso do modelo de linguagem *Generative Pre-Training Transformer 4* (GPT-4) desenvolvido pela OpenAI. O GPT-4 é atualmente um dos maiores modelos de processamento de linguagem natural baseado em rede neural, treinado em 175 bilhões de parâmetros para gerar texto semelhante ao humano.⁽⁶⁾ Ele funciona como um *chatbot* de uso geral e foi projetado para executar várias tarefas, incluindo tradução de linguagem, resumo e resposta a perguntas. Entre seus muitos usos possíveis, avaliamos sua capacidade de formar diagnósticos e laudos de OCT em três situações: dada a imagem de OCT sem a descrição do caso; adicionando dados epidemiológicos do caso; adicionando dados clínicos do caso. Como pode haver alguma variabilidade nas respostas geradas usando o mesmo texto de entrada, incluímos apenas os resultados obtidos na primeira tentativa.

Avaliação dos resultados

Todas as imagens foram previamente laudadas por oftalmologistas para comparação com os laudos emitidos pela IA. Apenas os laudos emitidos por oftalmologistas foram entregues aos pacientes.

Definições adotadas

Acurácia

Refere-se ao número de sujeitos corretamente classificados pelo instrumento em uma dada população.

Especificidade

Avalia a capacidade de um teste detectar os verdadeiros negativos entre os indivíduos não doentes, ou seja, afasta a doença quando ela está de fato ausente.

Sensibilidade

Avalia a capacidade de um teste detectar os verdadeiros positivos entre os indivíduos doentes, ou seja, detectar a doença quando ela está de fato presente.

Análise estatística

Inicialmente, foi realizada análise descritiva das condições avaliadas, utilizando média, desvio-padrão, mediana e valores máximos e mínimos de proporção, frequências e porcentagens, conforme o tipo de variável.

Acurácia, especificidade e sensibilidade foram medidas de acordo com as definições reportadas acima para cada tipo de alteração na OCT (drusas, atrofia retiniana, fibrose sub-retiniana, buraco macular [BM], coriorretinopatia serosa central [CRSC], edema macular [EM], membrana epirretiniana [MER], membrana neovascular sub-retiniana [MNSR] e tração vitreomacular [TVM]) e *prompt* utilizado (sem dados, com dados epidemiológicos e com dados clínicos) para gerar o laudo.

Calculamos a taxa de acerto do ChatGPT (razão entre o número de diagnósticos corretos e o total de imagens de OCT analisadas) para cada tipo de *prompt* utilizado, tipo de alteração presente na OCT e para a sequência dos casos (dez primeiros casos, dez casos intermediários e os dez últimos casos apresentados). Por fim, avaliamos a influência do tipo de *prompt*, tipo de alteração na OCT e ganho de experiência da IA comparando as taxas de acerto do ChatGPT em cada cenário por meio do teste do qui-quadrado. Em todas as análises foi considerado o nível de significância $< 0,05$. A análise estatística foi realizada usando o programa *Statistical Package for the Social Sciences* (SPSS), versão 20.0, SPSS Inc, Chicago, 2011.

RESULTADOS

Foram selecionadas 300 imagens de OCT previamente laudadas por oftalmologistas para serem avaliadas pelo ChatGPT. No total, foram realizadas 900 análises de imagem de OCT pelo ChatGPT, sendo 300 (33,33%) com o *prompt* contendo apenas a imagem de OCT, 300 (33,33%) com o *prompt* adicionando dados epidemiológicos do caso e 300 (33,33%) com o *prompt* adicionando dados clínicos do caso.

Nas análises das imagens de OCT com *prompt* contendo apenas a imagem e o comando “forneça o laudo

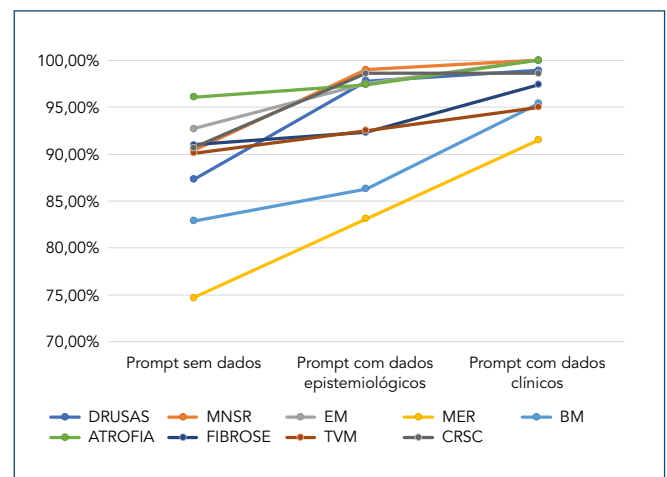
dessa OCT”, as acurácias para o diagnóstico das seguintes alterações foram: drusas = 87,30%; MNSR = 90,40%; EM = 92,70%; MER = 74,70%; BM = 82,90%; atrofia retiniana = 96,10%; fibrose sub-retiniana = 91%; TVM = 90,10%; CRSC = 90,70% (Tabela 1). Já nas análises das imagens de OCT com *prompt* adicionando dados epidemiológicos, as acurácias foram: drusas = 97,80%; MNSR = 99%; EM = 97,50%; MER = 83,10%; BM = 86,30%; atrofia = 97,40%; fibrose = 92,30%; TVM = 92,50%; CRSC = 98,60%. Por fim, nas análises das imagens de OCT com *prompt* adicionando dados clínicos, as acurácias foram: drusas = 98,90%; MNSR = 100%; EM = 100%; MER = 91,50%; BM = 95,40%; atrofia = 100%; fibrose = 97,40%; TVM = 95%; CRSC = 98,60%. A evolução das acurácias para cada alteração e *prompt* utilizado está representada na figura 1.

Tabela 1. Acurácia do ChatGPT no diagnóstico de alterações na tomografia de coerência óptica

Alteração	Prompt apenas com a imagem de OCT	Grupos Prompt com dados epidemiológicos	Prompt com dados clínicos
Drusas	87,30	97,80	98,90
MNSR	90,40	99	100
EM	92,70	97,50	100
MER	74,70	83,10	91,50
BM	82,90	86,30	95,40
Atrofia	96,10	97,40	100
Fibrose	91	92,30	97,40
TVM	90,10	92,50	95
CRSC	90,70	98,60	98,60

Resultados expressos como %.

MNSR: membrana neovascular sub-retiniana; EM: edema macular; MER: membrana epirretiniana; BM: buraco macular; TVM: tração vitreomacular; CRSC: coriorretinopatia serosa central.



MNSR: membrana neovascular sub-retiniana; EM: edema macular; MER: membrana epirretiniana; BM: buraco macular; TVM: tração vitreomacular; CRSC: coriorretinopatia serosa central.

Figura 1. Acurácia do ChatGPT no diagnóstico de cada alteração na tomografia de coerência óptica por tipo de *prompt* utilizado.

Em relação à especificidade do ChatGPT no diagnóstico das imagens de OCT, em todas as 204 vezes que

uma imagem de OCT normal foi enviada para análise do ChatGPT, foi constatado que se tratava de uma OCT normal em 100% dos casos, independentemente do *prompt* utilizado; logo, a especificidade do ChatGPT no diagnóstico das imagens de OCT foi de 100% (Tabela 2).

Tabela 2. Sensibilidade e Especificidade do ChatGPT no diagnóstico de alterações na tomografia de coerência óptica

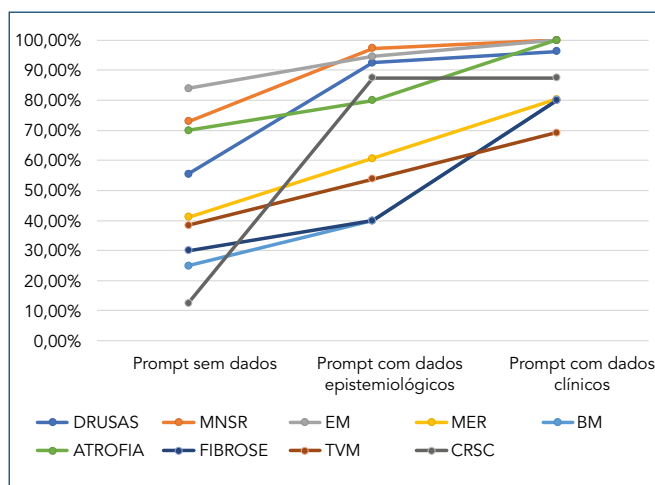
Alteração	Prompt apenas com a imagem de OCT	Grupos Prompt com dados epidemiológicos	Prompt com dados clínicos
Sensibilidade			
Drusas	55,50	92,50	96,20
MNSR	72,90	97,30	100
EM	83,90	94,60	100
MER	41,10	60,70	80,30
BM	25,00	40,00	80,00
Atrofia	70,00	80,00	100
Fibrose	30	40,00	80,00
TVM	38,40	53,80	69,20
CRSC	12,50	87,50	87,50
Especificidade			
100%: independente do <i>prompt</i> o ChatGPT reconheceu uma OCT normal em todas as tentativas			

Resultados expressos como %.

MNSR: membrana neovascular sub-retiniana; EM: edema macular; MER: membrana epirretiniana; BM: buraco macular; TVM: tração vitreomacular; CRSC: coriorretinopatia serosa central; OCT: tomografia de coerência óptica.

Quanto à sensibilidade, nas análises das imagens de OCT com *prompt* contendo apenas a imagem e o comando “forneça o laudo dessa OCT”, as sensibilidades para o diagnóstico das seguintes alterações foram: drusas = 55,50%; MNSR = 72,90%; EM = 83,90%; MER = 41,10%; BM = 25,00%; atrofia = 70,00%; fibrose = 30%; TVM = 38,40%; CRSC = 12,50%. Já nas análises das imagens de OCT com *prompt* adicionando dados epidemiológicos, as sensibilidades foram: drusas = 92,50%; MNSR = 97,30%; EM = 94,60%; MER = 60,70%; BM = 40,00%; atrofia = 80,00%; fibrose = 40,00%; TVM = 53,80%; CRSC = 87,50%. Por fim, nas análises das imagens de OCT com *prompt* adicionando dados clínicos, as sensibilidades foram: drusas = 96,20%; MNSR = 100%; EM = 100%; MER = 80,30%; BM = 80,00%; atrofia = 100%; fibrose = 80,00%; TVM = 69,20%; CRSC = 87,50%. A evolução das sensibilidades para cada alteração e o *prompt* utilizado estão representados na figura 2.

Quanto à taxa de acerto do ChatGPT no diagnóstico das imagens de OCT para cada tipo de *prompt*, observamos: 66,33% (199/300) de taxa de acerto para o *prompt* apenas com a imagem de OCT, 82,33% (247/300) de taxa de acerto para o *prompt* adicionando dados epidemiológicos e 92,67% (278/300) de taxa de acerto para o *prompt* adicionando dados clínicos, sendo essa diferença na taxa de acerto de acordo com o *prompt* utilizado estatisticamente significativa $p < 0,0001$ (Tabela 3 e Figura 3).



MNSR: membrana neovascular sub-retiniana; EM: edema macular; MER: membrana epirretiniana; BM: buraco macular; TVM: tração vitreomacular; CRSC: coriorretinopatia serosa central.

Figura 2. Sensibilidade do ChatGPT no diagnóstico de cada alteração na tomografia de coerência óptica por tipo de *prompt* utilizado.

Tabela 3. Análise da taxa de acerto do ChatGPT e fatores associados

Variável	Acertos	Erros	Valor de p
Prompt			< 0,0001*
Sem dados	199 (66,33)	101 (33,67)	
Com dados epidemiológicos	247 (82,33)	53 (17,67)	
Com dados clínicos	278 (92,67)	22 (7,33)	0,114
Sequência dos casos independente do prompt			
10 primeiros casos	236 (78,67)	64 (21,33)	
10 casos intermediários	253 (84,33)	47 (15,67)	
10 últimos casos	235 (78,33)	65 (21,67)	< 0,0001*
Alteração independente do prompt			
Drusas	66 (81,48)	15 (18,52)	
MNSR	100 (90,1)	11 (9,9)	
EM	156 (92,86)	12 (7,14)	
MER	93 (60,78)	60 (39,21)	
BM	29 (48,33)	31 (51,67)	
Atrofia	25 (83,33)	5 (16,67)	
Fibrose	15 (50)	15 (50)	
TVM	21 (53,84)	18 (46,16)	
CRSC	15 (62,5)	9 (37,5)	

Resultados expressos como n (%).

*Estatisticamente significativo.

MNSR: membrana neovascular sub-retiniana; EM: edema macular; MER: membrana epirretiniana; BM: buraco macular; TVM: tração vitreomacular; CRSC: coriorretinopatia serosa central.

Na investigação dos outros fatores associados a taxa de acerto do ChatGPT no diagnóstico das imagens de OCT, houve uma associação entre o tipo de alteração presente na imagem de OCT e a taxa de acerto ($p < 0,0001$), sendo a melhor e a pior taxa de acerto, respectivamente, EM = 92,86% e BM = 48,33%. Por fim, não houve associação da taxa de acerto com a sequência de envio dos casos ($p = 0,114$).

DISCUSSÃO

Os resultados deste estudo destacam o potencial do ChatGPT como uma ferramenta auxiliar no diagnóstico

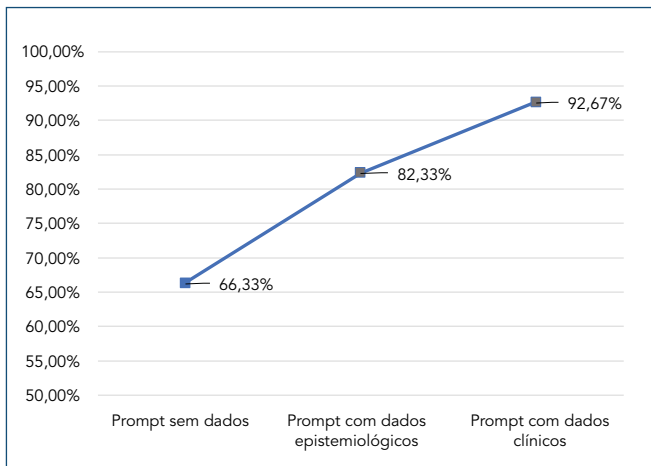


Figura 3. Taxa de acerto do ChatGPT por tipo de prompt utilizado.

de alterações em OCT, apresentando desempenho variável conforme o tipo de *prompt* utilizado. De forma geral, as acurácias observadas demonstraram que a integração de dados clínicos e epidemiológicos aumenta significativamente a precisão diagnóstica da IA, corroborando estudos anteriores sobre a relevância do contexto no processamento de dados pela IA.⁽⁷⁻⁹⁾

A análise da acurácia demonstrou que, sem dados adicionais, o ChatGPT atingiu taxas consideráveis para alterações como atrofia (96,1%), EM (92,7%) e membrana neovascular sub-retiniana (90,4%). Contudo, a inclusão de dados epidemiológicos e clínicos elevou as acurácias para quase 100% em várias condições, como EM e atrofia. Esses achados indicam que o uso de informações adicionais aprimora a compreensão contextual do modelo, alinhando-se à literatura que evidencia a superioridade da IA em condições com maior disponibilidade de dados estruturados.⁽¹⁰⁻¹²⁾

Especificamente, observa-se que a MER e o BM apresentaram taxas de acurácia menores em comparação a outras alterações, mesmo com a adição de dados clínicos (91,5% e 95,4%, respectivamente). Uma tendência inicial identificada foi a confusão de membranas epirretinianas incipientes com exames normais, sugerindo que alterações mais sutis podem ser desafiadoras para a IA sem dados contextuais adequados. Esse comportamento reflete a necessidade de interpretação mais detalhada das imagens por especialistas, conforme estudos já publicados.^(13,14)

Estudos anteriores também relatam desafios semelhantes no diagnóstico de alterações complexas com IA. Por exemplo, Schlegl et al. descreveram um modelo de *deep learning* que atingiu acurácia de 87% para identificação de fluidos maculares, similar às taxas iniciais observadas para EM neste estudo.⁽¹⁾ Além disso, Hormel

et al. demonstraram que a IA pode superar especialistas humanos em alguns diagnósticos quando dados completos são fornecidos, reforçando a importância de *prompts* detalhados.⁽²⁾

Ao não fornecer dados adicionais, erros comuns incluíram a confusão de membrana neovascular sub-retiniana com EM e de BM com edema macular. Esses erros diminuíram significativamente à medida que os dados epidemiológicos e clínicos foram incorporados. A confusão de tração vitreomacular com edema macular foi outro padrão observado, reforçando a necessidade de descrições mais completas para minimizar diagnósticos incorretos.

Em termos de sensibilidade, o ChatGPT demonstrou ganhos expressivos ao incluir dados epidemiológicos e clínicos, especialmente em condições como drusas (de 55,5% para 96,2%), EM (de 83,9% para 100%) e BM (de 25% para 80%). Essas melhorias são consistentes com a literatura sobre *deep learning*, que aponta maior eficiência na detecção de padrões quando os modelos recebem descrições contextuais detalhadas.^(15,16)

Em contrapartida, Bond et al. avaliaram ferramentas de diagnóstico diferencial baseadas em IA e observaram sensibilidade média de 65% para aplicações médicas gerais.⁽⁵⁾ Comparativamente, os resultados deste estudo demonstram avanços significativos na utilização de IA no campo oftalmológico.

A especificidade do ChatGPT foi notavelmente alta, alcançando 100% em todas as tentativas de identificação de exames normais. Este resultado está alinhado com pesquisas anteriores que destacam a capacidade de modelos de IA em excluir diagnósticos errôneos quando nenhuma anormalidade está presente, reduzindo o risco de alarmes falsos.^(17,18) Schlegl et al. também relataram especificidade de 98% em seu modelo, reforçando a comparação positiva.⁽¹⁾

A evolução das taxas de acerto conforme o *prompt* utilizado também é um ponto a ser destacado. Com apenas a imagem de OCT, a taxa de acerto foi de 66,33%, enquanto com a adição de dados clínicos, chegou a 92,67%. Esse aumento reforça a ideia de que a IA depende diretamente da qualidade e da quantidade de dados fornecidos para otimização do diagnóstico, sendo operador-dependente.

As alterações, como EM (92,86%) e membrana neovascular sub-retiniana (90,1%), apresentaram as melhores taxas de acerto globais, enquanto BM (48,33%) e fibrose sub-retiniana (50%) tiveram as piores. Essa variação pode ser atribuída à complexidade intrínseca das imagens e à necessidade de maior refinamento nos algoritmos da

IA para tais condições. No entanto, os dados também sugerem uma curva de aprendizado, com assertividade crescente observada nos últimos pacientes analisados, reforçando o impacto do treinamento e da experiência no desempenho do modelo.

No contexto de aplicações clínicas, a taxa de acerto geral do ChatGPT sugere que ele pode ser usado como uma ferramenta complementar à interpretação médica, especialmente em cenários onde há necessidade de segunda opinião ou triagem de grandes volumes de exames. Estudos comparativos com outras plataformas de IA, como as desenvolvidas especificamente para diagnósticos oftalmológicos, são necessários para validar ainda mais sua eficiência.

Uma limitação deste estudo foi a variabilidade intrínseca nas respostas do modelo, uma vez que a IA pode apresentar inconsistências em suas análises dependendo da forma como os *prompts* são estruturados. Assim, a padronização no fornecimento de dados é essencial para garantir maior consistência nos resultados.

Outro aspecto a ser considerado é que, embora o ChatGPT tenha demonstrado alta acurácia em várias condições, ele não foi projetado especificamente para análise de imagens médicas. Modelos treinados exclusivamente com dados de OCT poderiam apresentar desempenhos ainda melhores, conforme sugerido por estudos sobre IA especializada.⁽¹⁹⁻²²⁾

Comparando os achados deste estudo com a literatura existente, observa-se que o ChatGPT possui desempenho competitivo em relação a outras tecnologias baseadas em IA, como redes neurais convolucionais. Estudos anteriores demonstraram acurácias semelhantes para condições como EM e drusas, reforçando a viabilidade do uso do ChatGPT em cenários reais.⁽⁵⁾

Adicionalmente, os dados apontam para a importância do treinamento de modelos com bases de dados diversificadas e representativas. Uma maior heterogeneidade nos conjuntos de dados utilizados para treinamento poderia ampliar a aplicabilidade do ChatGPT para populações distintas e diferentes tipos de equipamentos OCT.⁽¹⁵⁾

Ademais, a integração de modelos de IA como o ChatGPT em sistemas de saúde também levanta questões éticas e legais, particularmente no que diz respeito à responsabilização por diagnósticos incorretos.⁽¹⁷⁾ Desse modo, ressalta-se a importância do uso correto dessas ferramentas, sendo que os laudos emitidos devem ser sempre avaliados e corrigidos por um especialista antes de qualquer conduta ser instaurada. Diretrizes claras para o uso dessas ferramentas devem ser estabelecidas para garantir sua implementação segura e eficaz.⁽¹⁹⁾

A implementação dessa ferramenta em larga escala requer estudos adicionais que explorem a reprodutibilidade dos resultados em diferentes cenários clínicos, além de comparações diretas com outros modelos especializados de IA.

CONCLUSÃO

Este estudo demonstra que o ChatGPT tem um desempenho promissor na análise de tomografias de coerência óptica, especialmente quando dados clínicos e epidemiológicos são integrados. Seu uso como ferramenta complementar pode contribuir para aumentar a eficiência diagnóstica, reduzindo a carga de trabalho dos especialistas e melhorando o acesso à saúde em regiões menos assistidas.

É fundamental que futuros estudos se concentrem na avaliação do impacto dessa tecnologia na prática clínica, investigando não apenas a precisão diagnóstica, mas também a aceitação pelos profissionais de saúde e pacientes, além dos desfechos clínicos associados ao seu uso.

Contribuição dos autores

Alice Gomes de Almeida: Concepção e delineamento do estudo, análise e interpretação de dados, redação e revisão final do manuscrito; Vitor Barbosa Cerqueira: Supervisão, validação dos resultados e revisão crítica do manuscrito.

REFERÊNCIAS

- Schlegl T, Waldstein SM, Bogunovic H, Endstraßer F, Sadeghipour A, Philip AM, et al. Fully automated detection and quantification of macular fluid in OCT using deep learning. *Ophthalmology*. 2018;125(4):549-58.
- Hormel TT, Hwang TS, Bailey ST, Wilson DJ, Huang D, Jia Y. Artificial intelligence in OCT angiography. *Prog Retin Eye Res*. 2021;85.
- Assi L, Chamseddine F, Ibrahim P, Sabbagh H, Rosman L, Congdon N, et al. A Global assessment of eye health and quality of life: A systematic review of systematic reviews. *JAMA Ophthalmol*. 2021;139(5):526-41.
- Ren LY. Isabel Pro use interface. *JCHLA/JABSC*. 2029; 40: 63-9.
- Bond WF, Schwartz LM, Weaver KR, Levick D, Giuliano M, Graber ML. Differential diagnosis generators: An evaluation of currently available computer programs. *J Gen Intern Med*. 2012;27(2):213-9.
- Brown T, Mann B, Ryder N, Subbiah M, Kaplan JD, Dhariwal P, et al. Language models are few-shot learners. *Adv Neural Inf Process Syst*. 2020;33:1877-901.
- Akpınar MH, Sengur A, Faust O, Tong L, Molinari F, Acharya UR. Artificial intelligence in retinal screening using OCT images: A review of the last decade (2013-2023). *Comput Methods Programs Biomed*. 2024;254:108253.
- Muntean GA, Marginean A, Groza A, Damian I, Roman SA, Hapca MC, et al. The predictive capabilities of artificial intelligence-based oct analysis for age-related macular degeneration progression – A systematic review. *Diagnostics*. 2023;13(14):2464.
- Balas M, Ing EB. Conversational AI models for ophthalmic diagnosis: Comparison of ChatGPT and the Isabel Pro Differential Diagnosis Generator. *JFO Open Ophthalmology*. 2023 Mar 1;1:100005.
- Madadi Y, Delsoz M, Lao PA, Fong JW, Hollingsworth T, Kahook MY, et al. ChatGPT assisting diagnosis of neuro-ophthalmology diseases based on case reports. *medRxiv*. 2023;2023.09.13.23295508.

11. Hashemian H, Peto T, Ambrósio R, Lengyel I, Kafieh R, Muhammed Noori A, et al. Application of artificial intelligence in ophthalmology: An updated comprehensive review. *J Ophthalmic Vis Res.* 2024;19(3):354-67.
12. Li Z, Wang L, Wu X, Jiang J, Qiang W, Xie H, et al. Artificial intelligence in ophthalmology: The path to the real-world clinic. *Cell Rep Med.* 2023;4(7).
13. Krishnan G, Singh S, Pathania M, Gosavi S, Abhishek S, Parchani A, et al. Artificial intelligence in clinical medicine: catalyzing a sustainable global healthcare paradigm. *Front Artif Intell.* 2023;6.
14. Popescu Patoni SI, Muşat AAM, Patoni C, Popescu MN, Munteanu M, Costache IB, et al. Artificial intelligence in ophthalmology. *Rom J Ophthalmol.* 2023;67(3):207-13.
15. Kras A, Celi LA, Miller JB. Accelerating ophthalmic artificial intelligence research: the role of an open access data repository. *Curr Opin Ophthalmol.* 2020;31(5):337-50.
16. Ting DS, Cheung CY, Lim G, Tan GS, Quang ND, Gan A, et al. Development and validation of a deep learning system for diabetic retinopathy and related eye diseases using retinal images from multiethnic populations with diabetes. *JAMA.* 2017;318(22):2211-23.
17. Esteve A, Robicquet A, Ramsundar B, Kuleshov V, DePristo M, Chou K, et al. A guide to deep learning in healthcare. *Nat Med [Internet].* 2019;25(1):24-9.
18. Soenksen LR, Ma Y, Zeng C, Boussieux L, Villalobos Carballo K, Na L, et al. Integrated multimodal artificial intelligence framework for healthcare applications. *NPJ Digit Med.* 2022;5(1).
19. Iqbal J, Cortés Jaimes DC, Makineni P, Subramani S, Hemaïda S, Thugu TR, et al. Reimagining healthcare: unleashing the power of artificial intelligence in medicine. *Cureus.* 2023;15(9).
20. Jiao C, Edupuganti NR, Patel PA, Bui T, Sheth V. Evaluating the artificial intelligence performance growth in ophthalmic knowledge. *Cureus.* 2023;15(9).
21. Perepelkina T, Fulton AB. Artificial Intelligence (AI) Applications for age-related macular degeneration (AMD) and other retinal dystrophies. *Semin Ophthalmol.* 2021;36(4):304-9.
22. Murugan S, Sanjay S, Somanath A, Mahendradas P, Patil A, Kaur K, et al. Artificial intelligence in uveitis: innovations in diagnosis and therapeutic strategies. *Clin Ophthalmol.* 2024;18:3753-66.