Eye Tracking e Aprendizado Profundo no Rastreamento do Transtorno do Espectro Autista: Uma Revisão da Literatura

Osmar Santana Ribeiro[1], José Valentim dos Santos Filho[2]

[1]Programa de Pós-Graduação em Engenharia Elétrica e Computação
Universidade Federal do Recôncavo da Bahia (UFRB) – Cruz das Almas – BA – Brasil

[2]Centro de Ciências Exatas e Tecnológicas (CETEC)
Universidade Federal do Recôncavo da Bahia (UFRB)

osmar.ribeiro@aluno.ufrb.edu.br, jose.valentim@ufrb.edu.br

Abstract. *Autism Spectrum Disorder (ASD) is characterized by persistent differences in social communication and interaction and by restricted or repetitive behaviors. Conventional screening may be limited by subjectivity, cost, and assessment time. This paper reviews studies published from 2022 to 2026 on Eye Tracking, Computer Vision, Machine Learning, and Deep Learning for ASD screening. The analyzed approaches include CNN, LSTM, Vision Transformers, YOLO, and MediaPipe. The literature reports promising results, but also high equipment costs, limited dataset diversity, and insufficient validation in real-world settings. The review supports the development of a low-cost Eye Tracking system for higher education institutions.*

Resumo. *O Transtorno do Espectro Autista (TEA) caracteriza-se por diferenças na comunicação e interação social e por comportamentos restritos ou repetitivos. Métodos convencionais de rastreamento podem ser limitados por subjetividade, custo e tempo de avaliação. Este artigo revisa estudos de 2022 a 2026 sobre Eye Tracking, Visão Computacional, Machine Learning e Aprendizado Profundo no rastreamento do TEA. As abordagens incluem CNN, LSTM, Vision Transformers, YOLO e MediaPipe. A literatura apresenta resultados promissores, mas também custos elevados, pouca diversidade de dados e validação insuficiente em ambientes reais. A revisão fundamenta um sistema de Eye Tracking de baixo custo para instituições de ensino superior.*

1 Introdução

O Transtorno do Espectro Autista (TEA) é um transtorno do neurodesenvolvimento caracterizado por déficits persistentes na comunicação e interação social, associados a padrões restritos e repetitivos de comportamento. A heterogeneidade clínica do transtorno torna o rastreamento e o diagnóstico processos complexos, reforçando a necessidade de métodos objetivos que complementem a avaliação clínica tradicional e favoreçam a

identificação precoce dos indivíduos em risco (Ahmed et al., 2022, 2023).

Nesse contexto, tecnologias baseadas em Machine Learning e Eye Tracking têm se destacado por possibilitarem a análise automática de biomarcadores comportamentais relacionados aos padrões de atenção visual. Os sistemas de Eye Tracking registram métricas como tempo de fixação, movimentos sacádicos, trajetórias do olhar (*scanpaths*) e mapas de calor (*heatmaps*), que, quando combinadas com técnicas de aprendizado profundo, permitem identificar padrões comportamentais associados ao TEA com maior objetividade (Ahmed et al., 2023; Al-Adhaileh et al., 2025).

Apesar dos avanços, a maioria das pesquisas utiliza equipamentos comerciais de alto custo, como Tobii Pro, EyeLink e Pupil Labs, além de bases de dados obtidas em ambientes laboratoriais, o que limita a aplicação dessas tecnologias em instituições públicas e cenários reais (Ahmed et al., 2022; Al-Adhaileh et al., 2025). No Brasil, embora existam estudos demonstrando o potencial do Eye Tracking em contextos educacionais e inclusivos (Cavalcante et al., 2020; Souza et al., 2025), ainda são escassas as pesquisas voltadas ao desenvolvimento de equipamentos nacionais integrados a técnicas de Machine Learning.

Diante desse cenário, este trabalho apresenta uma revisão bibliográfica sobre a aplicação de Eye Tracking, visão computacional e aprendizado profundo no apoio ao rastreamento do TEA. A análise principal abrange estudos publicados entre 2022 e 2026, complementados por trabalhos anteriores relevantes para a contextualização. Com base na literatura, propõe-se o desenvolvimento de um equipamento inteligente de Eye Tracking de baixo custo, utilizando câmera RGB, OpenCV, MediaPipe e modelos baseados em CNN e LSTM, visando ampliar o acesso a tecnologias assistivas e apoiar pesquisas sobre aplicações de Machine Learning à saúde e à educação.

2 Revisão Bibliográfica

2.1 Transtorno do Espectro Autista e biomarcadores digitais

O TEA caracteriza-se por alterações persistentes na comunicação social e pela presença de padrões restritos e repetitivos de comportamento, interesses e atividades, manifestando-se em diferentes níveis de suporte e comprometimento funcional. Sua elevada heterogeneidade clínica dificulta a identificação precoce, especialmente em indivíduos com sintomas mais sutis ou em fases iniciais do desenvolvimento. Dessa forma, pesquisadores têm buscado biomarcadores objetivos capazes de complementar os métodos clínicos tradicionais, reduzir a subjetividade da avaliação e contribuir para o rastreamento precoce do transtorno (Ahmed et al., 2022; Lord et al., 2020).

Entre os biomarcadores digitais investigados, os padrões de atenção visual obtidos por sistemas de Eye Tracking têm recebido destaque por permitirem quantificar aspectos relacionados à percepção social, à atenção compartilhada e ao processamento de estímulos visuais. Estudos demonstram que indivíduos com TEA apresentam diferenças no tempo de fixação, na frequência das sacadas, na exploração de rostos humanos e na observação da região dos olhos quando comparados à população neurotípica, indicando que essas características podem apoiar o rastreamento do transtorno (Ahmed et al., 2023; Jaradat et al., 2024).

2.2 Eye Tracking aplicado ao rastreamento do TEA

Eye Tracking é uma tecnologia destinada ao monitoramento contínuo da direção do olhar por meio da estimativa da posição da pupila e da córnea durante a observação de estímulos visuais. Os dados obtidos permitem calcular métricas como tempo de fixação, movimentos sacádicos, trajetórias do olhar, mapas de calor e áreas de interesse (*Areas of Interest* – AOIs), amplamente empregadas em pesquisas sobre cognição social e comportamento visual (Ahmed et al., 2022).

Al-Adhaileh et al. (2025) avaliaram modelos LSTM e CNN–LSTM em uma base pública de dados de Eye Tracking, empregando validação cruzada estratificada em cinco partes. O modelo híbrido alcançou acurácia média de 99,78%; entretanto, a ausência de validação externa limita a generalização do resultado para outras populações e contextos.

No contexto brasileiro, Cavalcante et al. (2020) demonstraram a viabilidade do Eye Tracking como estratégia de avaliação educacional inclusiva em estudantes com TEA. Os resultados evidenciaram que os padrões de atenção visual podem fornecer informações relevantes sobre a interação dos estudantes com ambientes digitais de aprendizagem, reforçando o potencial da tecnologia para aplicações além do contexto clínico.

2.3 Visão Computacional aplicada ao Eye Tracking

Os primeiros sistemas de Eye Tracking utilizavam câmeras infravermelhas e sensores ópticos específicos para determinar a direção do olhar com elevada precisão. Equipamentos como Tobii Pro, EyeLink e Pupil Labs continuam sendo referência em pesquisas experimentais devido à taxa de amostragem e à precisão espacial. Entretanto, seu custo constitui uma limitação para a adoção em instituições públicas de ensino e pesquisa (Tobii AB, 2025).

O avanço da Visão Computacional permitiu desenvolver soluções baseadas em câmeras RGB convencionais. Frameworks como MediaPipe Face Mesh e bibliotecas como OpenCV possibilitam detectar pontos faciais (*facial landmarks*), incluindo regiões da íris e da pupila, e estimar a direção do olhar em tempo real sem sensores infravermelhos dedicados (Google Research, 2024; OpenCV Team, 2024). Além da redução de custos, essas tecnologias apresentam vantagens de portabilidade, instalação e integração com sistemas embarcados.

2.4 Aprendizado Profundo para análise dos movimentos oculares

O Aprendizado Profundo representa uma das principais abordagens para análise automática de dados provenientes do Eye Tracking. Diferentemente dos métodos tradicionais, essas arquiteturas aprendem representações relevantes diretamente dos dados, reduzindo a necessidade de engenharia manual de atributos (Goodfellow et al., 2016).

As Redes Neurais Convolucionais (CNN) são empregadas para extrair características espaciais presentes nas imagens dos olhos, mapas de calor e regiões de interesse. As redes Long Short-Term Memory (LSTM), por sua vez, representam relações temporais entre fixações consecutivas e movimentos sacádicos, caracterizando a dinâmica do comportamento visual (Ahmed et al., 2023). Mais recentemente, Vision Transformers (ViT) passaram a ser investigados por sua capacidade de capturar relações globais entre

regiões da imagem. Algoritmos como YOLO podem localizar a face e os olhos, enquanto o MediaPipe fornece pontos faciais para estimar a direção do olhar.

2.5 Síntese da literatura e lacuna científica

A literatura demonstra avanços na utilização de Eye Tracking, visão computacional e aprendizado profundo no apoio ao rastreamento do TEA. Estudos recentes reportam bons resultados com arquiteturas CNN, LSTM e modelos híbridos (Ahmed et al., 2023; Al-Adhaileh et al., 2025). Uma meta-análise de 25 estudos, com 2.319 participantes, estimou acurácia agregada de 85%, além de apontar heterogeneidade metodológica e insuficiência de validação externa (Han et al., 2026). Também são escassos os estudos sobre dispositivos acessíveis que integrem hardware de baixo custo, visão computacional e aprendizado profundo para instituições públicas de ensino superior, especialmente no Brasil (Cavalcante et al., 2020; Jaradat et al., 2024).

Nesse contexto, propõe-se um equipamento inteligente de Eye Tracking baseado em câmera RGB convencional, integrado às bibliotecas OpenCV e MediaPipe para aquisição e processamento dos movimentos oculares, e associado a modelos CNN e LSTM para análise automática dos padrões de atenção visual. A proposta busca reduzir custos e ampliar o acesso a tecnologias de apoio ao rastreamento do TEA em ambientes acadêmicos e de pesquisa.

3 Trabalhos Relacionados

Nos últimos anos, a integração entre Machine Learning e Eye Tracking tem se consolidado como uma abordagem promissora para o apoio ao rastreamento do TEA. O avanço da Visão Computacional e do Aprendizado Profundo possibilitou modelos capazes de identificar padrões de atenção visual associados ao transtorno. Entretanto, a disponibilidade de bases de dados, o custo dos equipamentos e a limitada validação em ambientes reais ainda restringem a adoção dessas tecnologias.

de Belen et al. (2020) destacam o crescimento das pesquisas em Visão Computacional aplicada ao TEA, mas apontam baixa padronização metodológica e bases de dados pequenas. Elbattah et al. (2022) demonstraram que técnicas de transferência de aprendizado, com arquiteturas como VGG-16, ResNet e DenseNet, podem melhorar a classificação mesmo em conjuntos de dados reduzidos. No Brasil, Cavalcante et al. (2020) evidenciaram o potencial do Eye Tracking em ambientes educacionais inclusivos, porém sem desenvolver hardware dedicado ou integrar modelos avançados de Aprendizado Profundo.

A Tabela 1 sintetiza os estudos considerados e explicita a lacuna explorada pela proposta.

Tabela 1: Comparação dos trabalhos relacionados ao Eye Tracking aplicado ao TEA.

Estudo	Ano	Abordagem	Equipamento/base	Contribuição	Limitação
de Belen et al.	2020	Revisão sistemática	Diversas	Estado da arte em Visão Computacional para TEA	Poucas bases padronizadas
Ahmed et al.	2022	ANN, ResNet-18	547 imagens de trajetórias	Acurácia de até 99,8%	Base pública restrita
Elbattah et al.	2022	Transfer Learning	Imagens de *scanpaths*	ROC-AUC de até 0,78	Sem validação em ambiente real
Ahmed et al.	2023	LSTM, BiLSTM, CNN–LSTM	Base pública	Modelagem temporal	Base reduzida
Jaradat et al.	2024	MobileNet e *ensemble*	Duas bases públicas	Acurácia de até 98,1%	Sem validação externa
Al-Adhaileh et al.	2025	CNN–LSTM	Base pública; 59 participantes	Acurácia média de 99,78%	Sem validação externa
Han et al.	2026	Revisão e meta-análise	25 estudos; 2.319 participantes	Acurácia agregada de 85%	Heterogeneidade metodológica
Trabalho proposto	2026	MediaPipe, CNN e LSTM	Câmera RGB; UFRB	Plataforma integrada de baixo custo	Em desenvolvimento

Observa-se uma evolução de técnicas tradicionais de aprendizado de máquina para CNN, LSTM, transferência de aprendizado e modelos híbridos. Embora estudos individuais relatem resultados elevados, a síntese de evidências indica desempenho agregado mais moderado e grande variação entre protocolos. Bases públicas reduzidas e experimentos controlados podem favorecer o desempenho dos algoritmos sem assegurar sua generalização para cenários reais (Han et al., 2026).

Modelos híbridos combinam CNN para extração de características espaciais e LSTM para modelar a dinâmica temporal dos movimentos oculares. Essa combinação representa simultaneamente a posição do olhar e a sequência de fixações. A transferência de aprendizado também pode reduzir a necessidade de grandes bases específicas (Elbattah et al., 2022).

A análise comparativa evidencia a escassez de equipamentos inteligentes de baixo custo, baseados em câmeras RGB e frameworks abertos, capazes de operar em ambientes reais. A proposta deste trabalho integra aquisição, processamento e classificação dos movimentos oculares com MediaPipe, OpenCV, CNN e LSTM. A arquitetura conceitual é apresentada na Figura 1.

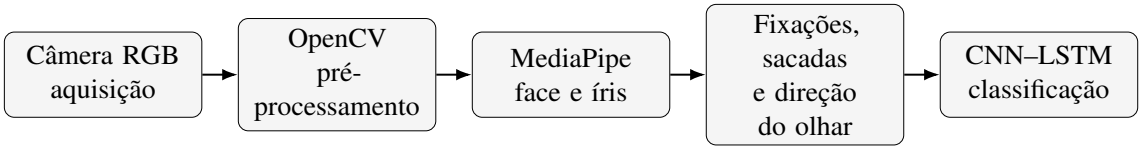


Figura 1: Arquitetura proposta para o equipamento inteligente de Eye Tracking.

Inicialmente, a câmera RGB realiza a aquisição contínua das imagens da face do participante. O OpenCV executa a conversão dos quadros, normalização da iluminação, redução de ruídos e preparação dos dados. Em seguida, o MediaPipe Face Mesh detecta marcos da face, da íris e da pupila. A partir dessas informações, são extraídas características relacionadas à posição e à direção do olhar, ao tempo de fixação e aos movimentos sacádicos. Por fim, modelos CNN–LSTM analisam os padrões visuais.

4 Discussão

A revisão evidencia que tecnologias de Eye Tracking associadas a Machine Learning têm avançado no apoio ao rastreamento do TEA. Padrões como tempo de fixação, movimentos sacádicos, trajetórias do olhar e mapas de calor constituem biomarcadores digitais potencialmente capazes de diferenciar indivíduos com TEA da população neurotípica (Ahmed et al., 2022, 2023).

Também se observa a evolução das arquiteturas usadas para analisar dados oculares. As pesquisas recentes empregam CNN, LSTM, modelos híbridos e transferência de aprendizado, que extraem características espaciais e temporais e reduzem a necessidade de engenharia manual de atributos (Elbattah et al., 2022; Al-Adhaileh et al., 2025).

Apesar desses avanços, muitos modelos foram validados somente com bases públicas ou experimentos em laboratórios controlados, o que dificulta avaliar sua robustez em cenários reais (de Belen et al., 2020; Han et al., 2026). Além disso, equipamentos comerciais de alta precisão podem impor barreiras de custo. A proposta apresentada diferencia-se por buscar um equipamento de baixo custo para apoio ao rastreamento do TEA em uma instituição pública de ensino superior, integrando câmera RGB, OpenCV, MediaPipe e modelos CNN–LSTM.

Além da contribuição tecnológica, a validação do equipamento em ambiente universitário permitirá investigar usabilidade, robustez e aplicabilidade em condições reais. Essa abordagem pode aproximar a pesquisa acadêmica das demandas da sociedade e fortalecer a integração entre Machine Learning, Engenharia Biomédica, Educação e Saúde Digital. Ressalta-se, contudo, que a solução se destina ao apoio ao rastreamento e não substitui a avaliação diagnóstica realizada por profissionais habilitados.

5 Conclusão

Este trabalho apresentou uma revisão bibliográfica sobre Eye Tracking, visão computacional e aprendizado profundo no apoio ao rastreamento do TEA. A análise principal dos estudos publicados entre 2022 e 2026, complementada por trabalhos anteriores de contextualização, indicou que movimentos oculares constituem biomarcadores digitais promissores e podem complementar métodos tradicionais de avaliação clínica.

Arquiteturas CNN, LSTM, transferência de aprendizado e modelos híbridos apresentam potencial para a análise automática dos movimentos oculares. Entretanto, predominam equipamentos comerciais de alto custo e protocolos em ambientes controlados, o que limita a adoção em instituições públicas. A comparação dos trabalhos evidenciou uma lacuna no desenvolvimento de plataformas acessíveis que integrem aquisição, processamento por Visão Computacional e classificação por Aprendizado Profundo.

Como trabalhos futuros, pretende-se desenvolver, calibrar e validar experimentalmente o protótipo, comparar arquiteturas CNN–LSTM, Vision Transformers e modelos multimodais, e construir uma base nacional de Eye Tracking com participantes com e sem TEA. Estudos com participantes deverão ser previamente submetidos ao sistema brasileiro de ética em pesquisa aplicável e conduzidos mediante consentimento informado.

6 Referências

Ahmed, I. A., Senan, E. M., Rassem, T. H., Ali, M. A., Shatnawi, H. S., Alwazer, S. M., and Alshahrani, M. (2022). Eye tracking-based diagnosis and early detection of autism spectrum disorder using machine learning and deep learning techniques. *Electronics*, 11(4):530.

Ahmed, Z. A. T., Albalawi, E., Aldhyani, T. H. H., Jadhav, M. E., Janrao, P., and Obeidat, M. R. M. (2023). Applying eye tracking with deep learning techniques for early-stage detection of autism spectrum disorders. *Data*, 8(11):168.

Al-Adhaileh, M. H., Alsubari, S. N. M., Al-Nefaie, A. H., Ahmad, S., and Alhamadi, A. A. (2025). Diagnosing autism spectrum disorder based on eye tracking technology using deep learning models. *Frontiers in Medicine*, 12:1690177.

Cavalcante, T. d. A., Soares, J. F., Paiva, A. C. d., Maia, I. M. O., Benitez, P., and Soares, A. (2020). Eye tracking como estratégia educacional inclusiva: Avaliação com estudantes com autismo. *Revista Brasileira de Informática na Educação*, 28:1181–1204.

de Belen, R. A. J., Bednarz, T., Sowmya, A., and Del Favero, D. (2020). Computer vision in autism spectrum disorder research: A systematic review of published studies from 2009 to 2019. *Translational Psychiatry*, 10(1):333.

Elbattah, M., Guérin, J.-L., Carette, R., Cilia, F., and Dequen, G. (2022). Vision-based approach for autism diagnosis using transfer learning and eye-tracking. In *Proceedings of the 15th International Conference on Biomedical Engineering Systems and Technologies (HEALTHINF)*, pages 256–263. SCITEPRESS.

Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep Learning*. MIT Press, Cambridge.

Google Research (2024). Mediapipe: Cross-platform, customizable ml solutions for live and streaming media. Disponível em: `https://mediapipe.dev`. Acesso em: 30 jun. 2026.

Han, W. et al. (2026). Diagnostic performance of eye-tracking-based artificial intelligence for autism spectrum disorder: A systematic review and meta-analysis. *International Journal of Medical Informatics*, 208:106235.

Jaradat, A. S., Wedyan, M., Alomari, S., and Barhoush, M. M. (2024). Using machine learning to diagnose autism based on eye tracking technology. *Diagnostics*, 15(1):66.

Lord, C., Brugha, T. S., Charman, T., et al. (2020). Autism spectrum disorder. *Nature Reviews Disease Primers*, 6(1).

OpenCV Team (2024). Opencv: Open source computer vision library. Disponível em: `https://opencv.org`. Acesso em: 30 jun. 2026.

Souza, A. C., Carmo, J. S., Benitez, P., and Silva, J. C. (2025). O que me diz seu olhar: Rastreamento ocular de crianças com e sem autismo em jogos educacionais digitais de matemática. *Revista Educação Especial*, 38(1):e20/1–21.

Tobii AB (2025). *Tobii Pro Fusion: User Manual*. Tobii Pro, Danderyd, Sweden. Versão 3.3.