Machine learning for polyp detection in colonoscopy: an integrative review

Stanrley Fontenelle Martins; Faculdade Matias Machline; fontenellemartins0@outlook.com;
Ronald Américo da Silva; Centro Universitário do Norte;
Jhonnatan Silva de Souza; PPGIBA-UFAM.

1. Introduction

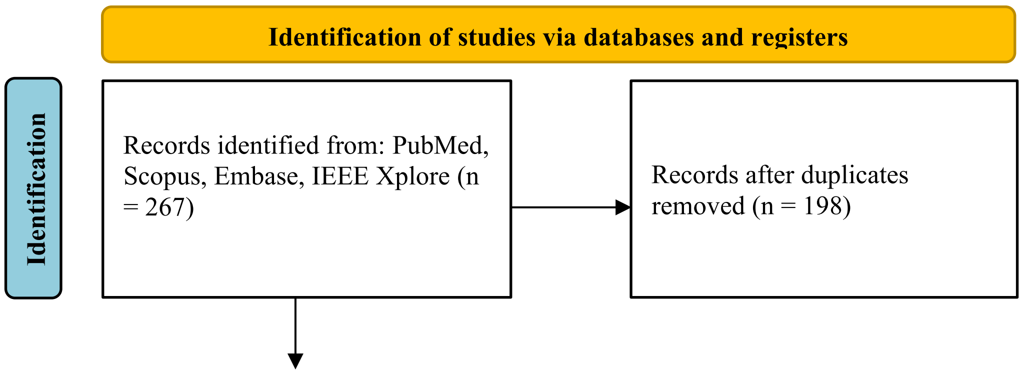
Colorectal cancer (CRC) represents a major global health burden, ranking among the leading causes of cancer-related morbidity and mortality. Since the adenoma–carcinoma sequence is well established, the prompt recognition and removal of colonic polyps are critical preventive strategies. Colonoscopy remains the gold standard technique for screening and intervention; however, studies indicate that up to 22% of adenomas may be missed during routine procedures, particularly flat or diminutive lesions that are challenging to identify[1].

Machine learning (ML), especially deep learning-based computer-aided detection (CADe) systems, has been increasingly applied to colonoscopy in order to improve adenoma detection rates (ADR) and reduce inter-operator variability. These approaches provide real-time assistance by flagging suspicious mucosal regions, thereby supporting the endoscopist in minimizing oversight. Although current evidence highlights substantial improvements in detection metrics, concerns regarding external validity, false-positive alerts, and workflow integration persist[2-4]. This integrative review synthesizes the most recent evidence (2018–2025) regarding ML-based systems for polyp detection in colonoscopy, emphasizing methodological rigor, clinical outcomes, and persisting research gaps.

2. Material and Methods

This integrative review was carried out according to the framework of Whittemore and Knafl, which allows the inclusion of empirical and theoretical studies to provide a comprehensive synthesis. Literature searches were performed in PubMed/MEDLINE, Scopus, Embase, and IEEE Xplore databases, covering publications from January 2018 to July 2025. The search strategy combined the descriptors "colonoscopy," "polyp detection," "machine learning," "deep learning," "artificial intelligence," and "CADe".

Eligibility criteria included original articles that employed ML algorithms explicitly for polyp detection in colonoscopy images or videos, reporting quantitative outcomes such as sensitivity, specificity, precision, recall, F1-score, ADR, or PDR. Studies focusing solely on classification, segmentation, or histopathological prediction were excluded. After independent screening and full-text analysis, data were extracted concerning algorithmic architecture, dataset characteristics, validation strategy, and performance outcomes (Figure 1). Given the heterogeneity of designs and metrics, a narrative synthesis was adopted.



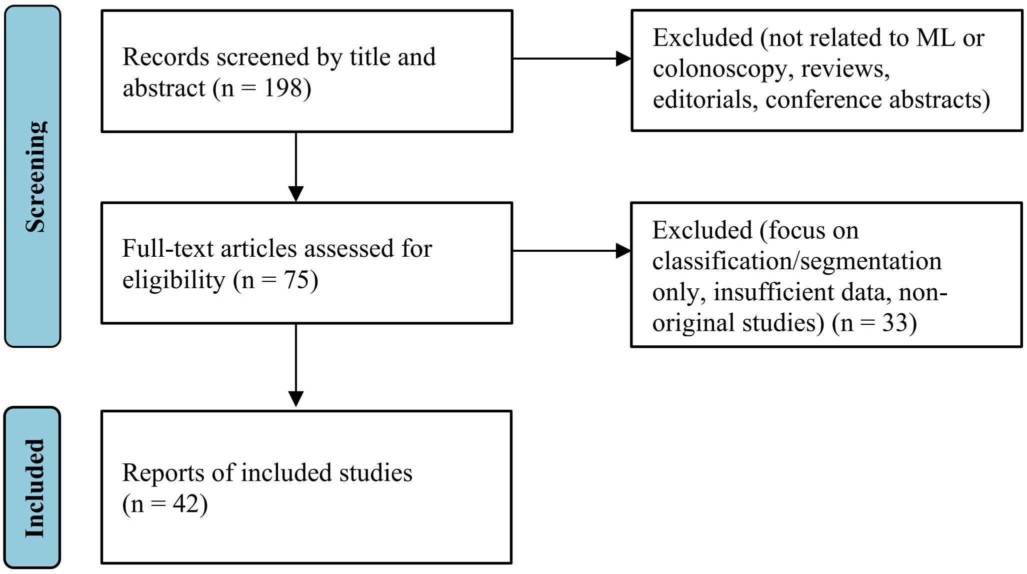


Figure 1. Flowchart showing the study selection process.

3. Results and Discussion

From 267 retrieved records, 42 studies fulfilled the inclusion criteria. Deep convolutional neural networks (CNNs) and derivative object detection architectures, including Faster R-CNN, RetinaNet, and the YOLO family, were predominant.

The performance of these models has been consistently high. Lee et al. (2020) reported that a YOLOv8-based system, trained on more than 8,000 colonoscopic images, achieved an accuracy of 93.4% and successfully identified polyps overlooked by experienced endoscopists (Figure 2) [5]. Krenzer et al. (2023) introduced ENDOMIND-Advanced, an open-source system validated with benchmark datasets, achieving an F1-score of 90.2% and recall of 100% in video sequences, underscoring its robustness for real-time application[6].

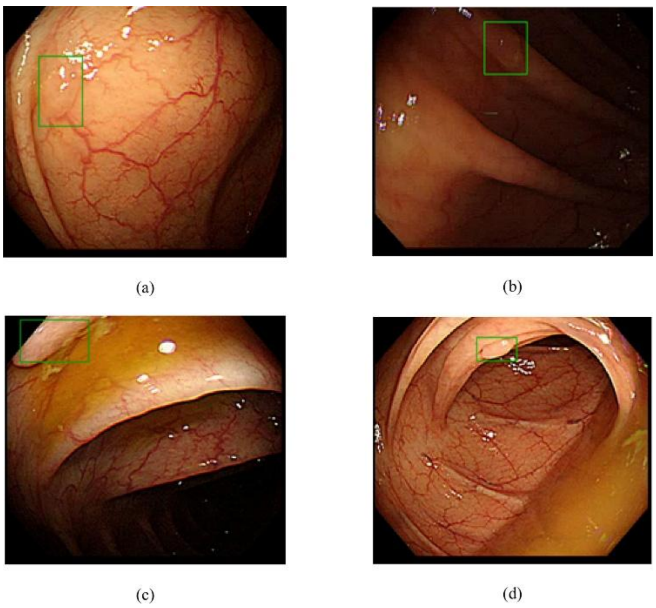


Figure 2. Examples of algorithmic polyp detection in video-image analysis. Green boxes indicate polyps detected by the algorithm. (a, b) Polyps detected under variable illumination conditions. (c) Partially visible polyp detected despite partial appearance. (d) Diminutive polyp detected under suboptimal bowel preparation. Reproduced from Lee et al.[5]

Evidence from clinical trials corroborates these findings. In a large multicenter randomized controlled trial including 805 patients, Park et al. (2024) demonstrated that AI-assisted colonoscopy significantly increased ADR and PDR compared to standard procedures (OR: 1.50)[7]. Similarly, Zhang et al. (2022) evaluated the DeFrame system on 681 full-length colonoscopy videos and reported sensitivity of 95.4% and precision of 92.1%, confirming feasibility in routine practice[8].

Despite these advances, several issues remain unresolved. Many models have been validated exclusively in single-center datasets, raising concerns about generalizability across diverse populations, endoscopic equipment, and imaging protocols. While high sensitivity and recall are desirable to minimize missed lesions, elevated false-positive rates may inadvertently increase procedure times and operator fatigue. Furthermore, recent observational studies have raised the possibility of "deskilling," whereby endoscopists become overly reliant on AI assistance, potentially impairing independent diagnostic performance[9].

It is noteworthy that current evidence highlights a trade-off between sensitivity and precision, with most systems prioritizing sensitivity to avoid missed lesions. Although this is clinically justified, strategies to mitigate false alarms, such as post-processing filters or hybrid human-AI decision frameworks, are being investigated. Additionally, the establishment of standardized public datasets and external validation benchmarks is urgently required to facilitate comparability and transparency in the field.

4. Conclusions

Integrative evidence suggests that ML-based CADe systems constitute a transformative adjunct to colonoscopy, consistently improving polyp detection rates across both retrospective datasets and prospective clinical trials. Cutting-edge architectures such as YOLOv5/YOLOv8 and RetinaNet exhibit high sensitivity and robust real-time applicability. Clinical validation indicates tangible benefits in increasing ADR and PDR, outcomes directly linked to reduced colorectal cancer incidence.

Nevertheless, translation into widespread clinical practice requires caution. Generalizability remains limited by dataset homogeneity, false-positive burden, and the risk of operator deskilling. Future research should prioritize multicenter randomized designs, standardized benchmarking, and strategies ensuring that AI complements rather than replaces clinical expertise. Only through rigorous validation and carefully planned implementation can ML-based systems be integrated as safe, reliable, and sustainable tools in colonoscopy.

Keywords: Colonoscopy; Polyp detection; Machine learning; Deep learning; Artificial intelligence.

5. References

1. van Rijn JC, Reitsma JB, Stoker J, Bossuyt PM, van Deventer SJ, Dekker E. Polyp Miss Rate Determined by Tandem Colonoscopy: A Systematic Review. Official journal of the American College of Gastroenterology | ACG [Internet]. fevereiro de 2006 [citado 31 de agosto de 2025];101(2):343. Disponível em: https://journals.lww.com/ajg/abstract/2006/02000/polyp_miss_rate_determined_by_tandem_colonoscopy_.25.aspx

2. Amin MA, Paul BK. Colon Polyps Detection from Colonoscopy Images Using Deep Learning [Internet]. arXiv; 2025 [citado 31 de agosto de 2025]. Disponível em: http://arxiv.org/abs/2508.13188

3. Barua I, Vinsard DG, Jodal HC, Løberg M, Kalager M, Holme Ø, et al. Artificial intelligence for polyp detection during colonoscopy: a systematic review and meta-analysis. Endoscopy. março de 2021;53(3):277–84.

4. Ali S, Ghatwary N, Jha D, Isik-Polat E, Polat G, Yang C, et al. Assessing generalisability of deep learning-based polyp detection and segmentation methods through a computer vision challenge. Sci Rep [Internet]. 23 de janeiro de 2024 [citado 31 de agosto de 2025];14(1):2032. Disponível em: https://www.nature.com/articles/s41598-024-52063-x

5. Lee JY, Jeong J, Song EM, Ha C, Lee HJ, Koo JE, et al. Real-time detection of colon polyps during colonoscopy using deep learning: systematic validation with four independent datasets. Sci Rep [Internet]. 20 de maio de 2020 [citado 31 de agosto de 2025];10(1):8379. Disponível em: https://www.nature.com/articles/s41598-020-65387-1

6. Krenzer A, Banck M, Makowski K, Hekalo A, Fitting D, Troya J, et al. A Real-Time Polyp-Detection System with Clinical Application in Colonoscopy Using Deep Convolutional Neural Networks. Journal of Imaging [Internet]. fevereiro de 2023 [citado 31 de agosto de 2025];9(2):26. Disponível em: https://www.mdpi.com/2313-433X/9/2/26

7. Park DK, Kim EJ, Im JP, Lim H, Lim YJ, Byeon JS, et al. A prospective multicenter randomized controlled trial on artificial intelligence assisted colonoscopy for enhanced polyp detection. Sci Rep. 26 de outubro de 2024;14(1):25453.

8. Chen S, Lu S, Tang Y, Wang D, Sun X, Yi J, et al. A Machine Learning-Based System for Real-Time Polyp Detection (DeFrame): A Retrospective Study. Front Med (Lausanne). 2022;9:852553.

9. Budzyń K, Romańczyk M, Kitala D, Kołodziej P, Bugajski M, Adami HO, et al. Endoscopist deskilling risk after exposure to artificial intelligence in colonoscopy: a multicentre, observational study. Lancet Gastroenterol Hepatol. 12 de agosto de 2025;S2468-1253(25)00133-5.