Where the Machines Fall Short: A Comparative Gap Analysis of AI Methods for Early Colorectal Cancer Screening

Uma Hombal[1], Shashi Kant Gupta[2]

[1] Post-Doctoral Researcher, LGPR Lincoln University; [2] Professor, LGPR Lincoln University

pdf.uma@lincoln.edu.my ; shashigupta@lincoln.edu.my

Abstract: Machine learning has been vigorously tackling the problem of early colorectal cancer (CRC) diagnosis, and through the previous four years, several methods with exceptional accuracy have been created. Nevertheless, when considered collectively, these methods have several blind spots that cannot be resolved by a single study. Seven recent works on multimodal fusion, deep learning based on spectroscopy, and blood-based liquid biopsy are examined in this review. For each, we describe the basic algorithm, its key drawback, and the specific gap it leaves unfilled. We observe a recurring pattern: high-accuracy methods almost never offer explanations that a physician can act upon, and they usually employ tissue instead of blood, limited, single-site datasets, or data types too costly for population screening. The most complete of the seven, the Colon Scope X framework is the closest to a deployable design, although it still has problems with demographic breadth, polyp recall, and explanation stability. Instead of another little gain in accuracy, we argue that the field needs a technique that is both non-invasive, inexpensive, demographically robust, and transparent. Instead of focusing on just one statistic, this analysis identifies a deficiency in that combination.

Keywords: Colorectal cancer; explainable AI; early screening; Raman spectroscopy; gap analysis; multimodal fusion; liquid biopsy.

Introduction

The second globally in terms of cancer-related deaths is the Colorectal cancer, and its deadliness is mainly dependent on the stage of its detection. A patient recognized at Stage I has an 80% to 95% five-year survival probability; by Stage IV, that percentage falls to around 10% [1]. Due to the disease's inactive progression and well-understood adenoma-to-carcinoma sequence, which can last up to ten years, there is an remarkably large window of chance for intervention. That window is wasted on screening uptake. The gold standard, colonoscopy, is invasive and resource-intensive, and stool-based diagnostics have poor compliance, particularly among younger and

less fortunate individuals [7]. Blood-based tests are specifically designed to address this compliance issue, according to Shaukat and Levin [11], who examine existing and new screening procedures. Artificial intelligence has been called upon to fill this gap between what is medically feasible and what actually reaches people. There are now two main approaches. In the first, a neural network trained to detect the resultant spectral fingerprint is paired with an inexpensive physical measurement, often Raman spectroscopy of blood or tissue [2, 3]. Because no one modality fully represents the heterogeneity of cancer, the second approach combines many data kinds simultaneously [5, 6]. Machine learning is used to liquid-biopsy data, such as circulating tumor DNA, in a third, more clinical field of study [7]. Each has generated astounding figures. None have created a screening tool that gastroenterologists use on a regular basis. We look at seven recent studies, determine the method each one uses, and pinpoint the flaw and ensuing research gap in each instance. The papers and the comparison are presented in tabular form in Section 2. In order to identify recurring patterns, Section 3 scans over the table. The final section of Section 4 explains why we view an optimal solution as one that simultaneously fulfills four criteria rather than maximizing just one of them, and why this combination represents the real gap in the existing literature.

Comparative Analysis of Existing Methods

The seven works reviewed here were selected because each represents a distinct methodological approach to CRC or closely related cancer detection, and because each has been peer-reviewed or archived with verifiable provenance. Table 1 summarises them along five axes: the paper, its authors, the algorithm it employs, the chief limitation of that algorithm, and the gap that limitation opens. The discussion that follows the table reads down its columns to draw out the shared weaknesses.

Table 1. Comparative gap analysis of seven AI methods for colorectal and related cancer screening

Paper (Ref.)	Authors	Algorithm / Method	Limitation / Drawback	Identified Gap
ColonScopeX [1]	Sikora et al. (2025)	Multimodal fusion of Raman spectra +	Single-site Welsh cohort (1,035 patients); late-fusion	Needs demographically diverse, multi-centre

Paper (Ref.)	Authors	Algorithm / Method	Limitation / Drawback	Identified Gap
		metadata (Random Forest + ANN); SHAP & LIME for explanation	polyp recall only 0.156; SHAP/LIME explanations vary across runs; no genomic or microbiome modality	data and stable explanations before the joint-fusion model can be trusted for population polyp screening
Serum Raman + 1D-CNN [2]	Du et al. (2023)	One-dimensional convolutional neural network on serum Raman spectra for gastric, colon, rectal and lung cancer	Small cohort (192 samples); cancer-vs-control only, no polyp / pre-cancer stage; no external validation; unimodal spectra with no clinical metadata	A pure-spectral CNN cannot separate pre-cancerous lesions from cancer and lacks the patient context needed for triage decisions
Raman + CNN colon tissue [3]	Wu et al. (2021)	CNN classifying three colon tissue types (adenomatous polyp, adenocarcinoma, normal) versus KNN, RF, SVM baselines	Requires resected tissue, not blood; small dataset; no interpretability; not a screening test as biopsy is already invasive	High accuracy on tissue does not transfer to non-invasive screening; a blood-based equivalent with explanation is missing
Exosome-SERS-AI [4]	Shin et al. (2023)	Deep-learning multiple-instance learning on SERS profiles of plasma exosomes; multi-cancer + tissue-of-origin	Needs exosome isolation (extra lab step, cost); retrospective design; black-box deep model with no clinician-facing explanation	Sample preparation burden and opaque predictions block routine primary-care use despite strong AUC (0.970)
Multimodal Raman CNN [5]	Wang et al. (2023)	Multimodal CNN combining serum SERS spectra with clinical features (age, PSA) for prostate cancer	Prostate not colorectal; risk of biopsy false negatives when linking Gleason score to metabolites; no XAI; needs SERS substrate	Demonstrates multimodal gain but leaves colorectal application and transparent reasoning unaddressed
MultiCoFusion [6]	Tan et al. (2022)	Multi-task fusion (ResNet-152 for pathology images	Depends on histopathology images and genomic	Powerful fusion, but its data and compute demands are

Paper (Ref.)	Authors	Algorithm / Method	Limitation / Drawback	Identified Gap
		+ sparse graph CNN for mRNA) for prognosis and grading	sequencing (costly, invasive); prognosis not early screening; heavy compute	unrealistic for cheap population-level screening
cfDNA blood test [7]	Chung et al. (2024)	Machine-learning classifier on cell-free DNA methylation and fragmentation signals (ECLIPSE / Shield)	Sensitivity for advanced pre-cancerous lesions only 13.2%; requires sequencing infrastructure; no per-patient explanation	Misses most pre-cancerous lesions, so it cannot prevent cancer; an early, interpretable, low-cost alternative is required

Reading Across the Table: Shared Weaknesses

Du et al. [2] report over 94% accuracy, Shin et al. [4] an AUC of 0.970, and ColonScopeX [1] an AUC of 0.975 for cancer detection. But the cohorts behind those numbers are small and geographically confined. Du and colleagues trained on 192 serum samples; ColonScopeX drew its entire 1,035-patient dataset from a single health board in South Wales [1]. A model that learns the metabolic baseline of one population encodes that baseline as normal, and Raman serum spectra shift with diet, gut flora, and metabolic health, all of which vary across ethnic and regional lines. d'Elia et al. [9] showed in a systematic review that socioeconomic and demographic homogeneity in training data is a recurring cause of poor generalisation in clinical AI. The accuracy, in other words, may not survive contact with a different population.

Tissue is not blood, and prognosis is not screening

Two of the strongest performers solve a different problem than screening. Wu et al. [3] classify colon tissue into polyp, adenocarcinoma, and normal with high accuracy, but the input is a resected specimen; if a biopsy has already been taken, the invasive step screening was meant to avoid has already happened. Tan et al. [6] fuse histopathology images with mRNA expression through a ResNet-152 and a sparse graph network, an elegant design, but it predicts prognosis and grade for patients who are already diagnosed, and it needs sequencing and a pathology slide to run.

Pre-cancerous lesions are where the methods break

The most substantial common mistake relates to polyps and progressive precancerous lesions, which are the exact objectives that validate initial screening. These lesions are important since the metabolic reprogramming that Hanahan [12] lists as one of the features of cancer starts slowly, long before a tumor develops, and a screening tool needs to detect this weak early signal. As per to Chung et al.'s [7] report on the massive ECLIPSE validation of a cfDNA blood test, the test's sensitivity for established colorectal cancer was 83.1%, while it was only 13.2% for advanced precancerous lesions. Even in its optimal setup, ColonScopeX's late-fusion model missed almost one-third of polyps, with a recall of just 0.156 [1]. The pre-cancerous stage was not modeled by the spectral-only CNNs [2, 3], which were trained to distinguish cancer from control. In a crucial way, a screening method that accurately identifies cancer but overlooks the lesions that precede it has come too late.

The models cannot explain themselves

Regulators increasingly treat medical AI as a high-risk category that must justify its outputs, yet transparency is the exception across this table. CNNs for serum and exosomes [2, 3, 4, 5] are "black boxes," classifying without providing an explanation. Only ColonScopeX [1] makes an effort to explain things to clinicians, and it does so by using SHAP and LIME, both of which have stability issues that have been documented: On repeated runs, LIME's sensitivity to the selected neighborhood and Kernel SHAP's stochastic sampling may produce distinct explanations [1]. In their evaluation of multimodal fusion for biomarker identification, Steyaert et al. [8] identified interpretability as the discipline's enduring weakness. A model demands a clinician to rely only on it without providing a solid explanation of its logic, which is not how triage choices are made.

Cost and workflow are rarely counted

Before analysis, Shin et al. [4] require an exosome-isolation phase; Wang et al. [5] and SERS-based methods require a specialized nanoparticle substrate; and Tan et al. [6] require extensive GPU computation and genome sequencing. The majority of primary-care settings lack the sequencing infrastructure that Chung et al. [7] rely on. Even the preprocessing of Raman spectra, prior to any model runs, benefits from high-performance computers, as Woods et al. [10] showed. These are

not footnotes for a test intended to reach all populations; rather, they establish the methods deployable.

Conclusion

When arranged side by side, the seven techniques create an eye-opening pattern. Each is weak in the other dimensions and powerful in one or two. Although the spectral CNNs are inexpensive and non-invasive, they are unaware of their own logic and the precancerous state. Richer signals are captured by the multimodal and liquid-biopsy systems, but they need expensive data, provide no explanation, and two of them completely ignore screening. The exosome approach has an excellent AUC, but it is obscured by a black-box model and an isolation phase. ColonScopeX, which combines non-invasive spectra, patient metadata, and an explanation layer, is the most balanced approach; yet, it is still limited by a single-site cohort, poor polyp memory, and the instability of the explainability tools that it relies on.

An optimal early-CRC screening method should be non-invasive, drawing on a routine blood sample rather than tissue or an added isolation step; low-cost and computationally light, so it can run in ordinary primary-care settings without sequencing or GPU clusters; demographically robust, validated across multiple sites and populations so its accuracy is not an artefact of one cohort; and transparent, offering explanations stable enough that a clinician can rely on them during triage. Every method reviewed here satisfies some of these and sacrifices others. None satisfies all four. We have chosen to frame the problem this way, rather than pursuing a further improvement to any single algorithm, because the evidence across these seven works points to diminishing returns on isolated metrics. Another two points of AUC on a single-site spectral model does not move the field closer to a deployable test. What would move it is a design that treats the four constraints as simultaneous requirements from the outset, most plausibly by building on the multimodal, explainable direction of ColonScopeX while deliberately correcting its cohort narrowness, its polyp recall, and its reliance on unstable post-hoc explanation. That is the gap this comparative analysis identifies, and it is the direction we regard as worth pursuing.

References

1. Sikora, N., Manschke, R.L., Tang, A.M., Dunstan, P., Harris, D.A. and Yang, S. (2025) ColonScopeX: Leveraging Explainable Expert Systems with Multimodal Data for Improved Early Diagnosis of Colorectal Cancer. arXiv preprint arXiv:2504.08824. Presented at the AAAI-25 Bridge Program.
2. Du, Y., Hu, L., Wu, G., Tang, Y., Cai, X. and Yin, L. (2023) Diagnoses in Multiple Types of Cancer Based on Serum Raman Spectroscopy Combined with a Convolutional Neural Network: Gastric Cancer, Colon Cancer, Rectal Cancer, Lung Cancer. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 298, 122743.
3. Wu, X., Li, S., Xu, Q., Yan, X., Fu, Q., Fu, X., Fang, X. and Zhang, Y. (2021) Rapid and Accurate Identification of Colon Cancer by Raman Spectroscopy Coupled with Convolutional Neural Networks. Japanese Journal of Applied Physics, 60(6), 067001.
4. Shin, H., Choi, B.H., Shim, O., Kim, J., Park, Y., Cho, S.K., Kim, H.K. and Choi, Y. (2023) Single Test-Based Diagnosis of Multiple Cancer Types Using Exosome-SERS-AI for Early Stage Cancers. Nature Communications, 14(1), 1644.
5. Wang, Y., Qian, H., Shao, X., Zhang, H., Liu, S., Pan, J. and Xue, W. (2023) Multimodal Convolutional Neural Networks Based on the Raman Spectra of Serum and Clinical Features for the Early Diagnosis of Prostate Cancer. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 293, 122426.
6. Tan, K., Huang, W., Liu, X., Hu, J. and Dong, S. (2022) A Multi-Modal Fusion Framework Based on Multi-Task Correlation Learning for Cancer Prognosis Prediction. Artificial Intelligence in Medicine, 126, 102260
7. Chung, D.C., Gray, D.M., Singh, H., Issaka, R.B., Raymond, V.M., Eagle, C., Hu, S., Chudova, D.I., Talasaz, A., Greenson, J.K., Sinicrope, F.A., Gupta, S. and Grady, W.M. (2024) A Cell-Free DNA Blood-Based Test for Colorectal Cancer Screening. New England Journal of Medicine, 390(11), pp. 973-983
8. Steyaert, S., Pizurica, M., Nagaraj, D., Khandelwal, P., Hernandez-Boussard, T., Gentles, A.J. and Gevaert, O. (2023) Multimodal Data Fusion for Cancer Biomarker Discovery with Deep Learning. Nature Machine Intelligence, 5(4), pp. 351-362.
9. D'Elia, A., Gabbay, M., Rodgers, S., Kierans, C., Jones, E., Durrani, I., Thomas, A. and Frith, L. (2022) Artificial Intelligence and Health Inequities in Primary Care: A Systematic Scoping Review and Framework. Family Medicine and Community Health, 10(Suppl 1), e001670.
10. Woods, F.E.R., Jenkins, C.A., Jenkins, R.A., Chandler, S., Harris, D.A. and Dunstan, P.R. (2022) Optimised Pre-Processing of Raman Spectra for Colorectal Cancer Detection Using High-Performance Computing. Applied Spectroscopy, 76(4), pp. 496-507.
11. Shaukat, A. and Levin, T.R. (2022) Current and Future Colorectal Cancer Screening Strategies. Nature Reviews Gastroenterology & Hepatology, 19(8), pp. 521-531.
12.] Hanahan, D. (2022) Hallmarks of Cancer: New Dimensions. Cancer Discovery, 12(1), pp. 31-46.