Explainable Machine Learning for Early Colorectal Cancer Detection: A Review of Recent Advances and Open Research Directions

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

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

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

Abstract: Colorectal cancer continues to be the leading causes of death globally. The techniques of screening pathways and invasive procedures are limited early-stage sensitivity. This review explores the recent advances in explainable artificial intelligence (XAI) and machine learning (ML) techniques applied to colorectal cancer (CRC) screening, with an emphasis on multimodal fusion frameworks, and federated learning. We analyze typical clinical trials that highlight the advantages and disadvantages of the main methods (XGBoost, LightGBM, SHAP, LIME, Grad-CAM, and ResNet-style deep models) and show their mathematical underpinnings. Comparative analysis across technique families reveals limitations, such as restricted cross-site validation, inconsistent treatment of class imbalance, and limited evaluation of computational sustainability. In order to facilitate early detection, we advocate for a sustainable, comprehensible, multimodal architecture that combines structured patient metadata with blood-based biomarker signals, such as serum Raman spectroscopy.

Keywords: Colorectal cancer screening; explainable artificial intelligence; XGBoost; SHAP; LIME; multimodal fusion; Raman spectroscopy; federated learning; deep learning

---

Introduction

Most commonly diagnosed cancers worldwide is the colorectal cancer and remains a major contributor to cancer mortality. GLOBOCAN estimates placed colorectal cancer third by incidence and second by mortality [1]. Survival is strongly stage-dependent, whereas survival in distant-metastatic disease falls below 15% [2]. Colonoscopy remains the reference standard owing to its high sensitivity and faces real-world adherence problems in many populations [3]. Stool-based assays are non-invasive and more acceptable to patients but vary in sensitivity [4]. The ECLIPSE study of a cell-free DNA test reported 83% sensitivity for colorectal cancer in an average-risk population but only 13% sensitivity [5]. Topol illustrates that deep learning approaches were reaching expert-level performance, while also flagging concerns about bias, privacy, and interpretability [6]. Convolutional neural networks have demonstrated the ability to detect polyps in colonoscopy video at clinically relevant sensitivity and specificity [7]. The opacity problem is somewhat addressed by the Explainable Artificial Intelligence (XAI) approach[8]. Local Interpretable Model-Agnostic Explanations (LIME) use a straightforward, locally faithful surrogate model to approximate the classifier's behavior in the vicinity of a particular case [9]. Gradient-weighted Class Activation Mapping (Grad-CAM) produces coarse spatial localization maps for image classifiers that show the areas that a convolution model has used to make predictions [10]. This paper makes the following

contributions: (i) a systematic description of the main ML and XAI approaches that are currently pertinent to early CRC detection; (ii) a concise mathematical formalization of the underlying algorithms; (iii) a comparative analysis of representative recent studies, emphasizing the insights each provides regarding the design space; and (iv) a set of research directions for a sustainable, comprehensible, multimodal CRC screening framework.

Related work

Shaukat and Levin examined the existing and new approaches to colorectal cancer screening and found that, despite strong evidence of efficacy, screening is still underutilized. Detailed risk-stratified guidelines, from ESGE specify the intervals at which patients with previously diagnosed polyps should return for additional evaluation [11]. The BLUE-C study assessed a multi-target stool DNA test and found that it was more sensitive than a commercial faucal immunochemical test [4]. Hanna et al. reviewed the application of Raman spectroscopy and noted its capacity to capture early biochemical changes[12]. Feng and co-workers demonstrated that SERS measurements on blood plasma could discriminate colorectal cancer patients from healthy controls, providing label-free spectral evidence of the disease state [13]. Sikora and colleagues introduced ColonScopeX, that integrates serum Raman spectra with patient metadata and accompanies its predictions with SHAP- and LIME-based explanations [14]. XGBoost, introduced by Chen and Guestrin, formulates tree boosting as a regularised additive expansion that allow it to scale to large datasets[15]. By introducing Gradient-based One-Side Sampling (GOSS), which compresses sparse feature spaces by combining mutually exclusive features, Ke and colleagues' LightGBM aims for comparable accuracy with significantly shorter training times [16]. Residual networks (ResNet), introduced by He and colleagues, address the degradation problem in very deep networks by learning residual mappings [17]. Transformers, introduced by Vaswani and colleagues are adapted to a broad range of non-textual tasks[18]. The attention operation provides a set of weights that can be interrogated for partial explanations of model behaviour.

Multimodal Fusion and Federated Learning

Steyaert and colleagues surveyed multimodal data fusion approaches and argued combining modalities is increasingly necessary if predictive models are to generalise across patients [19]. Rieke and colleagues set out the case for federated learning and discussed the residual risks needs to be addressed for clinical deployment [20].

EXPLAINABLE AI METHODS

SHapley Additive exPlanations (SHAP) and Local Interpretable Model-Agnostic Explanations (LIME)

SHAP, proposed by Lundberg and Lee, unifies a class of additive feature attribution methods within the framework of Shapley values from cooperative game theory [8]. LIME, by Ribeiro, Singh, and Guestrin, elobrates any classifier by fitting interpretable model to the predictions of the black-box model in a local neighborhood of the instance being explained [9].

Gradient-weighted Class Activation Mapping (Grad-CAM)

Grad-CAM, by Selvaraju etal. explained a method for CNN predictions by weighting the activations of the final convolution layer by the gradients of the target class score with respect to those activations [10].

LITERATURE SEARCH AND COMPARATIVE DISCUSSION

Table I summarises the principal categories of machine learning approaches with representative real-world studies for each category.

TABLE I: METHODOLOGICAL CATEGORIES AND REPRESENTATIVE STUDIES IN CRC-RELATED MACHINE LEARNING

Category	Data Modality / Setting	Representative Study	Reported Performance / Key Finding
Deep learning for endoscopic imaging	Colonoscopy still images and video	Wang et al. (2018) [7]	Per-image sensitivity 94.4%, specificity 95.9%, AUC 0.984; real-time inference
Cell-free DNA (blood-based test)	Plasma cfDNA, average-risk cohort (n = 7,861 analytical set)	Chung et al., ECLIPSE (2024) [5]	CRC sensitivity 83%, specificity 90% for advanced neoplasia, 13% sensitivity for advanced precancerous lesions
Multi-target stool DNA	Stool sample, average-risk cohort (n = 20,176)	Imperiale et al., BLUE-C (2024) [4]	Higher sensitivity than commercial FIT for both CRC and advanced precancerous lesions; less specificity
Surface-enhanced Raman spectroscopy (SERS)	Blood plasma (proof-of-concept, n = 46)	Feng et al. (2015) [13]	Discriminates CRC and precancerous lesions from healthy controls using label-free SERS signatures
Multimodal XAI framework	Serum Raman + patient metadata	Sikora et al., ColonScopeX (2025) [14]	Joint-fusion ML model with SHAP/LIME explanations for CRC and polyp detection
Multimodal data fusion (review)	Imaging, EHR, molecular data	Steyaert et al. (2023) [19]	Synthesises early / intermediate / late fusion strategies for cancer biomarker discovery
Federated learning for digital health	Multi-institutional medical data	Rieke et al. (2020) [20]	Framework for collaborative training without centralising raw patient data

DISCUSSION AND FUTURE RESEARCH DIRECTIONS

According to the empirical data, multimodal, interpretable, and sustainable architecture is recommended for early CRC screening. Blood-based modalities, cfDNA or serum Raman spectra are useful at population scale; gradient-boosted models with Tree SHAP attributions offer both performance and clinically inspectable explanations; federated training protocols allow cross-institutional learning without centralized data movement. Therefore, a sustainable, explainable, multimodal framework that integrates serum Raman spectra, structured clinical metadata, and a gradient-boosted classifier with SHAP-based explanations is a reasonable next research direction.

Conclusions

This study has compiled recent advancements in explainable AI and machine learning as they pertain to early colorectal cancer detection. The following are three conclusions. First, as each screening modality has a distinct position in the sensitivity, specificity, invasiveness, and cost space, a population-level

program is best served by integrating them. Second, the existing performance gap in non-invasive testing at the polyp stage. Third, concrete tools for converting model outputs into therapeutically usable narratives are available through SHAP, LIME, and Grad-CAM, and their combination with multimodal pipelines like Colon ScopeX. Future research should focus on comprehensive cross-institutional validation, measurement of computational sustainability, and clinical evaluation of the explanations themselves to guarantee that any deployed system is accurate, understandable, sustainable, and clinically reliable.

References

1. F. Bray, M. Laversanne, H. Sung, J. Ferlay, R. L. Siegel, I. Soerjomataram, and A. Jemal, "Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," CA: A Cancer Journal for Clinicians, vol. 74, no. 3, pp. 229–263, 2024. DOI: 10.3322/caac.21834.
2. E. Dekker, P. J. Tanis, J. L. A. Vleugels, P. M. Kasi, and M. B. Wallace, "Colorectal cancer," The Lancet, vol. 394, no. 10207, pp. 1467–1480, 2019. DOI: 10.1016/S0140-6736(19)32319-0.
3. A. Shaukat and T. R. Levin, "Current and future colorectal cancer screening strategies," Nature Reviews Gastroenterology & Hepatology, vol. 19, no. 8, pp. 521–531, 2022. DOI: 10.1038/s41575-022-00612-y.
4. T. F. Imperiale, K. Porter, J. Zella, Z. D. Gagrat, M. C. Olson, S. Statz, J. Garces, P. T. Lavin, H. Aguilar, D. Brinberg, C. Berkelhammer, J. B. Kisiel, and P. J. Limburg, "Next-generation multitarget stool DNA test for P. Wang, X. Xiao, J. R. Glissen Brown, T. M. Berzin, M. Tu, F. Xiong, X. Hu, P. Liu, Y. Song, D. Zhang, X. Yang, L. Li, J. He, X. Yi, J. Liu, and X. Liu, "Development and validation of a deep-learning algorithm for the detection of polyps during colonoscopy," Nature Biomedical Engineering, vol. 2, no. 10, pp. 741–748, 2018. DOI: 10.1038/s41551-018-0301-3.colorectal cancer screening," New England Journal of Medicine, vol. 390, no. 11, pp. 984–993, 2024. DOI: 10.1056/NEJMoa2310336.
5. D. C. Chung, D. M. Gray, H. Singh, R. B. Issaka, V. M. Raymond, C. Eagle, S. Hu, D. I. Chudova, A. Talasaz, J. K. Greenson, F. A. Sinicrope, S. Gupta, and W. M. Grady, "A cell-free DNA blood-based test for colorectal cancer screening," New England Journal of Medicine, vol. 390, no. 11, pp. 973–983, 2024. DOI: 10.1056/NEJMoa2304714.
6. E. J. Topol, "High-performance medicine: the convergence of human and artificial intelligence," Nature Medicine, vol. 25, no. 1, pp. 44–56, 2019. DOI: 10.1038/s41591-018-0300-7.
7. P. Wang, X. Xiao, J. R. Glissen Brown, T. M. Berzin, M. Tu, F. Xiong, X. Hu, P. Liu, Y. Song, D. Zhang, X. Yang, L. Li, J. He, X. Yi, J. Liu, and X. Liu, "Development and validation of a deep-learning algorithm for the detection of polyps during colonoscopy," Nature Biomedical Engineering, vol. 2, no. 10, pp. 741–748, 2018. DOI: 10.1038/s41551-018-0301-3.
8. S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in Advances in Neural Information Processing Systems 30 (NIPS 2017), Long Beach, CA, USA, Dec. 2017, pp. 4765–4774. arXiv:1705.07874.
9. M. T. Ribeiro, S. Singh, and C. Guestrin, "'Why should I trust you?': Explaining the predictions of any classifier," in Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16), San Francisco, CA, USA, Aug. 2016, pp. 1135–1144. DOI: 10.1145/2939672.2939778.

10. R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual explanations from deep networks via gradient-based localization," in Proceedings of the IEEE International Conference on Computer Vision (ICCV), Venice, Italy, Oct. 2017, pp. 618–626. DOI: 10.1109/ICCV.2017.74.
11. C. Hassan, G. Antonelli, J.-M. Dumonceau, J. Regula, M. Bretthauer, S. Chaussade, E. Dekker, M. Ferlitsch, A. Gimeno-Garcia, R. Jover, M. Kalager, M. Pellisé, C. Pox, L. Ricciardiello, M. Rutter, L. M. Helsingen, A. Bleijenberg, C. Senore, J. E. van Hooft, M. Dinis-Ribeiro, and E. Quintero, "Post-polypectomy colonoscopy surveillance: European Society of Gastrointestinal Endoscopy (ESGE) guideline — update 2020," Endoscopy, vol. 52, no. 8, pp. 687–700, 2020. DOI: 10.1055/a-1185-3109.
12. K. Hanna, E. Krzoska, A. M. Shaaban, D. Muirhead, R. Abu-Eid, and V. Speirs, "Raman spectroscopy: current applications in breast cancer diagnosis, challenges and future prospects," British Journal of Cancer, vol. 126, no. 8, pp. 1125–1139, 2022. DOI: 10.1038/s41416-021-01659-5.
13. S. Feng, W. Wang, I. T. Tai, G. Chen, R. Chen, and H. Zeng, "Label-free surface-enhanced Raman spectroscopy for detection of colorectal cancer and precursor lesions using blood plasma," Biomedical Optics Express, vol. 6, no. 9, pp. 3494–3502, 2015. DOI: 10.1364/BOE.6.003494.
14. N. Sikora, R. L. Manschke, A. M. Tang, P. Dunstan, D. A. Harris, and S. Yang, "ColonScopeX: Leveraging explainable expert systems with multimodal data for improved early diagnosis of colorectal cancer," Proceedings of the First AAAI Bridge Program on AI for Medicine and Healthcare, Proceedings of Machine Learning Research, vol. 281, pp. 144–154, 2025. arXiv:2504.08824.
15. T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16), San Francisco, CA, USA, Aug. 2016, pp. 785–794. DOI: 10.1145/2939672.2939785.
16. G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T.-Y. Liu, "LightGBM: A highly efficient gradient boosting decision tree," in Advances in Neural Information Processing Systems 30 (NIPS 2017), Long Beach, CA, USA, Dec. 2017, pp. 3146–3154.
17. K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, Jun. 2016, pp. 770–778. DOI: 10.1109/CVPR.2016.90.
18. A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in Advances in Neural Information Processing Systems 30 (NIPS 2017), Long Beach, CA, USA, Dec. 2017, pp. 5998–6008. arXiv:1706.03762.
19. S. Steyaert, M. Pizurica, D. Nagaraj, P. Khandelwal, T. Hernandez-Boussard, A. J. Gentles, and O. Gevaert, "Multimodal data fusion for cancer biomarker discovery with deep learning," Nature Machine Intelligence, vol. 5, no. 4, pp. 351–362, 2023. DOI: 10.1038/s42256-023-00633-5.
20. N. Rieke, J. Hancox, W. Li, F. Milletarì, H. R. Roth, S. Albarqouni, S. Bakas, M. N. Galtier, B. A. Landman, K. Maier-Hein, S. Ourselin, M. Sheller, R. M. Summers, A. Trask, D. Xu, M. Baust, and M. J. Cardoso, "The future of digital health with federated learning," npj Digital Medicine, vol. 3, no. 1, art. 119, 2020. DOI: 10.1038/s41746-020-00323-1.