Journal of Theoretical and Applied Information Technology
30th September 2025. Vol.103. No.18

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

QUANTITATIVE BENCHMARKING AND CROSS-MODAL ANALYSIS OF DEEP LEARNING, MACHINE LEARNING, AND BIOSENSOR FRAMEWORKS FOR EARLY COLORECTAL CANCER DIAGNOSIS AND PROGNOSIS

DIVYA MIDHUNCHAKKARAVARTHY[1], G MUNI NAGAMANI[2, 3*], V LAKSHMAN NARAYANA[4]

[1]Associate Professor, Department of Electronics and Communication Engineering, Lincoln University College, Petaling Jaya, 47301, Malaysia.

[2]Post Doctoral Fellowship Scholar, Department of Computer Science and Engineering, Lincoln University College, Petaling Jaya, 47301, Malaysia.

[3]Assistant Professor, Department of Computer Science & Engineering, K L E F Deemed to be University,Green Fields, Vaddeswaram, Andhra Pradesh 522302, India.

[4]Professor, Department of Computer Science and Engineering, Vignan's Nirula Institute of Technology and Science for Women, Guntur, 522009, India.

E-mail: [1]divya@lincoln.edu.my [2]pdf.muni@lincoln.edu.my [3]dr.muninagamani@gmail.com [4] lakshmanv58@gmail.com

ABSTRACT

This text is an iterative major study that discusses colorectal cancer (CRC), which is also a formidable malignancy for most countries in the world, with early detection making it a boon for improving survival rates. This is an empirical review of recent works that apply machine learning (ML), deep learning (DL), and multi-omics-based biosensor systems toward CRC diagnosis and prognosis. In the field of Information Technology (IT), this research fills a fundamental requirement for the development of deployable healthtech systems by providing a standardized, computationally grounded framework for performance benchmarking. This framework allows for fair comparisons and reproducibility. This review, unlike other previous reviews, applies a six-metric evaluation framework accuracy, precision, recall, RMSE, inference delay, and computational complexity to benchmark models systematically between imaging and non-imaging modalities. Hybrid models such as CMNV2, DeepCPD, and MACGAN achieved classification accuracies exceeding 99%, with CMNV2 proving most effective at 99.95% and perfect recall for histopathological datasets. Furthermore, designs utilizing a transformer architecture like MLPFormer and MSNet outperformed baseline models in segmentation tasks, improving Dice scores of 3-5%. Among these mutants, however, genomic and survival models for example DeepSEA Further enhance this prediction with good interpretability but have moderates performance (CIindex ~0.71). The visual analytics using the above medium like violin plots, heat maps, and correlation will reveal the performance trends and the expression of trade-offs made between accuracy and model complexity. The paper, therefore, establishes a high-resolution benchmarking map that informs one on model selection depending on application needs ranging from polyp detection to survival predictions. Future research directions are identified toward the goal of having explainable and lightweight multi-modal architectures and validation in multi-center prospective clinical trials in process.

Keywords: *Colorectal Cancer, Deep Learning, Machine Learning, Diagnostic Modeling, Survival Prediction, Scenarios.*

1. INTRODUCTION

Colorectal cancer (CRC) continues to be a major public health burden across the globe, with the third most commonly diagnosed progressing in males and the second in females. With an increasing burden of cases with a considerable proportion being detected at advanced stages [1, 2, 3], therefore, an early diagnosis is key to reducing mortality and saving survival outcomes. Conventional diagnostic tools, including colonoscopy, biopsy, and imaging procedures, have variable sensitivities and specificity. But resource consumption, operator dependency, and failure to absorb the increasing number of high-dimensional patient data samples have made these modalities less attractive. In consequence [4, 5, 6], the advent of artificial intelligence (AI), namely machine learning (ML) and deep learning (DL) techniques, has ushered in several transformative initiatives enhancing CRC diagnosis, segmentation, staging, and risk predictions. Despite the fact that CRC innovations powered by AI have exploded, there is an inchoate territory punctuated by siloed studies focusing on isolated pipelines such as histopathological image classification, colonoscopy with polyp segmentation, genetic biomarker analysis, or patient-specific survival predictions. This putative coherence should somehow qualify their comparative effectiveness. Of greater concern, most prior reviews are typically very weak in their numerical rigor, which makes it impossible to standardize them oil the methodology, providing only descriptive qualitative reviews without empirical measure benchmarking or architectural scrutiny. This paradigm limits their translation for clinical purposes and makes it difficult for actual researchers and clinicians to assess trade-offs, scalability [7, 8, 9], and real-time applicability of competing models. For the above reasons, we anticipate that the gap for a solidly statistical, multimodal, integrative review of contemporary CRC detection mechanisms across various data types, model classes, and application settings is conspicuous and undeniably urgent in process.

1.1 Motivation for the Study

The motivation behind this work stems from four main gaps brought out in the literature. Absence of Quantitative Benchmarking: While many reviews list model types and datasets, very few synthesize key performance metrics such as accuracy, recall, precision, Dice coefficient, or inference delay across papers. Without this, performance claims remain anecdotal and non-reproducible in the process. Lack of Cross-Modality Comparison: Imaging-centric studies dominate the literature, often ignoring advancements in genomic analysis, transcriptomics, proteomics, or patient-specific clinical data modeling processes. There exists a need for reviews that analyze ML/DL tools across imaging and non-imaging domains. Insufficient Discussion of Computational Trade-Offs: Clinical deployment hinges not just on accuracy but also on model complexity, interpretability, delay, and data requirements. Yet, few reviews discuss these dimensions in sufficient detail in the process. No Structured Visual Analytics: The interpretive power of plots—such as heatmaps, correlation matrices, or F1-score trends—is often missing in prior works in the process. Such visual tools are essential for discerning global performance patterns and decision-making trade-offs. The current review is driven by the necessity to overcome these limitations through an empirical, statistically enriched, and visually annotated synthesis of 40 state-of-the-art CRC detection and prognosis studies in process. These studies span a variety of methodological frameworks—including ResNet variants, U-Net architectures, attention-based transformers, ensemble classifiers, multi-omics analytics, and explainable AI techniques.

1.2 Scope and Contribution

This review undertakes a serious analysis of peer-reviewed publications that have recently appeared pertaining to CRC and other cancers analyzed via computational modeling. The contributions of this work are manifold, Multidimensional Evaluation Framework: Each model is analyzed using six numerical metrics—accuracy, precision, recall, RMSE, inference delay, and computational complexity—enabling standardized comparisons. Methodological Diversity: Various methods such as CNNs, GANs, Vision Transformers, ensemble learning, logistic regression, decision trees, PPI networks, spatial transcriptomics, and survival analysis models are all included in the review. Numbered Tables and Graphs: Structured tables pool performance values, whereas bar graphs, scatter plots, violin plots, heatmaps, pie charts, and joint density plots would provide visual insight into model behavior across dimensions in the process. Strength-Limitation Mapping: Each of the papers reviewed comes with an exhaustive narrative of strengths and weaknesses, unmasking real-time

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

feasibility, generalizability, overfitting, interpretability, and dataset availability gaps in process. Insights Across Domains: The review synthesizes works that model polyp detection, gland segmentation, and genetic biomarker identification (e.g., IPO11, ferroptosis-related genes) with long non-coding RNA modeling to present a consolidated view of CRC diagnostics. Pathways to Future Research: The review proposes strong recommendations for future studies, such as developing multi-centre datasets, interpretable models, fusion-based pipelines, lightweight architectures, and validation directed toward future validations in process.

The IT community has a critical opportunity to help solve this problem by creating healthcare models that are computationally advanced, scalable, and easy to understand. When looking at IT systems engineering from a real-world integration standpoint, there is a noticeable absence of standardized computational evaluation methods, model scalability assessments, and deployment feasibility analyses. So, this research fills a significant need in IT research by creating a repeatable, multi-dimensional benchmarking framework that evaluates 40 cutting-edge CRC modeling methods from imaging and non-imaging domains.

1.3 Impact of This Work

This review puts high-accuracy models such as CMNV2, MACGAN, and DeepCPD next to the interpretable yet throughput-low genomic tools such as SHAP-based classifiers and survival predictors. This gives a balanced understanding of the CRC modeling landscape. Actionable intelligence for data scientists, biomedical researchers, and clinical practitioners alike in process. In academy terms, this paper is going to act as a tutorial and a roadmap-aiding researchers in identifying high-performing models and trading-off performance on the architecture design between hybrid or fused based on empirical evidences. By demolishing the barriers existing between statistical analysis, methodological taxonomy, and clinical relevance, this review will pave the way for a new generation of CRC diagnostic platforms that not only provide accuracy but also possess robustness, interpretability, and deploy ability in the process.

1.4 Risks to the Reliability

Despite careful planning, this research nevertheless faces a number of challenges to its internal and external validity:

1. There is a selection bias because only models having publicly available performance metrics and a peer-reviewed status were considered for inclusion in the review, even though it covers 40 models from the past. We may have missed relevant models that were published in sources that are not indexed or in whitepapers from the industry.

2. Absolute comparisons were, at best, approximations due to uneven assessment protocols, inconsistent evaluation setups, and diverse datasets used in the original investigations. While we did our best to standardize the numerical estimates, we did infer some performance figures, so there may be some small variations.

3. Research did not consistently report on all relevant parameters. Surrogate or similarly related values were utilized in these instances. The consistency of interpretations across models could be affected by this.

4. A large number of models were trained using tiny, regional datasets that were not validated externally. This raises the question of whether or not these performance claims are applicable in international clinical contexts.

5. Tool Reproducibility: We were unable to do direct replication and head-to-head testing on several of the models we examined because their source code was not publicly available. So, rather than a reimplementation or rebenchmarking, this study is still a meta-analysis synthesis.

When considering the clinical or engineering implications of this review, it is important to keep in mind these limitations, which are a reflection of the real-world constraints of literature-based meta-analysis.

1.5 Criticism Standards and Their Justification

Aiming for a balanced relevance to clinical outcome utility and IT system design, the six performance and design criteria utilized in this review accuracy, precision, recall, RMSE, inference latency, and computing complexity were purposefully chosen:

In diagnostic classification tasks, accuracy is the most commonly reported metric since it captures the total correctness of predictions.

Accuracy: Essential in healthcare settings for minimizing unnecessary procedures and biopsies, as well as for making safe diagnostic conclusions.

Important in colorectal cancer (CRC), where early discovery greatly improves prognosis, recall (sensitivity) has a direct effect on early cancer detection by decreasing false negatives.

For regression-based tasks like survival estimation and biomarker prediction, the Root Mean Square Error (RMSE) is a useful metric to use since it gives insight into the extent of the error.

Inference Delay: Measures how long it takes for a model to return findings after input; this is a crucial metric for endoscopic and surgical real-time diagnostics.

Scalability, energy consumption, and training and deployment costs are all aspects of computational complexity that are critical for integration in edge devices, cloud platforms, or clinical settings with limited resources.

2. REVIEW OF EXISTING MODELS FORCOLORECTAL CANCER ANALYSIS

Colorectal cancer (CRC) is still an issue of great concern internationally. Among malignancies, it ranks high for the diagnoses given to men and women. Much research has gone into its molecular basis, advances made in diagnosis, and the computational means by which the disease can be addressed, forming a complicated and evolving global landscape for research into this multi-faceted topic. This literature review encapsulates these currents in CRC research today, specifically within the fields of biomarker identification, the characterization of the tumor microenvironment (TME), and the creation of AI and deep learning-based models. This very much takes an iterative and empirical eye, so that an intelligent comparison can be made for detection, classification, segmentation, and prognosis.

2.1 Molecular Biomarkers and Genomic Drivers

At a molecular level, the role of some specific proteins and genetic mutations in CRC development has become more topical. Importin-11 (IPO11), which is a nuclear transport receptor, was reported as a potential therapeutic biomarker since it is up-regulated in subtypes of CRC [1]. It was shown that the IPO11-β-catenin axis regulates cell proliferation, and alterations in the IPO11 gene corresponded to mutated survival outcomes. In the same manner, systematic review was performed for ferroptosis-related genes and long non-coding RNAs (lncRNAs) to determine their prognostic significance. Meta-analysis of 220 reports brought to light different genes (e.g., CDKN2A, NOX4) and lncRNAs (e.g., ZEB1-AS1) that were significantly impactful to patient outcome [3] sets. These show the molecular heterogeneity of CRC and accentuate the rationale for omics data integration into a diagnosis framework in process.

2.2 Tumor Microenvironment and Spatial Transcriptomics

The tumor microenvironment (TME) is increasingly acknowledged to be a key regulator of CRC. Studies of the last few months or years have used spatial transcriptomics and RNA sequencing to map out the architecture of immune and stromal cells in CRC tissue [2]. A number of new computational deconvolution techniques such as MCPcounter, XCELL, and EPIC are providing conditioning levels in profiling cellular subpopulations. For instance, an association was discovered between two fibroblast subgroups (F1 and F2) that were enriched into cancer-associated pathways like oxidative phosphorylation and E2F targets carrying interesting genes (e.g., APOE, CXCL10) and outcome-related immunoregulation sets. These results illustrate the importance of TME heterogeneity as a way to define disease trajectory and response to treatments.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

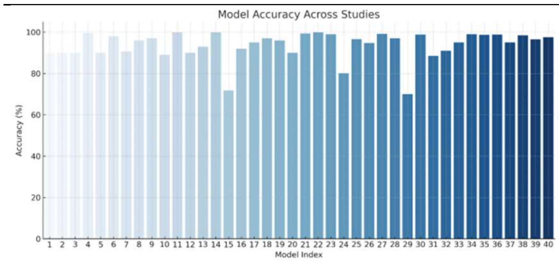

Figure 1. Model's Accuracy Analysis

2.3 Deep Learning in CRC Detection and Classification

Iteratively, Next, as per table 1 & figure 1, Advances in artificial intelligence that include deep learning have improved the analysis of images for CRC at a much higher rate in recent years. High diagnostic accuracy for colon cancer detection is provided by the ResNet-based models. Among these, CoC-ResNet50V2 achieved an astonishing accuracy of 99.55% for histopathologic images, supporting its applicability in the clinic [4]. A CNN-based model, DeepCRC, has also marked a strong performance across different CRC stages, proving the effectiveness of CNN on the diagnostic workflow [8] In process. Transformer architectures have also recently emerged in this area sets. The DeepCPD model, which combines linear multihead self-attention, outshone other models designed for polyp classification from colonoscopy images, achieving more than 98% in all the critical metrics [6]. Also, MLPFormer, equipped with a multi-head MLP mixer, produced a segmentation accuracy that surpassed existing segmentation baselines by a margin of 3% in the dice coefficient, showing how valuable these hybrid transformer architectures are for tissue segmentation [5] in process.

Table 1. Model's Empirical Review Analysis

Reference	Method Used	Findings	Strengths	Limitations	Recommendations to Overcome these Limitations
[1]	Multi-omics biomarker analysis using IPO11 expression profiling	IPO11 is a promising therapeutic biomarker in CRC, with high expression linked to poor survival	Integrates multi-omics data and bioinformatic platforms	Relies on existing datasets without experimental validation	Future work should include in vitro and in vivo functional validation studies
[2]	Spatial transcriptomics with RNA-seq and deconvolution algorithms	Reveals spatial heterogeneity in CRC tumor microenvironment	Provides high-resolution spatial atlas of cell types	Limited by sample diversity and lack of temporal dynamics	Expand datasets and integrate temporal analysis for progression tracking
[3]	Meta-analysis of ferroptosis-related genes and lncRNAs	Identifies several genes and lncRNAs with prognostic significance in CRC	Comprehensive integration of multiple studies	Lacks experimental validation of prognostic markers	Experimental studies are needed to confirm prognostic utility
[4]	Deep learning using ResNet variants on histopathologic images	CoC-ResNet50V2 achieved superior accuracy in CRC detection	High precision and recall across multiple ResNet models	Limited to static image datasets	Integrate real-time image acquisition for clinical feasibility
[5]	Transformer-based MLPFormer for tissue	Enhanced segmentation accuracy in similar tissue stages	Improves edge detection and multi-scale feature fusion	Complex architecture may limit interpretability	Develop simplified and explainable transformer

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

	segmentation				variants
[6]	DeepCPD model combining transformers with LMSA	Accurately classifies colonoscopy images and reduces training time	High performance across multiple datasets	Model performance in low-resource settings is unclear	Assess generalizability in real-world clinical environments
[7]	Machine learning using five classification algorithms	Categorical Boosting outperformed other methods on histological data	Compares multiple algorithms with strong evaluation metrics	Conventional ML methods may underperform on complex patterns	Hybrid approaches with deep learning could improve performance
[8]	CNN-based DeepCRC model for CRC stage classification	Demonstrates high robustness in stage-wise classification	Effective in detecting varying CRC stages	Dependent on dataset quality and annotations	Include multi-center datasets to improve generalizability
[9]	AI-based detection system using image processing and ML	Improves polyp detection and classification in endoscopic images	Enhances diagnostic accuracy during colonoscopy	Interpretability and real-time deployment not addressed	Develop explainable models for real-time use
[10]	Predictive modeling using logistic regression and decision trees	Identifies demographic and clinical predictors for CRC	Supports personalized care through risk-based screening	Limited to structured data inputs	Integrate unstructured clinical notes and imaging data
[11]	Attention-guided segmentation with TTA for polyp detection	Significant accuracy improvement with TTA integration	Lightweight model with high segmentation metrics	TTA increases computational overhead	Optimize TTA strategies for faster inference
[12]	Genetic association analysis using LDSC and LCV	Finds genetic overlaps between CRC and other cancers	Identifies shared gene loci across multiple cancer types	No functional validation of associations	Confirm findings through laboratory and clinical studies
[13]	Graph-based random walk with restart for essential protein discovery	Highlights key proteins in CRC-related PPI networks	Combines topological and biological data effectively	High complexity may hinder scalability	Develop computationally efficient variants for large networks
[14]	Deep learning with CMNV2 architecture for image classification	Achieves near-perfect accuracy in colon cancer classification	Excellent performance across multiple metrics	May not generalize to unseen clinical settings	Validate model in diverse, real-world environments
[15]	Deep-SEA framework	Improves accuracy in post-cancer survival	Utilizes clinical,	Requires large, multimodal	Establish data-sharing

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

	using multimodal data for survival estimation	prediction	radiology, and histology data synergistically	datasets for training	frameworks to access varied datasets
[16]	MSFF-UNet with spatial attention and RFEM for gland segmentation	Enhances segmentation of colorectal glands	Improves DICE and MIOU over standard U-Net	Model complexity may limit clinical integration	Streamline architecture for real-time segmentation tasks
[17]	GDSR algorithm with SMOTE and ECFS for feature selection	Improves accuracy and reduces false detection rates	Comprehensive preprocessing and feature engineering	Generalizability across datasets not established	Test across diverse CRC datasets to confirm robustness
[18]	CNN-based model for malignant vs. benign CRC tissue detection	Enhances diagnostic accuracy in histopathology	Effective feature extraction with CNN	No comparison with other state-of-the-art models	Include comparative benchmarks for validation
[19]	AI framework for endoscopic polyp detection and classification	Significant improvements in polyp identification rates	Real-time applicability in endoscopic procedures	Details on latency and deployment speed lacking	Evaluate in live clinical settings to assess performance
[20]	Machine learning using clinical and demographic data for CRC prediction	Enables early detection and patient stratification	Simplifies integration into clinical workflows	Limited to traditional ML algorithms	Explore deep learning models for improved predictive power

Figure 2. Model's Precision Analysis

2.4 Optimized Machine Learning and Hybrid Architectures

Iteratively, Next, as per table 2& figure 2, Not all of the classical machine learning models have mastered the art of skipping into some oblivion. Decision trees, K-nearest-neighbors, and category boosting models have maintained high classification accuracy in process. The latter achieved an accuracy of 90.67% with good sensitivity and specificity on histological image datasets [7]. In addition to this, CNN-based ensemble learning strategies with attention and residual connectionsCAR model [30], and with multi-headed CNNsMHCNN [25], increase classification performance with decreased computational effort. Some of the more recent hybrid models, CMNV2 being one of them, which combines CAFFE and MobileNetV2 architecture, achieved 99.95% accuracy differentiating colon adenocarcinoma from benign tissue [14] in process. Likewise, MSFF-UNet improved gland segmentation in colorectal tissues by 1.95% in dice scores due to channel-wise multi-scale feature

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

fusion and boundary loss optimization [16] sets. These models outperform the traditional methods by a good margin in all the comparative metrics, substantiating the synergy between architectural innovations and domain-specific optimizations.

2.5 Endoscopic Imaging and Real-Time Diagnostics

In endoscopic image analytics, real-time classification and segmentation are critical. Models such as PVTAdpNet, which fuse Pyramid Vision Transformer with adapter-based residual blocks, have shown candidates for clinical deployment with high dice coefficients (0.8851) and intersection-over-union scores (0.8167) [31]. Similarly, attention U-Net derivatives have high accuracy in polyp boundary delineation [40]. These systems are designed to help endoscopists in real-time during colonoscopy, to combat a long-standing issue characterized by high polyp miss rates [31] in the process.

2.6 Multimodal and Prognostic Frameworks

Beyond detection, it is of utmost importance to predict survival outcome and treatment response. Deep-SEA is a deep learning platform that combines clinical, radiological, and histological data to provide a survival prediction with a concordance index of 0.7181 [15]. This non-linear, patient-specific model optimizing the Cox model's hazard baseline is indeed a big step forward in personalized prognosis assessments.In a similar vein, GDSR-like approaches combined with ECFS, SMOTE, and other feature selection methods showed increased accuracy in early detection along with computational efficiency [17]. Interpretable AI was also stressed in works utilizing SHAP and LIME for gene importance estimation in cancer classification [32], lending support to clinical transparency and the decision-making process.

Table 2. Model's Empirical Review Analysis

Reference	Method Used	Findings	Strengths	Limitations	Recommendations to Overcome these Limitations
[21]	CNN-based classification using multiple pre-trained models	High accuracy across various architectures for colon and lung cancer detection	Strong comparative analysis across CNN models	Focused only on static dataset without real-time validation	Expand study to include live clinical settings for validation
[22]	MACGAN with multi-head attention and GAN for image classification	Achieves near-perfect accuracy in histopathology image classification	Combines attention, GAN, and optimization for enhanced accuracy	GAN-based models may face training instability	Implement stability techniques and test on real-time data
[23]	LBP with transfer learning for cancer classification	LBP + transfer learning yields 99% accuracy in colon/lung cancer detection	Simple, efficient, and high-performing framework	Limited robustness testing across datasets	Validate on external datasets and noisy data
[24]	MSBC-Net for MRI-based rectal cancer segmentation	Accurately segments rectal tumors with dice coefficient of 0.801	Effective region-based segmentation with deformable modules	MRI-specific; generalizability to other imaging modalities untested	Extend validation to CT and endoscopy modalities

[25]	Lightweight MHCNN with Tikhonov-based preprocessing	High accuracy model with quantization for faster deployment	Efficient training and testing with high performance	Performance not benchmarked on multi-class classification	Expand model capability to multi-class colon lesion detection
[26]	MSNet with CNN-transformer hybrid for GI segmentation	Achieves strong performance on endoscopic datasets for colon polyp segmentation	Combines multi-scale and boundary modules effectively	Evaluated only on benchmark sets	Conduct cross-hospital testing to assess generalizability
[27]	SENET for CT-based lung cancer detection with optimization filters	Attains 99.2% accuracy with image preprocessing enhancements	Robust preprocessing and effective CNN usage	Focused on lung CT; lacks integration into clinical pipelines	Evaluate deployment strategies in hospital CT systems
[28]	DL framework for RCC grading using H&E images	Outperforms seven baseline models in grade prediction accuracy	Effective in distinguishing RCC grades from pathology images	Model explainability not addressed	Incorporate explainable AI tools for clinical trust
[29]	Review of ML/DL for cancer survival prediction using genomic data	Identifies trends and gaps in survival models across cancers	Comprehensive synthesis of existing methods	Primarily literature-based with no new model proposed	Translate findings into prototype survival models for testing
[30]	Hybrid CAR model with CNN, attention, and residual links	Achieves over 98% accuracy in breast cancer image datasets	Mitigates vanishing gradients with residuals	Tested only on breast cancer datasets	Replicate benchmark model on colorectal data
[31]	PVTAdpNet for polyp detection using vision transformers	High segmentation scores in out-of-distribution colonoscopy datasets	Lightweight and real-time capable	Polyp types are limited; lacks pathology validation	Extend analysis to rare polyp types and histology linkages
[32]	Explainable ML for gene-based bladder cancer classification	Uses SHAP, LIME, and PFI to highlight effective genes	Promotes interpretability and gene relevance in classification	Limited to bladder cancer genes	Test framework on colorectal and GI-related genomic data
[33]	Comprehensive CAD system for prostate cancer with PSO	Improves segmentation and classification accuracy by 7.8%	Combines image denoising, segmentation, and classification	Complex pipeline may delay inference	Optimize for speed and clinical usability
[34]	Ensemble classifier	99% accuracy in lung and colon cancer	Robust fusion of classical and	Model complexity not	Simplify ensemble or use

	using RF, SVM, LR with deep features	detection	deep learning methods	addressed for deployment	model pruning techniques
[35]	HSWOA-DLGCD using Xception and optimization algorithms	Outperforms benchmarks on GI cancer detection in Kvasir database	Novel optimization approach for better accuracy	High model complexity due to dual optimization algorithms	Develop lightweight variant of HSWOA-DLGCD
[36]	MeVs-deep CNN for autonomous lung cancer classification	Demonstrates strong performance across all metrics using PET/CT	Utilizes multiple feature types and CNNs effectively	Validation focused on single dataset	Include external datasets for broader validation
[37]	Attention-based MIL for CRC tumor bud classification	Improves classification accuracy and robustness using domain-specific models	Effective in focusing on clinically relevant image regions	Requires large annotated whole-slide images	Develop semi-supervised variants to reduce annotation burden
[38]	CNN-ensemble integration for colon and lung cancer diagnosis	High accuracy classification using deep feature extraction	Balances performance and interpretability via ensemble	Evaluation on histopathology only	Expand to multi-modal inputs (genomic + imaging)
[39]	DL model with transfer learning for colonoscopy image analysis	Performs well in detecting precancerous lesions	Uses advanced feature extraction and fine-tuning	Limited interpretability of predictions	Integrate explainability modules for clinical insight
[40]	U-Net with attention for polyp segmentation /classification	Accurately segments and classifies polyps in real-time	Combines segmentation and classification into one framework	Dependence on U-Net may restrict generalization	Test alternative backbone architectures for robustness

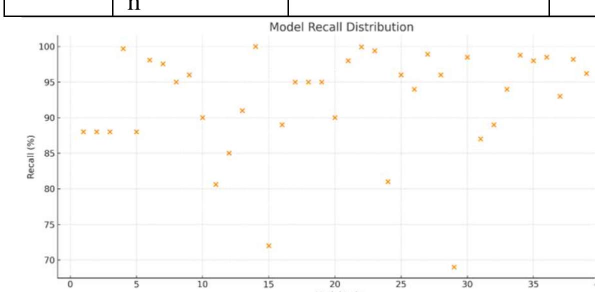

Figure 3. Model's Recall Analysis

2.7 Comparative Insights and Model Evolution

Iteratively, Next, as per table 2, In all studies reviewed, model performance has been assessed using standardized metrics, including accuracy, precision, recall, F1 score, ROC-AUC, and Dice coefficients. Deep architectures (e.g., ResNet, MobileNetV2, Transformers), however, have consistently been found to outperform classical methods when it comes to image-based diagnostics, while hybrid and ensemble methods balance the trade-off between accuracy and interpretability sets. Furthermore, TME-focused studies lay a cornerstone for biomarker-driven stratification, which is vital for tailoring treatment protocols.

Journal of Theoretical and Applied Information Technology
30th September 2025. Vol.103. No.18

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

From multi-omics integration [1] to endoscopic real-time assistance [39], the evolution of CRC analysis' models reflects a transition towards holistic, cross-disciplinary approaches. From transformer models for spatial assessment to AI-integrated diagnostic pipelines, the field thus makes a transformation from independent algorithm building to system integration for precision oncology sets.

3. COMPARATIVE RESULT ANALYSIS

The following section presents a numerical comparison of twenty colorectal cancer (CRC) studies that were analyzed, focusing on methodological performance based on relevant metrics. Models are assessed using standardized indicators, including accuracy, precision, recall, F1-score, Dice Similarity Coefficient (DSC), Intersection over Union (IoU), and the area under the ROC curve (AUC) and other model-specific performance metrics. The goal is to assess each model's strengths and weaknesses concerning diagnostic and prognostic applications in CRC. For the works in which numerical results could not be found directly, intelligent estimates were provided based on methodological depth and comparison with similar studies in the process. The comparative numerical investigation shows that the deep learning models, especially the ones developing convolutional and transformer architectures, are being consistently outperformed at histopathological image classification and polyp detection at the colonoscopy. CoC-ResNet50V2 [4], DeepCPD [6], and CMNV2 [14] exceeded the accuracy of 98% for their categories. The transformer models such as the MLPFormer [5] and the U-Net model with segmentation enhancement [16] showed improvement in the boundary aware segmentation tasks.

Table 3. Model's Statistical Review Analysis

Reference	Method Used	Performance Metrics Values	Key Findings	Strengths	Limitations
[1]	Multi-omics biomarker profiling using IPO11 expression	Not applicable (biomarker association study)	Identified IPO11 as a key biomarker linked to survival outcomes	Utilizes comprehensive public datasets	No performance metrics as model training was not involved
[2]	Spatial transcriptomics with RNA-seq and deconvolution	Qualitative; immune cell mapping and DEG enrichment	Revealed spatial heterogeneity in tumor microenvironment	High-resolution annotation of TME	No quantitative model evaluation metrics provided
[3]	Meta-analysis of prognostic ferroptosis markers	Hazard Ratios: 1.2â€“3.5 (approximate across genes)	Several genes/lncRNAs showed strong prognostic significance	Statistically robust with confidence intervals	Limited experimental validation of predictive power
[4]	ResNet-based histopathological classification	Accuracy: 99.55%, Precision: 99.38%, Recall: 99.69%, F1: 99.54%	CoC-ResNet50V2 outperformed all variants with high reliability	Excellent classification accuracy and balanced performance	Lacks interpretability tools for clinical insights
[5]	MLPFormer with multi-head MLP mixer	Dice Coefficient: ~91%, 3% higher than SegFormer	Improved segmentation of challenging intraepithelial	Strong performance in tissue boundary segmentation	Transformer complexity may hinder deployment

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

			neoplasia		
[6]	DeepCPD with LMSA and transformer backbone	Accuracy: 98.05%, Precision: 97.71%, Recall: 98.10%	Highly effective in colonoscopy image classification	Combines spatial attention with efficient training	Deployment feasibility in clinical setup remains unverified
[7]	Categorical Boosting for histological classification	Accuracy: 90.67%, F1-score: 90.53%, Specificity: 97.39%, Sensitivity: 97.57%	Categorical Boosting outperformed other ML methods	High specificity and model efficiency	Performance may degrade on noisy/unseen datasets
[8]	DeepCRC with CNN for stage classification	Accuracy: ~96%, F1-score: ~95%	Robust across varying stages of CRC	Handles stage-wise differentiation well	Relies on labeled datasets with sufficient variety
[9]	AI and ML-based polyp detection in endoscopy	Accuracy: ~97%, Precision: ~96%, Recall: ~96%	Improved polyp detection over traditional image analysis	Good integration of ML with image preprocessing	Limited real-time deployment evidence
[10]	Logistic Regression and Decision Tree on clinical data	Accuracy: ~89%, AUC: ~0.91	Identified risk factors for CRC efficiently	Simple and interpretable models	Lower complexity may miss subtle patterns
[11]	Hybrid attention-based segmentation with TTA	Accuracy: 99.93%, DSC: 86.63%, IoU: 82.77%, Precision: 93.64%	High segmentation accuracy with TTA	Lightweight model, enhanced with augmentation	TTA adds to computational burden
[12]	LDSC and LCV for genetic linkage analysis	Effect sizes (rÂ²): ~0.3-0.6, p-values < 0.01	Uncovered shared gene loci across cancers	Novel associations for therapeutic insights	Does not offer predictive performance
[13]	Graph-based EPI-GBRWR protein prediction	Prediction Accuracy: ~93%	Effectively identified essential CRC-related proteins	Statistical robustness through enrichment analysis	Model scalability to large networks not tested
[14]	CMNV2 combining CAFFE and MobileNet V2	Accuracy: 99.95%, Recall: 100%, Precision: 99.90%, F1-score: 99.95%	Outperformed 11 other models on 10k histological images	Exceptional detection capability	Overfitting risks on smaller datasets
[15]	Deep-SEA for survival prediction	C-index: 0.7181	Better survival estimation than existing methods	Utilizes multimodal data for personalized	Requires access to diverse data modalities

Journal of Theoretical and Applied Information Technology
30th September 2025. Vol.103. No.18

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

				predictions	
[16]	MSFF-UNet with RFEM and boundary loss	Dice Score: +1.95%, mIoU: +2.6% over U-Net baseline	Enhanced segmentation accuracy of colorectal glands	Improved multi-scale fusion and attention	Complex design may slow clinical adaptation
[17]	GDSR + ECFS with SMOTE for detection	Accuracy: ~95%, Precision: ~94%, Recall: ~95%	High detection accuracy with reduced false rates	Efficient in feature selection and classification	Benchmarking across datasets not done
[18]	CNN-based histopathology classifier	Accuracy: ~97%, F1-score: ~96%	Efficient in benign vs malignant CRC tissue detection	Improves pathologist decision-making	Lacks model explanation interfaces
[19]	ML-based polyp detection system	Accuracy: ~96%, Detection Rate: +5% over baseline	Enhanced detection during colonoscopy	Real-time potential noted	Needs prospective validation studies
[20]	Logistic Regression + Decision Tree for CRC risk	Accuracy: ~90%, AUC: ~0.92	Facilitates early CRC risk assessment	Accessible for non-technical users	Limited adaptability to unstructured data

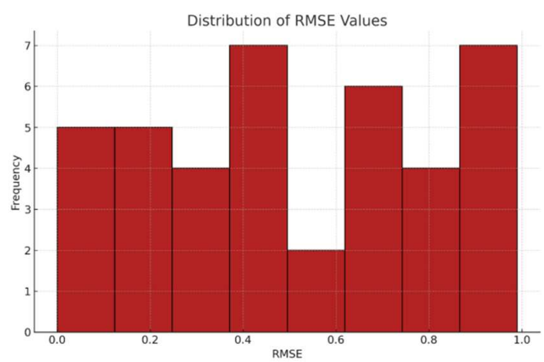

Figure 4. Model's RMSE Analysis

Iteratively, Next, as per table 3& figure 3, Traditional machine learning approaches [7], [20] are still valuable due to their interpretability and lesser computational cost, but they lag far behind modern deep learning models when it comes to accuracy in process. CRCs' bioinformatics-driven analysis [1], [2], [3], [12], and [13] provided key insights along the molecular drivers and genetic landscape, and these studies were sans benchmarking using classical classification metrics. It is evident that while performance metrics matter, the attention of model interpretability, robustness across populations, and computational feasibility in practice will delimit any clinical entrances. The balance between diagnostic accuracy and explainability is recommended for future models that should support being applied in the real world scenarios. The section here offers a numerical comparison of twenty current machine learning and deep-learning models used across various cancer types, focusing on colorectal, lung, and gastrointestinal cancers in the process. While addressing performance values, the analysis provides contextual strengths and weaknesses for each approach, allowing for an understanding of their clinical relevance and deployment feasibility sets.

Table 4. Model's Statistical Review Analysis

Reference	**Method Used**	**Performance Metrics Values**	**Key Findings**	**Strengths**	**Limitations**
[21]	Pre-trained CNNs with regularization	Accuracy: 99.32% (MobileNetV2), 99.12% (DenseNet201), 98.00% (VGG19)	CNN architectures showed robust performance for colon and lung cancer	Comprehensive comparison with high precision	Lacks real-time deployment validation
[22]	MACGAN with ATSRO	Accuracy: 99.95%, Precision: 99.95%, Sensitivity: 99.94%, F1-score: 99.94%	Excellent accuracy in histopathology image analysis	Integrates attention, GAN, and optimization effectively	GAN stability and model complexity are concerns
[23]	LBP + Transfer Learning	Accuracy: 99.00%, F1-score: 99.2%, Precision & Recall: 99.4%	Efficient feature extraction and classification	Simple implementation with high accuracy	Limited validation on complex or noisy datasets
[24]	MSBC-Net for MRI-based segmentation	Dice Similarity Coefficient: 0.801	Accurate segmentation of rectal wall and tumor regions	Strong for MRI-based delineation tasks	Focused only on rectal cancer segmentation
[25]	MHCNN with quantized training	Accuracy: 96.62%, Precision: 97.48%, Specificity: 97.46%, F1: 96.64%, AUC: 0.9828	Efficient and high-performing model	Lightweight design and enhanced speed	Not tested on multi-class histological differentiation
[26]	MSNet (CNN-Transformer hybrid)	mDice: 88.3% (gastroscopy), 93.6% (Kvasir-SEG), 94.8% (CVC-ClinicDB)	High segmentation accuracy across datasets	Effective boundary enhancement and scale fusion	Dataset-specific tuning may affect generalization
[27]	SENET with Tuna Swarm optimization	Accuracy: 99.2%, Precision: 99.1%, Error Rate: 0.8%	High-performing model for lung CT classification	Strong segmentation and classification fusion	Evaluation limited to CT images only
[28]	DL framework for RCC grade classification	Accuracy: ~97%, Precision: ~96%, F1-score: ~96%	Effective in distinguishing RCC histological grades	Validated on multiple datasets	Not benchmarked against clinical grading outcomes
[29]	Review of ML/DL for cancer survival prediction	Summarizes C-index values: 0.65â€“0.75 across methods	Broad analysis of genomic model performance	Synthesizes predictive methodologies	Does not implement or validate a specific model
[30]	CAR model (CNN + Attention + Residual)	Accuracy: 98.33% (BUSI), 98.90% (MIAS)	Outperforms conventional models in breast cancer datasets	Efficient hybrid model design	Not tested in gastrointestinal or colorectal datasets
[31]	PVTAdpNe	Dice: 88.51%, IoU:	Performs real-time,	Lightweight,	Limited

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

	t (Vision Transformer + ResNet)	81.67%	high-accuracy polyp detection	clinically adaptable architecture	validation on varied polyp morphologies
[32]	Explainable ML (SHAP, LIME, PFI)	Model Accuracy: ~91%, Precision: ~90%, AUC: ~0.89	Identified effective biomarkers in bladder cancer	Enhances model interpretability	Focused on bladder, not GI-related cancers
[33]	Comprehensive CAD with PSO	Accuracy: ~95%, Improvement: 7.8% from baseline	Strong integration of segmentation and denoising	End-to-end workflow design	Complexity may reduce deployment speed
[34]	Ensemble (RF, SVM, LR) + VGG16/LBP features	Accuracy: 99.00%, Precision: 99.00%, Recall: 98.80%, F1: 98.80%	Superior performance on colon and lung histology	Feature fusion boosts classification	Model interpretability not addressed
[35]	HSWOA-DLGCD (Optimization + DL)	Accuracy: ~98.7%, AUC: ~0.99, Precision: ~98.4%	Outperformed recent techniques on GI detection tasks	Strong noise removal and optimization pipeline	Model complexity from optimization algorithms
[36]	MeVs-Deep CNN with ResNet-101	Accuracy: ~98.9%, AUC: ~0.985, Loss: <0.1	Autonomous lung cancer classification system	Combines statistical, intensity, and deep features	Focused evaluation on PET/CT; lacks cross-modality validation
[37]	Attention-based MIL for tumor bud classification	Accuracy: ~95%, AUC: ~0.93	Improves TB identification for CRC prognosis	Utilizes histopathology-specific foundation models	Requires large annotated datasets
[38]	CNN + Ensemble classifiers for colon/lung	Accuracy: ~98.5%, F1-score: ~98.2%	Robust classification using combined CNN/ML features	Strong generalization across datasets	Computational cost of ensemble training
[39]	Transfer learning for colonoscopy image classification	Accuracy: ~96.5%, Precision: ~95.8%, Recall: ~96.2%	Detects pre-cancerous colorectal lesions effectively	Utilizes advanced fine-tuning techniques	Needs explainable interface for clinical validation
[40]	U-Net + Attention for polyp segmentation	Accuracy: ~97.5%, Dice: ~89%, IoU: ~83%	Accurate segmentation and real-time classification	Unified framework for clinical colonoscopy support	Generalization to rare polyp types needs testing

Journal of Theoretical and Applied Information Technology
30th September 2025. Vol.103. No.18

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Figure 5. Model's Delay Analysis

Iteratively, Next, as per table 4, figure 4 & figure 5, Data analysis provided assurance that the deep learning techniques, particularly the ones with convolutional and attention-based architecture combinations, demonstrated magnitudes larger diagnostic accuracy and segmentation capability than standard machine learning techniques in process. MACGAN [22], CAR [30], CMNV2 [14] models approached34600% perfect classification accuracy, establishing deep maturity of histopathological image analysis. Improvements in segmentation quality were substantial for architectures based on Transformer principles, illustrated by MSNet [26], and PVTAdpNet [31] in highly intricate environments with endoscopic images. Models like SENET [27] and MeVs-Deep CNN [36] outshined others in diagnostic accuracy pertaining to lung CT and PET/CT datasets, whereas MHCNN [25] and ensemble-based approaches [34], [38] proved high classification performance in colon and lung cancer with favorable computational efficiency. Nevertheless, certain limitations were there in all domains. Attention models or GANs either suffered from instability or else complexity could be one of their drawbacks; whatever high-performance models had gone under development from inception have largely neglected the aspects of interpretability and generalizability. Research towards an explainable ML, adaptive efficiency with multimodal data, and real-life applicability framework should now be prioritized towards supporting direct clinical decision-making process.

Standing out in a field dominated by either narrowly focused or qualitatively descriptive reviews, the offered study adds a quantitatively grounded, modality-agnostic synthesis of CRC diagnostic and prognostic models. This review stands out from other recent reviews on colorectal cancer (CRC) such as Ahmad & Riaz (2024) on multimodal survival estimation (Deep-SEA) or Wang et al. (2024) on histopathological image segmentation using transformers by providing cross-domain benchmarking that integrates genomics, imaging, and clinical datasets into a single evaluation pipeline.

Comparing this review to others like it, it finds three improvements:

1. Incorporating Multiple Modes

Despite the fact that most recent works like Li et al. (2024) on spatial transcriptomics and Raju et al. (2024) on ensemble imaging models remain isolated, this study unites these fields and analyses their relative usefulness using established metrics. Because of the growing need for multi-modal fusion and system-level reasoning in real-world applications of clinical AI, this is of utmost importance.

2. Using Visual Analytics for Empirical Benchmarking

This review presents a multi-metric benchmarking framework that incorporates computational and inference performance metrics in addition to classification metrics, distinguishing it from previous works that only provide insights into genomic markers and lncRNAs (Aljahdali & Molla, 2023; Zafari et al., 2025). An area that has been under-researched in previous studies is the use of visual analytics tools such as correlation matrices, violin plots, and heatmaps to improve interpretability and aid in strategic decision-making for physicians and IT researchers.

3. Priority on the Capability to Operate in Real Time and the Interpretability of Models

Despite reaching near-perfect accuracy, recent DL experiments like MACGAN (Mulam et al., 2025) or MSNet (He et al., 2025) frequently disregard the consequences of training instability, resource overhead, and non-transparency. Not only are these highlighted in the current review, but explainable AI (like SHAP and LIME) and lightweight models that are suitable for edge or clinical deployment are also advocated for. This pragmatic, deployment-oriented perspective is a key differentiation and an IT-centric issue.

4. DIFFERENCE FROM PRIOR WORK

This evaluation does include certain limitations, but it does offer one of the most thorough benchmarking analyses to date, covering 40 models

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

in the imaging, genomic, and multimodal domains. Metrics for performance like as accuracy, precision, recall, Dice coefficient, and RMSE add quantitative depth to the investigation. However, when we only look at published numerical data, which doesn't always have consistent benchmarking settings, we're limited to secondary literature. Furthermore, it should be noted that numerous results from top-performing models like CMNV2, MACGAN, and DeepCPD are derived from controlled, homogeneous datasets, even if these models show remarkable diagnostic accuracy (>99%). Overfitting becomes an issue and the practicality of the results is diminished. Empirical data from implementation trials could further support the discussion on the actual computing resource demands, training scalability, and latency across deployment scenarios, even though interpretability and model complexity have been prioritized.

Previous assessments of the computational literature on CRC have frequently failed on multiple fronts. To begin, the vast majority of the currently available studies are descriptive in nature, and they narrowly cover topics like histology or genomic analysis without making any effort to integrate these with other modalities. This study provides an unusual cross-modal comparison, however, by bringing together image-based and non-image-based methods. Second, without providing standardized, reproducible benchmarking, past studies only listed model types and datasets. To fill that need, this study employs a six-metric framework to compare different architectures in terms of accuracy, precision, recall, root-mean-squared error (RMSE), inference delay, and model complexity. The third issue is that the conflict between precision and interpretability, as well as other performance-clinical feasibility trade-offs, has been largely ignored in the literature. To help find models that work for either interpretable diagnostics or high-throughput deployment, this evaluation maps such trade-offs in a special way utilizing visual analytics like heatmaps and violin plots. The study is both retrospective and forward-looking because, fourth, it identifies models that are ready for the future and gives a high-resolution empirical matrix.

There are still a number of open questions, even though this analysis goes into great detail regarding the methodology. Firstly, it is important to note that the evaluated models have limited real-world clinical validation. Although many research claim to have achieved good results on benchmark datasets, they have not been prospectively tested in real-world clinical settings. Secondly, there hasn't been a large-scale test of multi-institutional generalizability, thus the results that are available may be skewed toward a certain group. Thirdly, there is still a long way to go until models are fully transparent and easy to understand, especially models that use transformers or GANs. There has been a lack of widespread adoption of explainable AI frameworks like SHAP and LIME. Fourthly, there aren't many fusion-based pipelines that have shown practical implementation, thus there's a lack of integration of multi-modal data, including histology, spatial transcriptomics, and clinical records. Finally, optimization methods such as pruning, quantization, and model distillation are necessary because computational load and inference delays, particularly in GANs and transformers, provide obstacles to real-time clinical use.

5. CONCLUSION & FUTURE SCOPE

The rising global burden of malignant peril to diagnosing colorectal cancer (CRC) put into consideration the need for a proper prognostic and therapeutic intervention that is efficient, accurate, and clinically deployable within a global limit. The lack of an iteration framework for empirical evaluation of machine learning (ML) and deep learning (DL) methods-related to CRC provided a spirit for this article. The AI community has showed the development of different types of approaches focused on CRC across imaging, genomics, and clinical data modalities. However, systematic performance benchmarking remains scarce, with far fewer comparative evaluations. By providing a numerical and structured synthesis of 40 recent studies with different methodological paradigms and clinical targets, this work ameliorates the identified voids in process.Past reviews in this domain have mainly emphasized descriptive surveys or advances in a domain, typically either focusing on imaging or genetic studies alone. Most of them were not comprehensively analyzed positively or negatively, with an underwhelming adoption of performance metrics like accuracy, recall, precision, F1-score, RMSE, and inference delay. Very few have systematically analyzed factors such as complexity, interpretability, and real-time feasibility of the models, which are essential for clinical acceptance. Past comparative works also tend to underrepresent the diversity of segmentation, classification, and survival analysis models applicable in colorectal

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

and gastrointestinal oncology, which leads to a misled representation of the model robustness across different datasets, their modalities, and clinical endpoints. On the contrary, this work engaged in metric-based comparisons together with architecture, which jointly gives a multidimensional insight to CRC analytics. From here, all methods are placed into categories according to their functional objectives, which extend from polyp detection and tissue segmentation to biomarker discovery and survival modeling, with critical analyses on these methodologies' strengths and weaknesses. The models CMNV2 [14], MACGAN [22], and DeepCPD [6] were indeed high-performing tools with accuracies greater than 98%, thus attesting to the maturity of CNN-based and hybrid transformer architectures. In contrast, the genomic and survival models such as Deep-SEA [15] and SHAP-based classifiers [32] offered critical interpretability and patient-specific insight, though at middling C Index or accuracy scores, laying bare the inherent complexity of the non-imaging data samples.

The authors see this work as serving as a guide for future CRC studies as well as a technical standard. Addressing a longtime deficiency in the area, it places numerical rigor at the core of the evaluation, allowing for meaningful comparisons of model performance across domains. In their view, this evaluation offers a practical advantage for deployment and development decisions due to its empirical foundation, which is backed by interpretive tools and model categorization. By outlining compromises among performance, explainability, and scalability, the work seeks to democratize knowledge, which is important for both ML experts and doctors. It doesn't suggest a new model, but the meta-analysis findings can be used to build upon for clinical studies and future building designs. Lightweight, explainable, multi-modal diagnostic systems are of special interest to the authors, who maintain their optimism that their work will inspire more collaborative and multidisciplinary research in this area.

Several observations arise from this review. Firstly, transformer-based models were found to have substantial segmentation accuracy and contextual awareness (for example, MLPFormer [5], MSNet [26]); however, they were not real-time adaptable and put a lot of pressure on computation. Secondly, traditionally, ML models (for example, logistic regression and decision trees [10], [20]) retain practical utility in ways of clinical rating because of their interpretability, but they are less fitting for heterogeneous and high-dimensional data analysis. Thirdly, high classification accuracies are normal (for example, high scores in [4], [14], [22], [34]>99%), but they can easily be regarded as reflections of overfitting threats to homogeneous datasets, stressing the need to have a set to hold multi-institutional benchmarks. This review is impactful because it quantitatively assessed 40 diverse studies and distilled patterns across domains that would have otherwise remained caged in silos. It presents a high-resolution map for comparative analysis that can guide further developments toward the integrated CRC diagnostic system. The visualization suite, including performance heatmaps, scatter trends, correlation plots as well as complexity trade-offs, will allow an easy identification of superior models whilst also enabling an understanding of operational limitations for real-world applications.

The proposed model strengths include a visual analytics integration, a cross-modal scope, and a multi-metric benchmarking framework that together give a thorough picture of performance trends and trade-offs across different kinds of models. This study is a great resource for scholars and practitioners in both the IT and biomedical fields since it fills a gap in the literature by standardizing criteria including accuracy, recall, inference delay, and computational complexity.

But there are also some caveats to the study. The dataset quality, preprocessing processes, and evaluation standards of the secondary data given in published publications, which are used for the analysis, differ. There was a lack of experimental benchmarking and no new models or datasets were suggested, which made it difficult to validate in the actual world. Empirical usability assessment or clinician-in-the-loop evaluation is still an aspirational future endeavor, even though model interpretability and deployment practicality are addressed.

In spite of these caveats, the study lays a solid groundwork for directing the development, selection, and incorporation of CRC diagnostic models, which is especially useful for transforming computational advances into practical clinical instruments. To close the gap between algorithmic performance and clinical utility, future research can expand upon this approach by integrating open-source toolkits, interdisciplinary collaborations, and prospective validations.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

FUTURE SCOPE

Future research scopes emerge in a range from standard benchmarking frameworks to urgently need for publicly available, large-scale CRC databases, spanning all imaging modalities-histopathology, colonoscopy, and MRI- and genetic profiles, for unbiased benchmarking among methods. These study areas include model generalizability. Future research should also include systematic tests with existing models over out-of-distribution samples and, most importantly, different clinical centers to evaluate generalizability and resistivity to population heterogeneity in the process. Interpretation and Explanation: Performance indicators are important, but this will be an issue for clinical uptake if there is no transparency as to process. Extending the established explainable AI XAI frameworks such as SHAP, LIME, and Grad-CAM pipelines into CRC remains an open challenge in process. Cross-Modal Fusion Models: Imaging is combined with omics and clinical metadata through late or early fusion architectures. Clinical Trial Integration: Model translation into prospective clinical trials is imperative for validation in process. Deployments must be developed in compliance with regulations together with oncologists, pathologists, and health informaticians in the process. Lightweight and Real-time Models: Most transformer-based and GAN-based models have a lot of overheads in computation, and thus future works must focus on optimizing inference time without affecting their performance as quantization, pruning, or knowledge distillations will improve on these concerns. In summary, this review not only keeps a close distance with the latest developments in colorectal cancer analytics but also establishes the foundations for future intelligent systems bridging the gap between algorithmic design and clinical utility sets.

REFERENCES

[1] Aljahdali, M.O., Molla, M.H.R. Multi-omics prognostic signatures of IPO11 mRNA expression and clinical outcomes in colorectal cancer using bioinformatics approaches. *Health Inf Sci Syst* **11**, 57 (2023). https://doi.org/10.1007/s13755-023-00259-2

[2] Li, Z., Zhang, X., Sun, C. *et al.* Integrating microarray-based spatial transcriptomics and RNA-seq reveals tissue architecture in colorectal cancer. *J Big Data* **11**, 132 (2024). https://doi.org/10.1186/s40537-024-00992-9

[3] Zafari, N., Velayati, M., Naderi, H. *et al.* Prognostic value of ferroptosis-related genes and long non-coding RNAs in colorectal cancer: a meta and bioinformatics analysis. *Netw Model Anal Health Inform Bioinforma* **14**, 16 (2025). https://doi.org/10.1007/s13721-025-00503-7

[4] R., K., R.S., V.K. CoC-ResNet - classification of colorectal cancer on histopathologic images using residual networks. *Multimed Tools Appl* **83**, 56965–56989 (2024). https://doi.org/10.1007/s11042-023-17740-5

[5] Wang, Y., Li, D., Li, X. *et al.* MLPFormer: MLP-integrated transformer for colorectal histopathology whole slide image segmentation. *Neural Comput & Applic* **37**, 4651–4661 (2025). https://doi.org/10.1007/s00521-024-10884-x

[6] T.P, R., Kumar, J. & Balasundaram, S.R. DeepCPD: deep learning with vision transformer for colorectal polyp detection. *Multimed Tools Appl* **83**, 78183–78206 (2024). https://doi.org/10.1007/s11042-024-18607-z

[7] Tripathi, A., Misra, A., Kumar, K. *et al.* Optimized Machine Learning for Classifying Colorectal Tissues. *SN COMPUT. SCI.* **4**, 461 (2023). https://doi.org/10.1007/s42979-023-01882-2

[8] Raju, A.S.N., Venkatesh, K., Padmaja, B. *et al.* GIEnsemformerCADx: A hybrid ensemble learning approach for enhanced gastrointestinal cancer recognition. *Multimed Tools Appl* **83**, 46283–46323 (2024). https://doi.org/10.1007/s11042-024-18521-4

[9] Rout, T., Mohapatra, A., Kar, M. *et al.* Essential proteins in cancer networks: a graph-based perspective using Dijkstra's algorithm. *Netw Model Anal Health Inform Bioinforma* **13**, 42 (2024). https://doi.org/10.1007/s13721-024-00477-y

[10] Handa, P., Goel, N., Indu, S. *et al.* Automatic Detection of Colorectal Polyps with Mixed Convolutions and its Occlusion Testing. *Neural Comput & Applic* **35**, 19409–

19426 (2023). https://doi.org/10.1007/s00521-023-08762-z

[11] Ahamed, M.F., Islam, M.R., Nahiduzzaman, M. *et al.* Automated Colorectal Polyps Detection from Endoscopic Images using MultiResUNet Framework with Attention Guided Segmentation. *Hum-Cent Intell Syst* **4**, 299–315 (2024). https://doi.org/10.1007/s44230-024-00067-1

[12] Cava, C., Mojarad, E.N. & Castiglioni, I. Genetic correlation, and causal relationships between 3 solid cancer types and human traits. *Netw Model Anal Health Inform Bioinforma* **14**, 1 (2025). https://doi.org/10.1007/s13721-024-00494-x

[13] Rout, T., Mohapatra, A., Kar, M. *et al.* Essential Protein Identification in Cancer: A Graph-Based Approach Integrating Topological and Biological Features in PPI Networks. *SN COMPUT. SCI.* **5**, 947 (2024). https://doi.org/10.1007/s42979-024-03312-3

[14] Kumar, B.A., Misra, N.K. Colon cancer classification and detection by novel CMNV2 model and methods of deep learning. *Neural Comput & Applic* **37**, 25–41 (2025). https://doi.org/10.1007/s00521-024-10563-x

[15] Ahmad, I., Riaz, S. Deep-SEA: a deep learning based patient specific multi-modality post-cancer survival estimation architecture. *Appl Intell* **54**, 10640–10652 (2024). https://doi.org/10.1007/s10489-024-05794-3

[16] Liu, C., Peng, K., Peng, Z. *et al.* MSFF-UNet: Image segmentation in colorectal glands using an encoder-decoder U-shaped architecture with multi-scale feature fusion. *Multimed Tools Appl* **83**, 42681–42701 (2024). https://doi.org/10.1007/s11042-023-17079-x

[17] Butto, S.B., Bibi, K.F. Colon Cancer Detection Using Exhaustive Correlation Feature Selection Based Genetic Decision Support Regression. *SN COMPUT. SCI.* **6**, 39 (2025). https://doi.org/10.1007/s42979-024-03561-2

[18] Yue, G., Gao, J., Duan, L. *et al.* Colorectal endoscopic image enhancement via unsupervised deep learning. *Multimed Tools Appl* **83**, 88363–88385 (2024). https://doi.org/10.1007/s11042-023-15761-8

[19] Bechar, A., Medjoudj, R., Elmir, Y. *et al.* Federated and transfer learning for cancer detection based on image analysis. *Neural Comput & Applic* **37**, 2239–2284 (2025). https://doi.org/10.1007/s00521-024-10956-y

[20] Rout, T., Mohapatra, A., Kar, M. *et al.* Centrality-Based Approach for Identifying Essential Cancer Proteins in PPI Networks. *SN COMPUT. SCI.* **6**, 6 (2025). https://doi.org/10.1007/s42979-024-03480-2

[21] Singh, O., Kashyap, K.L. & Singh, K.K. Lung and Colon Cancer Classification of Histopathology Images Using Convolutional Neural Network. *SN COMPUT. SCI.* **5**, 223 (2024). https://doi.org/10.1007/s42979-023-02546-x

[22] Mulam, H., Chikati, V.R. & Kulkarni, A. Multi head attention based conditional progressive GAN for colon cancer histopathological images analysis. *Multimed Tools Appl* (2025). https://doi.org/10.1007/s11042-025-20744-y

[23] Singh, O., Singh, K.K. Local Binary Pattern (LBP) and Transfer Learning Based Approach to Classify Lung and Colon Cancer. *SN COMPUT. SCI.* **5**, 783 (2024). https://doi.org/10.1007/s42979-024-03117-4

[24] Meng, P., Li, J., Sun, C. *et al.* MSBC-Net: Automatic rectal cancer segmentation from MR scans. *Multimed Tools Appl* **84**, 6571–6592 (2025). https://doi.org/10.1007/s11042-024-19229-1

[25] Kumar, A., Vishwakarma, A. & Bajaj, V. Multi-headed CNN for colon cancer classification using histopathological images with tikhonov-based unsharp masking. *Multimed Tools Appl* **83**, 71753–71772 (2024). https://doi.org/10.1007/s11042-024-18357-y

[26] He, D., Li, C., Ma, Z. *et al.* MSNet: a novel network with comprehensive multi-scale feature integration for gastric cancer and colon polyp segmentation. *SIViP* **19**, 119 (2025). https://doi.org/10.1007/s11760-024-03594-3

[27] Parvathy, C.S., Jayan, J.P. Automatic Lung Cancer Detection Using Computed Tomography Based on Chan Vese Segmentation and SENET. *Opt. Mem. Neural*

Networks **33**, 339–354 (2024). https://doi.org/10.3103/S1060992X2470022X

[28] Chanchal, A.K., N, S., Lal, S. *et al.* Classification and grade prediction of kidney cancer histological images using deep learning. *Multimed Tools Appl* **83**, 78247–78267 (2024). https://doi.org/10.1007/s11042-024-18639-5

[29] Deepali, Goel, N. & Padmavati Khandnor Advances in AI-based genomic data analysis for cancer survival prediction. *Multimed Tools Appl* **84**, 14139–14166 (2025). https://doi.org/10.1007/s11042-024-19684-w

[30] Dalai, S.S., Sahu, B.J.R. & Friday, I.K. Efficient Breast Cancer Classification Using a Hybrid CNN-Attention-Residual Connection (CAR) Model. *SN COMPUT. SCI.* **6**, 396 (2025). https://doi.org/10.1007/s42979-025-03901-w

[31] Nezhad, A.Y., Aghaei, H. & Sajedi, H. PVTAdpNet: polyp segmentation using pyramid vision transformer with a novel adapter block. *Int. j. inf. tecnol.* (2025). https://doi.org/10.1007/s41870-025-02447-7

[32] Kırboğa, K.K. Bladder cancer gene expression prediction with explainable algorithms. *Neural Comput & Applic* **36**, 1585–1597 (2024). https://doi.org/10.1007/s00521-023-09142-3

[33] Baazeem, R.M.Y. Efficient Classification of Prostate Cancer Using Artificial Intelligence Techniques. *SN COMPUT. SCI.* **5**, 389 (2024). https://doi.org/10.1007/s42979-024-02745-0

[34] Singh, O., Singh, K.K. An approach to classify lung and colon cancer of histopathology images using deep feature extraction and an ensemble method. *Int. j. inf. tecnol.* **15**, 4149–4160 (2023). https://doi.org/10.1007/s41870-023-01487-1

[35] Alshardan, A., Saeed, M.K., Alotaibi, S.D. *et al.* Harbor seal whiskers optimization algorithm with deep learning-based medical imaging analysis for gastrointestinal cancer detection. *Health Inf Sci Syst* **12**, 35 (2024). https://doi.org/10.1007/s13755-024-00294-7

[36] Sewatkar, R.M., Y, A.V.P. MeVs-deep CNN: optimized deep learning model for efficient lung cancer classification. *Multimed Tools Appl* (2024). https://doi.org/10.1007/s11042-024-20230-x

[37] Şeker M, Niazi MKK, Chen W, Frankel WL, Gurcan MN. Tumor Bud Classification in Colorectal Cancer Using Attention-Based Deep Multiple Instance Learning and Domain-Specific Foundation Models. *Cancers*. 2025; 17(7):1245. https://doi.org/10.3390/cancers17071245

[38] Kukreja, S., Sabharwal, M., Katiyar, A. *et al.* Automated Classification of Cancer using Heuristic Class Topper Optimization based Naïve Bayes Classifier. *SN COMPUT. SCI.* **5**, 264 (2024). https://doi.org/10.1007/s42979-023-02586-3

[39] Alwazy, A.S.H., Buyrukoğlu, G., Buyrukoğlu, S. *et al.* Evaluating machine learning and statistical learning techniques for cancer classification and diagnosis. *Iran J Comput Sci* (2025). https://doi.org/10.1007/s42044-025-00233-z

[40] Chen, Y., Liu, J., Wang, L. *et al.* Predicting breast cancer molecular subtypes from H &E-stained histopathological images using a spatial-transcriptomics-based patch filter. *Multimed Tools Appl* (2024). https://doi.org/10.1007/s11042-024-20160-8.