Frontiers in Health Informatics
ISSN-Online: 2676-7104
www.healthinformaticsjournal.com
2024; Vol 13: Issue 8
Open Access

Advancements in Colorectal Cancer Predication and Classification using Deep Learning and Machine Learning Models

Kalpana K[1],Dr.G.N.K. Suresh Babu[2]
[1]Research Scholar, Sristhi College of Science and Management
Bangalore
[2]Professor-CSE Research Supervisor(University of Mysore) Sristhi College of Science and Management
Bangalore

Cite this paper as: Kalpana K ,Dr.G.N.K. Suresh Babu (2024). Advancements in Colorectal Cancer Predication and Classification using Deep Learning and Machine Learning Models. *Frontiers in Health Informatics*, *13 (8) 813-818*

Abstract
The goal of this work is to provide a thorough assessment of the most recent research on AI-based machine learning and deep learning methods used in colorectal cancer modeling. A detailed synopsis and a list of the studies collected under each topic are then given. By focusing on the technical and medical viewpoints, we critically examine the prospects and difficulties in colorectal cancer prediction using ML and DL algorithms as we wrap up our work. Finally, we think that scientists who are thinking about using ML and DL techniques to diagnose colorectal cancer will find our study useful. This study provides a thorough analysis of the earlier advancements made by scientists in the prediction of colorectal cancer using both ML and DL algorithms.

1. INTRODUCTION

Deep learning approaches have shown significant advantages over traditional methods in predicting outcomes for colorectal cancer (CRC) patients. These modern techniques leverage complex algorithms to analyze vast datasets, leading to improved accuracy and personalized treatment strategies. The following sections outline the key comparisons between deep learning and traditional methods.

1.1 Predictive Accuracy

- Deep learning models, such as artificial neural networks (ANN) and long-short term memory (LSTM) networks, have demonstrated superior predictive accuracy compared to traditional Cox proportional hazards (CPH) models, with AUC values reaching up to 0.910 for deep learning versus 0.793 for CPH(Qu et al., 2024).
- In histopathology image analysis, deep learning models like Xception+ achieved an accuracy of 99.37% in cancer diagnosis, significantly outperforming manual methods prone to human error(Kar & Rowlands, 2024).

1.2 Prognostic Indicators

- The colorectal cancer risk score (CRCRS) developed through deep learning serves as an independent prognostic indicator, enhancing risk stratification beyond traditional clinical staging systems(Wei et al., 2024).
- Deep learning models have also been effective in predicting treatment responses, achieving a 97% success rate in tailoring treatment regimens for CRC patients(Vinudevi et al., 2024).

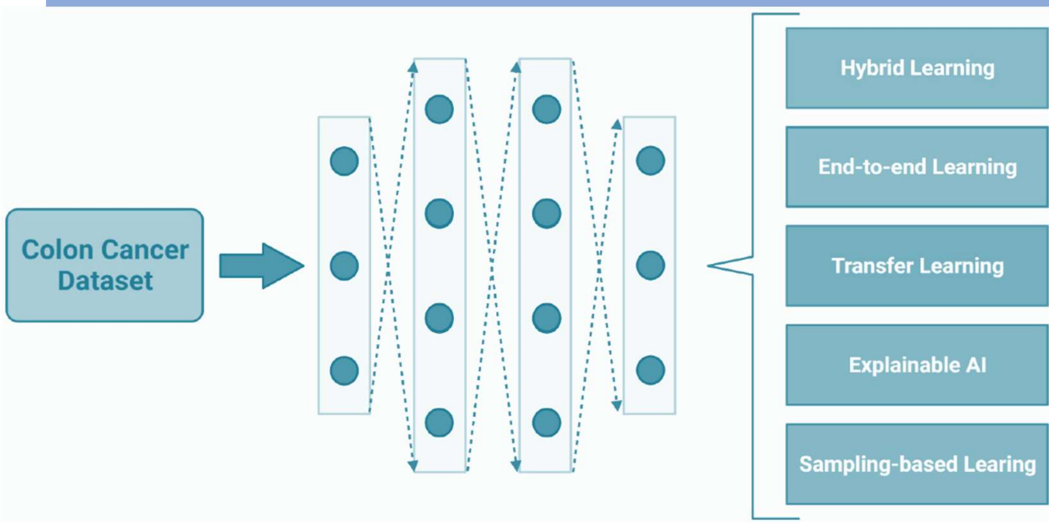

Figure 1 . Deep Learning Models for Colorectal Cancer

1.3 Automation and Efficiency

- Deep learning automates the detection of polyps, crucial for early CRC diagnosis, with models like DenseNet and EfficientNet achieving accuracies of 99% and 99.4%, respectively(Fanijo, 2024).
- This automation reduces the time and potential errors associated with traditional manual diagnosis, thereby improving patient outcomes(Kar & Rowlands, 2024).

Conversely, traditional methods like CPH models have been foundational in cancer prognosis but may lack the nuanced insights provided by deep learning. While they remain valuable, the integration of deep learning into clinical practice could redefine prognostic assessments and treatment strategies for CRC patients.

Deep learning models offer significant advantages in predictive accuracy for colorectal cancer outcomes compared to traditional methods, primarily through enhanced image analysis and biomarker identification. These models leverage complex algorithms to analyze histopathological images, leading to improved diagnostic precision and prognostic capabilities.

1.4 Enhanced Diagnostic Accuracy

- Deep learning models, such as GoogLeNet and Xception, have demonstrated high accuracy rates (up to 99.37%) in identifying cancerous tissues from histopathology images, outperforming traditional manual methods prone to human error(Kar & Rowlands, 2024).
- The integration of deep learning with multiomics data has resulted in the development of the colorectal cancer risk score (CRCRS), which serves as an independent prognostic indicator, enhancing predictive accuracy beyond conventional staging systems(Wei et al., 2024).

1.5 Automation and Speed

- Deep learning systems automate the detection process, significantly reducing the time required for diagnosis. For instance, models like DenseNet and EfficientNet achieved accuracy rates of 99% and 99.4%, respectively, in early detection of polyps, which are precursors to colorectal cancer(Fanijo, 2024).
- The use of convolutional neural networks (CNNs) allows for rapid classification of diverse tissue types, facilitating timely treatment decisions(Prezja et al., 2024).

While deep learning models show promise in improving predictive accuracy, traditional methods still play a role in clinical settings, particularly in cases where human expertise is essential for nuanced decision-making. Balancing these approaches may yield the best outcomes for colorectal cancer management.

Deep learning algorithms have significantly advanced the accuracy of colorectal cancer tissue identification through various innovative approaches. These algorithms leverage sophisticated architectures and techniques to enhance feature extraction and classification, leading to impressive accuracy rates in histopathological image analysis.

2. KEY ALGORITHMS AND TECHNIQUES

- **Atrous Convolution and Coordinate Attention**: The CCDNet model utilizes atrous convolution combined with a coordinate attention transformer, allowing it to capture both local and global features effectively. This approach achieved accuracy rates of 98.61% and 98.96% on different datasets(Khalid et al., 2024).
- **Triple Convolutional Neural Networks**: The Color-CADx system employs multiple CNNs (ResNet50, DenseNet201, and AlexNet) for feature extraction, followed by discrete cosine transform (DCT) for dimensionality reduction. This method reached an accuracy of 99.3% on the NCT-CRC-HE-100 K dataset(Sharkas & Attallah, 2024).
- **Divide and Conquer Strategy**: The DeepCon model implements a two-stage transfer learning approach, enhancing feature transferability and achieving an accuracy of 98.4% using the Xception network(Chughtai et al., 2024).
- **Hybrid Models**: The integration of CNNs with ensemble machine learning techniques, such as EfficientNetV2 paired with random forest classifiers, has shown promising results, achieving up to 99.89% accuracy on internal test sets(Prezja et al., 2024).

While these deep learning models demonstrate remarkable accuracy, challenges remain in ensuring their generalizability across diverse clinical settings and histopathological variations. Further research is needed to address these limitations and enhance the robustness of these systems.

3. DATA SETS

The availability of standardized datasets is crucial for enhancing the effectiveness of machine learning models in predicting colorectal cancer (CRC). Standardized datasets ensure consistency, improve model training, and facilitate the comparison of results across studies. This leads to more reliable predictions and better clinical decision-making. The following points highlight the significance of standardized datasets in CRC prediction.

Improved Model Training

- Standardized datasets provide a uniform structure, allowing for more effective training of machine learning algorithms, such as artificial neural networks and random forests, which have shown promising results in CRC risk stratification(Nartowt et al., 2020).
- The use of large, well-structured datasets enables the development of models with higher accuracy, as seen in studies achieving C-Index scores of 0.86 for overall survival predictions(Gründner et al., 2018).

Enhanced Data Integration

- Standardization facilitates the integration of diverse data types, including genomic, imaging, and clinical data, which is essential for comprehensive predictive modeling(Meldolesi et al., 2016).
- Automated data-mining processes can better handle the complexity of large datasets, leading to improved insights and interactions among variables(Meldolesi et al., 2016).

4. MACHINE LEARNING MODELS

Machine learning models have emerged as pivotal tools for risk stratification in colorectal cancer (CRC), enhancing the precision of patient management. Various algorithms have been developed and validated across different studies, demonstrating their effectiveness in predicting outcomes such as distant metastases, recurrence, and unplanned reoperations. The following sections outline the key models utilized in this domain.

4.1 Random Forest (RF) Models

- **Young-Onset Colorectal Cancer**: RF models achieved an AUC of 0.859 in internal validation and 0.888 in temporal validation for identifying high-risk individuals requiring colonoscopy(Zhen et al., 2024).
- **Distant Metastases Prediction**: In a large cohort, RF models demonstrated superior performance with AUCs of 0.843, 0.793, and 0.806 across training, test, and external validation sets(Wei et al., 2024).

5. DEEP LEARNING APPROACHES

- **Histological Analysis**: Deep learning models analyzing histological slides provided robust risk stratification, achieving similar performance to traditional binary predictors(Höhn et al., 2023).

CatBoost Classifier

- **Colon Cancer Recurrence**: This model excelled with an AUC of 0.92, highlighting the importance of various clinical factors in predicting recurrence(Kayikcioglu et al., 2024).

While machine learning models show promise in improving risk stratification in CRC, challenges remain in generalizing these models across diverse populations and clinical settings, indicating a need for further research and validation.
The integration of machine learning (ML) and deep learning (DL) techniques has significantly advanced the prediction of colorectal cancer (CRC) outcomes. These methodologies leverage various data sources, including histopathological images and clinical records, to enhance diagnostic accuracy and prognostic assessments. The following sections outline key contributions of ML and DL in CRC prediction.

Deep Learning in Histopathology

- **Model Performance**: Deep learning models, such as CNNs, have demonstrated high accuracy in analyzing histopathological images, achieving up to 99.89% accuracy in some studies(Prezja et al., 2024).
- **Prognostic Indicators**: The colorectal cancer risk score (CRCRS) developed through deep learning serves as an independent prognostic indicator, correlating with patient outcomes(Wei et al., 2024).

Machine Learning for Risk Assessment

- **Predictive Modeling**: Machine learning algorithms can effectively classify and predict CRC development, providing insights into influential variables affecting carcinogenesis(Chen & He, 2024).
- **Multimodal Approaches**: Combining imaging data with clinical records enhances predictive capabilities, achieving improved risk assessment outcomes(Jiang et al., 2024).

6. CHALLENGES AND FUTURE DIRECTIONS

Despite the promising results, challenges remain in standardizing these technologies for clinical use. Variability in histopathological interpretations and the need for extensive datasets for training models are critical considerations for future research. Standardizing machine learning (ML) and deep learning (DL) technologies for clinical use in colorectal cancer (CRC) prediction involves addressing several significant challenges. These challenges encompass data quality, algorithm interpretability, integration into clinical workflows, and regulatory compliance, which are crucial for ensuring effective and safe implementation in healthcare settings. While these challenges are significant, they also present opportunities for innovation in CRC prediction. Addressing these issues can lead to more robust, reliable, and clinically applicable AI solutions that enhance patient outcomes.

Consistency in Research Outcomes

- Utilizing standardized datasets allows for the replication of studies and validation of results, which is vital for establishing robust predictive models(Hoogendoorn et al., 2014).
- This consistency aids in the development of decision support systems that can be applied across various healthcare settings.

Conversely, while standardized datasets are beneficial, the reliance on them may limit the exploration of novel data sources or unique patient characteristics that could enhance predictive accuracy. This highlights the need for a balanced approach in data utilization.

7. CONCLUSION

The review presents both machine learning and deep learning models for colorectal cancer histology decomposition on external test sets. All these Machine Learning and Deep Learning approaches enhances the extraction of clinically relevant biomarkers from histologic samples. Finally, we think that scientists who are thinking about using ML and DL techniques to diagnose colorectal cancer will find our study useful. This study provides a thorough analysis of the earlier advancements made by scientists in the prediction of colorectal cancer using both ML and DL algorithms.

REFERENCES

1. Abiiro, G.A.; Alhassan, F.; Alhassan, B.P.; Alhassan, B.P.; Akanbang, B.A. Socio-demographic correlates of public awareness of patient rights and responsibilities in the Sagnarigu Municipality, Ghana. Int. J. Health Promot. Educ. **2020**, 60, 38–48.
2. Al-Jarrah, O.Y.; Yoo, P.D.; Muhaidat, S.; Karagiannidis, G.K.; Taha, K. Efficient machine learning for big data: A review. Big Data Res. **2015**, 2, 87–93
3. Al-Rajab, M.; Lu, J.; Xu, Q. A framework model using multifilter feature selection to enhance colon cancer classification. PLoS ONE **2021**, 16, e0249094
4. Alsanea, N.; Abduljabbar, A.S.; Alhomoud, S.; Ashari, L.H.; Hibbert, D.; Bazarbashi, S. Colorectal cancer in Saudi Arabia: Incidence, survival, demographics and implications for national policies. Ann. Saudi Med. **2015**, 35, 196–202.
5. Alzubi, J.; Nayyar, A.; Kumar, A. Machine learning from theory to algorithms: An overview. In Proceedings of the Journal of Physics: Conference Series; IOP Publishing: Bristol, UK, 2018; Volume 1142, p. 012012
6. Arabia, M.O.H.S. Cancer Facts and Guidelines.
7. Arabia, M.O.H.S. ChronicDisease.
8. Arabia, M.O.H.S. Colorectal Cancer Early Detection.
9. Bae, J.H.; Kim, M.; Lim, J.; Geem, Z.W. Feature selection for colon cancer detection using k-means clustering and modified harmony search algorithm. Mathematics **2021**, 9, 570.
10. Bardhi, O.; Sierra-Sosa, D.; Garcia-Zapirain, B.; Bujanda, L. Deep Learning Models for Colorectal Polyps. Information **2021**, 12, 245.
11. Bhattacharyya, D.K.; Kalita, J.K. Network Anomaly Detection: A Machine Learning Perspective; Crc Press: Boca Raton, FL, USA, 2013
12. Bychkov, D.; Linder, N.; Turkki, R.; Nordling, S.; Kovanen, P.E.; Verrill, C.; Walliander, M.; Lundin, M.; Haglund, C.; Lundin, J. Deep learning based tissue analysis predicts outcome in colorectal cancer. Sci. Rep. **2018**, 8, 3395
13. Cancer Survival Rates. Available
14. Chehade, A.H.; Abdallah, N.; Marion, J.M.; Oueidat, M.; Chauvet, P. Lung and Colon Cancer Classification Using Medical Imaging: A Feature Engineering Approach. Phys. Eng. Sci. Med. **2022**, 45, 729–746.
15. Chen, H.; Zhao, H.; Shen, J.; Zhou, R.; Zhou, Q. Supervised machine learning model for high dimensional gene data in colon cancer detection. In Proceedings of the 2015 IEEE International Congress on Big Data, Santa Clara, CA, USA, 29 October–1 November 2015; pp. 134–141
16. Chicco, D.; Jurman, G. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. BMC Genom. **2020**, 21, 6.
17. Choi, S.J.; Kim, E.S.; Choi, K. Prediction of the histology of colorectal neoplasm in white light colonoscopic images using deep learning algorithms. Sci. Rep. **2021**, 11, 5311.
18. Clinic, M. Digestive Diseases.

19. Collins, T.; Maktabi, M.; Barberio, M.; Bencteux, V.; Jansen-Winkeln, B.; Chalopin, C.; Marescaux, J.; Hostettler, A.; Diana, M.; Gockel, I. Automatic recognition of colon and esophagogastric cancer with machine learning and hyperspectral imaging. Diagnostics **2021**, 11, 1810
20. Davri, A.; Birbas, E.; Kanavos, T.; Ntritsos, G.; Giannakeas, N.; Tzallas, A.T.; Batistatou, A. Deep Learning on Histopathological Images for Colorectal Cancer Diagnosis: A Systematic Review. Diagnostics **2022**, 12, 837