Automated Approaches For Detecting And Classifying Colon Cancer: A Comprehensive Study

2025 2nd International Conference on Trends in Engineering Systems and Technologies (ICTEST) | 979-8-3315-0537-0/25/$31.00 ©2025 IEEE | DOI: 10.1109/ICTEST64710.2025.11042462

Namitha T M
Department of Electronics and Communication Engineering
Noorul Islam Centre for Higher Education
Kumaracoil,Tamil Nadu, India
namitha.t.m@outlook.com

Vinod Kumar R S
Department of Electronics and Communication Engineering
Noorul Islam Centre for Higher Education
Kumaracoil,Tamil Nadu, India
vinod.kumar.rs@outlook.com

Abstract— **Colon cancer is an important global health issue, which is ranked as the third most frequently diagnosed cancer and the second most prevalent reason for cancer mortality. The early recognition and accurate classification of colon cancer are critical for enhancing the outcomes of patients and treatment strategies. This research investigates the automated detection and classification techniques for colon cancer utilizing artificial intelligence, particularly deep learning and machine learning algorithms. It aims to assess the efficacy of these advanced methodologies in analyzing medical images, such as colonoscopy and histopathology, for the identification of cancerous tissues. The study also addresses some challenges, such as optimizing these models for real-world use, avoiding overfitting, and ensuring they function well across different patient groups. By reviewing the latest developments in this area, this research emphasizes the importance of using automated systems in everyday medical practice to improve the early diagnosis of colon cancer, resulting in better care and patient outcomes.**

Keywords— Colon cancer, Medical imaging, Colonoscopy, Deep learning, Machine learning, Histopathology.

I. INTRODUCTION

Cancer is defined as a group of diseases in which the body develops abnormal cells as a result of random mutations. After the development of the cells, they divide and circulate throughout the organs. If they are not treated properly, it can lead to severe death. Cancer affects millions of individuals globally and can take many different forms, such as prostate cancer, lung cancer, colon cancer, breast cancer, etc. [1]. Among the numerous variants of cancer, colon cancer is categorized as the third most frequently diagnosed cancer stated after breast and lung cancer, and it is the second most frequent leading cause of mortality worldwide. Males are more frequently diagnosed with this disease, compared with females. It can occur in the anus, colon, or rectum, and each of these three numerous variants of cancer are determined as colon or colorectal cancer (CRC). Colon cancer starts when unwanted cells in the lining of the colon or rectum multiply rapidly. In some cases, these abnormal cells can break away and spread over the different parts of the body, which can lead to life-threatening conditions. Many people don't show any symptoms at the beginning; later, the changes occurring in the bowel habits, bleeding in the stool, ongoing discomfort in the belly area, weakness, and losing weight lead to the diagnosis of colon cancer.

According to the reports from World Health Organizations (WHO), the number of cases of CRC has significantly risen in the southern and northern zones of the world when compared to the middle zone. Based on the International Agency for Research on Cancer, the regions of India and Africa show the lowest rates of cancer cases [2]. The risk of colon cancer is increased by inflammation in the intestine, such as ulcerative colitis and Crohn's disease. Patients having a family history of colon cancer and colonic polys are at greater risk. After 30 years, the chance of cancer in these patients is estimated between 15% and 20%. The use of tobacco increases the risk of cancer as well as the mortality rates. Consumption of less meat and intake of more fruits and vegetables reduced the risk of cancer occurrences.

Once the patient is diagnosed with colon cancer, the doctor proceeds to the cancer staging process. Similar to other cancers, colon cancer also has five stages, as stated by the National Cancer Institute [3]. In Stage 0, the cancer has just begun to grow and is remaining constrained in the innermost lining of the colon. At stage 1, the cancer reaches the middle layer of the colon, providing 90% of chances of survival. In stage 2, it spreads beyond the colon wall but has not reached the lymph nodes, having 60%-80% chance of survival. Stage 3 is a little more challenging since the disease has reached the lymph nodes and has progressed to other organs. At the final stage, the cancer grows and spreads to a few of the nearby organs, like the liver and lungs. At stages 3 and 4, the chance of survival is approximated at 30% or 40%, or sometimes it can even go down to 10% [4].

The burden of colon cancer is expected to rise in the coming decades, driven by aging populations, changing dietary habits, and increasing obesity rates. This demands for a greater focus on prevention, early detection, and improved treatment strategies worldwide. The advancements in artificial intelligence (AI) and machine learning (ML) have converted the field of medical imaging, enabling more accurate and efficient detection and classification of numerous diseases, including colon cancer. During the past few years, deep learning (DL) models, have shown remarkable success in medical image analysis, offering enhanced accuracy in tumor detection. The focus of this research is to explore the numerous automated techniques developed for colon cancer detection and classification in DL and ML. The aim of the research is to offer a comprehensive overview of the existing methods on automated colon cancer detection and classification, its limitations, and potential future developments in this field. By doing so, it highlights the importance of automated systems in improving the early diagnosis and treatment of colon cancer, thereby improving patient outcomes.

II. RELATED WORKS

A. *Colon Cancer Detection Using Machine Learning*

Tripathi et al. (2023) [5] evaluated the classification of colon cancer tissues utilizing various ML methods, emphasizing the importance of early illness detection to improve survival chances. They utilized a publicly accessible dataset, CRC-VAL-HE-7K, which includes images of nine types of colorectal tissues. For every image block, the Differential-Box-Count method was used to capture features.

979-8-3315-0537-0/25/$31.00 ©2025 IEEE

The dataset was then examined using several classifiers: K-NN, Decision Tree (DT), SVM, Extreme Gradient Boosting (XGBoost), RF, and Gaussian Naïve Bayes (GNB). The XGBoost algorithm, with extracted features, attained the highest classification accuracy of 91.25%, demonstrating enhanced classification accuracy and reduced validation loss. The study is limited by patch preselection representing only one class, reducing accuracy. A hybrid ensemble feature extraction model was presented by Talukder et al. (2022) [6] that integrates deep feature extraction from ML models to effectively determine lung and colon cancer. The framework was specifically trained on the LC25000 colon dataset, comprising colon histopathological images that have been augmented and categorized into two classes, such as benign colon tissues and colon adenocarcinoma. Empirical findings on the LC25000 dataset revealed that the developed model achieved 96.61% accuracy for identifying colon cancer.

Babu et al. (2022) [7] introduced a segmentation technique that operates independently of magnification, utilizing the area of connected components along with the double density dual tree discrete wavelet transform (DWT). The features obtained were further reduced using fuzzy c-means. Utilizing an artificial neural network (ANN) optimized with salp swarm optimization (SSO), images were categorized into normal and abnormal categories. This method was assessed across four different datasets with varying magnifications, showing substantial outcomes compared to existing techniques. The proposed framework achieved accuracy of 98.5% for the IPC datasets, respectively, demonstrating strong correlation for cancer detection and aiding pathologists. However, the study noted a limitation that when cells overlap, identifying the connected components can lead to inaccurate segmentation. In their study, Naeem et al. (2021) [8] examined 55 health and 55 cancer genes acquired from the National Center for Biotechnology Information GenBank for mutation identification, which is crucial in colon cancer diagnosis. When the sequence was first given, the electron-ion interaction pseudopotential (EIIP) numbering system was utilized. Next, the Haar wavelet was used to apply a single-level discrete DWT. The resulting values are applied to KNN and SVM. The results obtained from the model demonstrated 95% accuracy, 94.74% F1 score, and 90.45% Mathews correlation coefficient (MCC) for SVM and 97.5% accuracy, 95.12% MCC score, and 97.44% F1 for KNN.

Ghosh et al. (2021) [9] combined unsupervised and supervised learning techniques to address the complexities of colon cancer data, which is rare and complex. To tackle the issue of limited data and extract the most relevant information, they applied an oversampling method. A correlation-based technique was employed for feature selection, and principal component analysis (PCA) was employed for dimensionality reduction in order to simplify processing and remove unwanted features. The set of features obtained were given to ANN for classifying colon cancer. The findings demonstrated that the suggested methodology attained 98.50% accuracy. Bae et al. (2021) [10] proposed a methodology to categorize the absence or presence of colorectal cancer by analyzing gene information, as genetic mutations are a primary cause of this cancer. The methodology consisted of four key steps: Z-normalizing the actual data through preprocessing, candidate genes were clustered using K-means clustering with one representative gene selected from each cluster, feature selection was performed with a modified harmony search algorithm, and finally the selected gene combinations were applied to a classifications model which was validated through 5-fold cross-validation. This approach attained 94.36% classification accuracy.

Islam et al. (2020) [11] used ensemble approaches, which combine numerous classification techniques for increased performance, to determine colon cancer with substantial accuracy using microarray processed gene expression data. The study employed a public colon cancer gene expression dataset, implementing an adaptive preprocessing procedure involving linear discriminant analysis (LDA) and PCA to manage the high dimensionality of the data. Subsequently, they built an ensemble learning model incorporating KNN, RF, Bayes Generalized Linear Model (GLM), Kernel Support Vector Machines (KSVM), and XGBoost. This approach yielded superior accuracy compared to individual classifiers, achieving an overall accuracy of 91.67%, with precision, recall, and MCC scores of 0.75, 1.00, and 0.85, respectively. Shafi et al. (2020) [12] aimed to examine and predict colon cancer data using ML and a feature selection method based on an RF classifier. They combined "mean decrease accuracy" and "mean decrease Gini" as feature selection methods within the RF classifier to enhance the prediction model’s accuracy. A comparative analysis was conducted between models with and without feature selection. Findings presented that suggested framework with feature selection attained 95.16% accuracy. However, the study depends on a relatively small subset of genes for feature selection, which could potentially limit the model's ability to generalize to broader datasets or different patient populations.

Nirmalakumari et al. (2020) [13] conducted research to identify colon cancer from a dataset of microarray, aiming to help experts differentiate cancer cells from normal cells for early diagnosis and therapies. Initially, the ANOVA method was employed to determine which genes are the best, followed by fuzzy C-means clustering (FCM) and PCA methods to select relevant genes. The features derived from PCA and FCM were then categorized using various models, including regression, discriminant, and heuristic-based classifiers. The study found that the classifier using features from PCA achieved 97.92% of average classification accuracy in distinguishing between colon cancer and normal samples. Salmi and Rustam (2019) [14] presented a prediction method based on simple probabilistic techniques that classifies data for patients with or without colon cancer using the Naïve Bayes (NB) Classifier model as a classification tool. It achieved up to 95.24% classification accuracy, highlighting its potential as an efficient analysis tool. The study noted a limitation that the classifier’s assumption of independence among attributes can decrease the accuracy because some data contain interrelated attributes.

Xie et al. (2019) [15] utilized the dynamic modeling properties of the chameleon algorithm to propose a gene selection algorithm for colon cancers. The algorithm comprised three steps: identifying genes with elevated Fisher function values as potential candidates; employing the chameleon algorithm was used to group gene groups based on Euclidean distance and selecting the significant gene from each cluster to create a gene subset. The chameleon algorithm then used this gene subset to distinguish between colon cancer patients and normal individuals, achieving a clustering accuracy of up to 85.48%. Comparative analysis with other related studies demonstrated the effectiveness of the proposed algorithm in detecting differential genes associated with colon cancers. Sundaram and Santhiyakumari (2019) [16] introduced a computer-aided method utilizing color histograms based on ROI and SVM2 to identify cancer tumors in Wireless Capsule Endoscopy (WCE) images. Their technique involved preprocessing digital WCE images through ROI-based color histograms and filtering, focusing on the significant regions in the colon. Saliency region was estimated based on color and structure contrast within the colon image, facilitating clustering and tumor classification. K-means clustering was employed to detect tumors in the preprocessed digital images. Features were captured using the Spatial gray level dependence matrices (SGLDM) method and classified with the SVM2 classifier, which used selected feature vectors. Additionally, combining these features enhanced the hybrid feature vector for more accurate tumor categorization. Findings from the study demonstrated that this technique accurately detected colon tumors, achieving nearly 95% accuracy when compared to existing algorithms. The study noted the limitation that low-resolution or poorly captured images can hinder accurate tumor detection and classification. Table I shows the summary of existing approaches of colon cancer detection using ML.

TABLE I. SUMMARY OF EXISTING APPROACHES OF COLON CANCER DETECTION USING ML

AUTHORS	MODEL	DATASET	ACCURACY (%)	REMARKS
Tripathi et al. [5]	ML techniques	CRC-VAL-HE-7K	91.25	The preselection of patches represents only one class, which results in reduced accuracy.
Talukder et al. [6]	Hybrid ensemble ML model	LC25000	96.61	Leads to quick medical interventions
Babu et al. [7]	ANN	Colon Histopathological Image	98.5	Images of both normal and malignant structures lead to misclassification.
Naeem et al. [8]	KNN & SVM	National Center for Biotechnology Information GenBank	97.5	Effective mutation identification in colon cancer diagnosis.
Ghosh et al. [9]	ANN	Colon cancer image	98.5	Model enable automatic extraction of high-level features, eliminating the need for handcrafted feature extraction
Bae et al. [10]	K-means clustering with a modified harmony search algorithm	Not specified	94.36	Attained performance in differentiating colorectal cancer from normal patients
Islam et al. [11]	Ensemble Model	Public colon cancer gene expression dataset	91.67	Reduce computational complexity
Shafi et al. [12]	RF	Microarray dataset	95.16	The study depends on a relatively small subset of genes for feature selection, which potentially limits the model's ability.
Nirmalakumari et al. [13]	Heuristic Classifier	Microarray dataset	97.92	Achieved good classification accuracy
Salmi and Rustam [14]	NB Classifier	Colon cancer images	95.24	Attribute independence leads to accuracy reduction.
Xie et al. [15]	Chameleon Algorithm for Gene Selection	Colon cancer dataset	85.48	Effectively detects differential genes
Sundaram and Santhiyakumari [16]	SVM	WCE images	95	Low resolution of captured images hinders accurate tumor detection and classification

B. *Colon Cancer Detection Using Deep Learning*

In their research, Sinha et al. (2024) [17] developed a DL framework that utilized a convolutional neural network (CNN) to identify and categorize colon cancer. A dataset containing histopathological images of the colon was utilized for model training. The dataset was split into two classes. The model, trained using a batch size of 32 across 30 epochs, attained 98.79% of overall accuracy. Kumar et al. (2023) [18] conducted a comparative analysis of DL methods for colon cancer classification, including transfer learning (TL), CNN, recurrent neural networks (RNN), GoogLeNet, and AlexNet. They used a CT colonography dataset for image classification. A median filter was used during the preprocessing stage to eliminate the noise from the input image. The filtered image was subsequently processed with SegNet to identify and segment the affected areas. The results obtained from the classification showed that Google LeNet was the best classifier with 94.16% accuracy, 97.58 sensitivity, and 87.35 specificity.

A median filter was used by Pavan Kumar et al. (2023) [19] to remove noise from an input image of colon cancer. The filtered images from the ImageNet dataset are then segmented using SegNet. Finally, a variety of DL techniques, such as CNN and GoogLeNet were used to classify colon cancer. The results demonstrated that the GoogLeNet classifier was the best in classification, attaining 94.75% accuracy, 97.46% sensitivity, and 84.68% specificity, with the limitation that it requires more computation resources. Kumar et al. (2023) [20] conducted preprocessing using a median filter to remove noises from an input image of colon cancer using the CT colonography dataset. The filtered image was subsequently processed with SegNet to identify and segment the affected areas. Finally, a variety of DL methods, including RNN, CNN, and TL, were used for the categorization of colon cancer. The performance of the model demonstrated TL was the best classifier, attaining 88% accuracy, 82% sensitivity, and 78% specificity for a training dataset comprising 60% of the total data.

Mohamed et al. (2023) [21] presented a reliable colon cancer detection based on a feature selection approach. The framework suggested was split into three steps: CNN models, specifically SqueezeNet, GoogleNet, AlexNet, and ResNet-

50, were used in the first step to extract features from the dataset consisting of lung and colon cancer histopathological images. In the second step, a metaheuristic method was employed to reduce the feature set, utilizing the grasshopper optimization algorithm to find optimal features from the dataset. At last, ML methods were applied, resulting in accurate and successful colon disease diagnosis. In their study, Dermane and Torch (2023) [22] used the CRC-HE-VAL-7K database to classify various tissue types within colon cancer samples. They explored and evaluated three distinct techniques for CNN models. At first, they employed neural network (NN) training from the beginning, enabling the network to study from the provided data. Next, they utilized transfer learning with the VGG19 pretrained framework, known for its exceptional performance in image recognition tasks. Finally, they implemented an ensemble CNN method, combining ResNet50, VGG19, and Inceptionv3 frameworks. The two ensemble methods, combined through averaging and weighted averaging methods, achieved an impressive accuracy of 98%.

Hossain et al. (2022) [23] used CNN and digital pathology images to develop a computer-aided diagnosis system for discriminating between benign colon and adenocarcinomas tissues of the colon. The dataset used in this research was LC25000. A CNN architecture was used to categorize the histopathological slides of adenocarcinomas and benign cells in the colon. The results showed accurate classification with 94% accuracy. The study noted that the proposed model exhibited high levels of overfitting and highlighted the need for optimization of the pretrained models. Hasan et al. (2022) [24] created a system for the detection and classification of colon adenocarcinomas by employing a deep convolutional neural network (DCNN) framework along with various preprocessing techniques applied to digital histopathology images. The use of modern deep learning (MDL) and digital image processing (DIP) methods helped colon histopathologists to overcome the basic problem of separating benign from malignant diseases. The findings from the suggested method indicated that it can analyze cancer tissues with better accuracy. However, the study faced limitations, such as limited dataset, inappropriate hyper-parameter settings in their model, and the presence of noise and artifacts in the images.

Albashish (2022) [25] proposed two ensemble learning methods for classifying histopathology images of colon cancer into various classes. These techniques were based on modifying pretrained CNN models, such as MobileNetV2, DenseNet121, InceptionV3, and VGG16. The decisions made by these models were combined using majority voting and product rule aggregation methods. Using two publicly accessible datasets of colon histopathology images, the suggested model was tested against conventional pretrained models, achieving accuracies of 97.20% and 91.28%. However, a limitation of the study was noted, indicating that the limited number of training samples caused overfitting in classification tasks. Gupta et al. (2022) [26] analyzed the prediction performance of various DL methods to present an overview of cancer diagnosis techniques and therapies. The study carefully examined various DL models to determine the best algorithm. The results from the simulations represented that automated prediction models could efficiently predict the survival of colon cancer patients. Among the models tested, deep autoencoders exhibited the highest performance, achieving an accuracy of 97% and an AUC-ROC of 95%.

In order to examine the imaging data of colon cells, Tasnim et al. (2021) [27] applied a CNN framework. For the image classification in colon cells, CNN utilized models with maxpooling layers, average pooling layers, and MobileNetV2. The dataset consisted of colon cancer images obtained from the Kaggle repository. The model was trained and evaluated across a number of epochs to calculate the learning rate. It was determined that MobileV2 Net model outperformed other models. One problem cited was that, the dataset used in this study was small. Babu et al. (2021) [28] suggested a colon cancer diagnosis method that used TL to automatically capture high-level features from images of colon biopsy. In order to train the Bayesian optimal support vector machine (SVM) classifier, the features were extracted using a pretrained CNN. Various neural networks (NN), such as VGG-16, AlexNet, and Inception V3, were employed to identify the best detection network. The suggested framework was assessed using four datasets, two obtained from Indian hospitals and two from a public colon image dataset. Results demonstrated that Inception V3 was one of the best-performing models, with an accuracy range from 96.5% to 99%. The key limitation was the heterogeneity of histopathological representations of colon cancer, making it challenging to classify an image as benign.

Hamida et al. (2021) [29] focused on using DL architectures to categorize and emphasize regions of colon cancer reviewed on various CNN models. These CNN frameworks were tested and evaluated using the NCT-CRC-HE-100K, CRC-5000, and integrated datasets, along with RESNET attaining 96.77% accuracy. The study also introduced a pixel-wise segmentation method using both U-Net and SegNet models for colon cancer whole slide images (WSIs). These models were evaluated under different training conditions, such as data augmentation and TL, achieving 76.18% and 81.22% accuracy rates, respectively. Britto and Ali (2021) [30] studied CNN models to evaluate imaging data of colon cells and create models for predicting colon cancer. Deep neural networks (DNN) were applied, trained, and evaluated using data received from the Surveillance, Epidemiology, and End Results (SEER) Program. Using image processing methods, specifically CNN, were used to classify colon types as Type 1, Type 2, and Type 3, helping to eliminate additional noise. The suggested method demonstrated a notable increase in accuracy using DL algorithms when compared to traditional ML algorithms. Experimental results showed that the suggested CNN-based colon segmentation method beat Random Forest (RF) and K-Nearest Neighbors (KNN) at accuracies of 87%, 83%, and 85%, respectively.

Masud et al. (2021) [31] focused on improving the colon cancer detection, utilizing MDL and DIP methods, they developed a categorization framework to distinguish between five numerous variants of colon tissues by examining histopathological images from the LC25000 dataset. Their findings demonstrated that the suggested framework could accurately determine cancer tissues, achieving 96.33% of maximum accuracy. Qasim et al. (2020) [32] suggested a CNN framework for diagnosing colon adenocarcinoma. The images in the histopathological dataset were split into colon adenocarcinoma and benign colon tissues. The framework has two paths; each path generating 256 feature maps to enhance the sensitivity and accuracy of classification. A VGG16 model was additionally trained on the dataset to evaluate its performance against other models. The experimental results demonstrated that VGG16, attaining an accuracy of 96.2%

and 99.6%. However, the study had one drawback, as it focuses only on colon adenocarcinoma and is not trained on other forms of colorectal cancer.

To identify histopathological images of colon cancer, Liang et al. (2020) [33] presented a multi-scale feature fusion convolutional neural network (MFF-CNN) utilizing the Shearlet transform. The framework captured Shearlet coefficients from the histopathological image across various decomposition scales as supplementary features, which is then integrated with the real pathological image and given to the network. After fusion and feature learning, the MFF-CNN was able to identify colorectal adenocarcinoma epithelium (TUM) with 96% identification accuracy and a 0.9594 average F-1 score for normal colon mucosa (NORM). Moreover, the false negative rate decreased to 5.5%, while false positive rate dropped to 2.5%, respectively. The prime aim of the research by Bukhari et al. (2020) [34] was to observe the usage of DL for the histopathological diagnosis of colonic cancer by examining digitized pathology images. Two datasets were used to generate the images of colonic cancer and non-neoplastic colonic tissue: the first dataset consist of 10,000 images was used to train and validate the CNN network, while the second dataset used 40% of the images for training and 60% of the image for testing. Three variants of CNN utilized to examine the images, with ResNet-50 achieving the highest accuracy of 93.91%. However, a limitation of the study that it did not mark the segmented portion of the lesion, which may have provided additional support to the pathologist. Table II shows the Summary of existing approaches of colon cancer detection using DL. Figure 1 shows the comparison graph for ML and DL models.

TABLE II. SUMMARY OF EXISTING APPROACHES OF COLON CANCER DETECTION USING DL

AUTHORS	MODEL	DATASET	ACCURACY (%)	REMARKS
Sinha et al. [17]	CNN	Colon cancer histopathological images	98.79	Assist health care professionals in making timely decision
Kumar et al. [18]	Different DL techniques	CT colonography	94.16	Model has difficulty in determining non-cancerous cases
Pavan Kumar et al. [19]	CNN & GoogleNet	ImageNet	94.75	Required more computational resources
Kumar et al. [20]	CNN, RNN & TL	CT colonography	88	TL showed better performance
Mohamed et al. [21]	CNN	Lung and colon cancer histopathological images	-	CNNs extract large number of features, still pose challenges in terms of memory and processing time before optimization
Dermane and Torch [22]	VGG19, Inception V3 & ResNet 50	CRC-HE-VAL-7K database	98	Neural network training from scratch needs more computational resources and time
Hossain et al. [23]	CNN	LC25000	94	The model exhibited high levels of overfitting and highlighted the need for optimization of the pretrained models
Hasan et al. [24]	DCNN	Digital histopathology images	96.50	Limited number of input data and inappropriate hyper-parameter settings
Albashish [25]	Pretrained CNN models	Colon cancer histopathology images	97.20	Limited number of training samples caused overfitting in classification
Gupta et al. [26]	Deep Autoencoders	SEER	97	Model captured complex features in the data
Tasnim et al. [27]	CNN	Colon cancer images	-	Limited dataset
Babu et al. [28]	Pretrained CNN	Colon biopsy images	96	Difficult to classify an image as benign
Hamida et al. [29]	DL Architectures	AiCOLO, NCT-CRC-HE-100K, CRC-5000, WARWICK	96.77	Limited to offer transferable pathological data representations.
Britto and Ali [30]	DNN	Not specified	87	Better classification for colon segmentation
Masud et al. [31]	DL techniques	LC25000	96.33	Identified cancer tissues
Qasim et al. [32]	VGG16	Histopathological dataset	96.20	The model was not trained on all types of colorectal cancer.
Liang et al. [33]	MFF-CNN	Histopathological images	96.00	The false positive and false negative rates minimize diagnostic errors
Bukhari et al. [34]	Variants of CNN	Digitized pathology images	93.91	The model failed to define the segmented part of the lesion

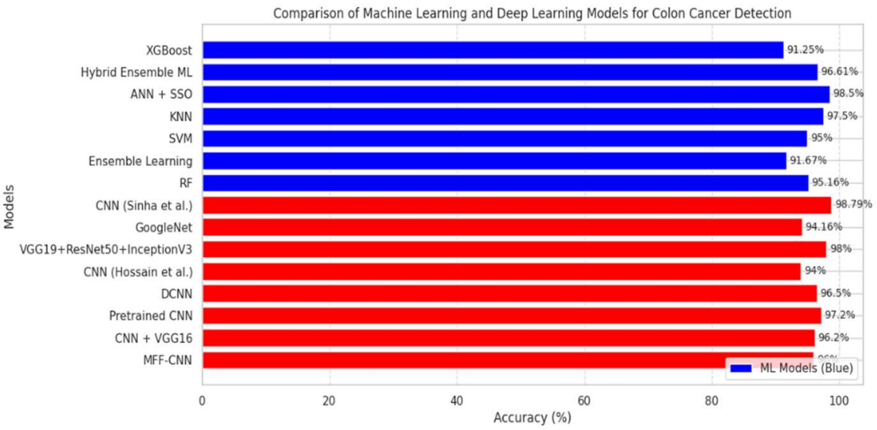

Fig. 1. Comparison graph for ML and DL models

III. RESEARCH GAP

Recent studies on colon cancer classification and detection have made significant advancements in accuracy and model performance across various datasets and methods. Despite these advancements, there remains a notable gap in the thorough optimization and validation of these models for practical clinical use. While various algorithms and models have demonstrated high accuracy in detecting and classifying colon cancer, many studies are limited by small sample sizes and imbalanced class distributions, leading to concerns about the generalizability and fairness of their findings across diverse populations and clinical settings. Models trained on imbalanced datasets are prone to bias toward the majority class, potentially reducing sensitivity to minority (often malignant) cases, critical in medical diagnostics. Many methods also face challenges with overfitting; as a result, there is a need for robust methods that can effectively manage variations in data quality, class representation, and input types to ensure reliable performance in real-world scenarios. Additionally, while researchers have emphasized feature extraction and selection, there remains a need for more integrative approaches that combine multiple data modalities, such as genetic data, medical images, and clinical records. This multi-modal integration could enhance diagnostic accuracy and clinical applicability. Therefore, further research is required to develop automated systems that are not only precise and generalizable but also capable of handling data imbalance, ultimately improving early detection and treatment outcomes for colon cancer patients.

IV. CONCLUSION

Colon cancer is one of the most prevalent causes of cancer-related deaths globally, making early detection and treatment important for improving patient outcomes. Automated detection and classification of colon cancer were important in helping doctors find and treat the disease early. The number of colon cancer cases is influenced by various factors, including genetics, lifestyle choices, and other health issues. This study emphasizes the need for better ways to diagnose colon cancer accurately. It highlights the benefits of using advanced technologies like ML and DL to improve how doctors detect cancer in medical images. By using these modern methods, healthcare providers can create more personalized treatment plans for patients. Integrating these technologies into everyday medical practice will help doctors identify colon cancer sooner and more accurately, leading to better care and outcomes for patients. This proactive approach is crucial for improving patient care and reducing the impact of colon cancer in the community.

REFERENCES

[1] Britto, C. F., & Ali, A. R. H. (2021). Performance analysis of Colon cancer using Neural Networks. EFFLATOUNIA-Multidisciplinary Journal, 5(3), 83-90.

[2] Alrushaid, N., Khan, F. A., Al-Suhaimi, E., & Elaissari, A. (2023). Progress and Perspectives in Colon Cancer Pathology, Diagnosis, and Treatments. Diseases, 11(4), 148.

[3] National Cancer Institute, "Stages of Colon Cancer," http://www.cancer.gov/cancertopics/pdq/treatment/colon/Patient/page2, June 2013.

[4] Ulanja, M. B., Rishi, M., Beutler, B. D., Sharma, M., Patterson, D. R., Gullapalli, N., & Ambika, S. (2019). Colon cancer sidedness, presentation, and survival at different stages. Journal of oncology, 2019(1), 4315032.

[5] Tripathi, A., Misra, A., Kumar, K., & Chaurasia, B. K. (2023, March). Colon cancer tissue classification using ml. In 2023 6th International Conference on Information Systems and Computer Networks (ISCON) (pp. 1-6). IEEE.

[6] Talukder, M. A., Islam, M. M., Uddin, M. A., Akhter, A., Hasan, K. F., & Moni, M. A. (2022). Machine learning-based lung and colon cancer detection using deep feature extraction and ensemble learning. Expert Systems with Applications, 205, 117695.

[7] Babu, T., Singh, T., Gupta, D., & Hameed, S. (2022). Optimized cancer detection on various magnified histopathological colon imagesbased on dwt features and fcm clustering. Turkish Journal of Electrical Engineering and Computer Sciences, 30(1), 1-17.

[8] Naeem, S. M., Mabrouk, M. S., Eldosoky, M. A., & Sayed, A. Y. (2021). Automated detection of colon cancer using genomic signal processing. Egyptian Journal of Medical Human Genetics, 22, 1-8.

[9] Babu, T., Singh, T., Gupta, D., & Hameed, S. (2021). Colon cancer prediction on histological images using deep learning features and Bayesian optimized SVM. Journal of Intelligent & Fuzzy Systems, 41(5), 5275-5286.

[10] Bae, J. H., Kim, M., Lim, J. S., & Geem, Z. W. (2021). Feature selection for colon cancer detection using k-means clustering and modified harmony search algorithm. Mathematics, 9(5), 570.

[11] Islam, A., Rahman, M. M., Ahmed, E., Arafat, F., & Rabby, M. F. (2020, January). Adaptive feature selection and classification of colon cancer from gene expression data: an ensemble learning approach. In Proceedings of the international conference on computing advancements (pp. 1-7).

[12] Shafi, A. S. M., Molla, M. I., Jui, J. J., & Rahman, M. M. (2020). Detection of colon cancer based on microarray dataset using machine learning as a feature selection and classification techniques. SN Applied Sciences, 2, 1-8.

[13] Nirmalakumari, K., Rajaguru, H., & Rajkumar, P. (2020). Performance analysis of classifiers for colon cancer detection from dimensionality reduced microarray gene data. International Journal of Imaging Systems and Technology, 30(4), 1012-1032.

[14] Salmi, N., & Rustam, Z. (2019, June). Naïve Bayes classifier models for predicting the colon cancer. In IOP conference series: materials science and engineering (Vol. 546, No. 5, p. 052068). IOP Publishing.

[15] Xie, J., Wang, Y., & Wu, Z. (2019). Colon cancer data analysis by chameleon algorithm. Health Information Science and Systems, 7, 1-8.

[16] Shanmuga Sundaram, P., & Santhiyakumari, N. (2019). An enhancement of computer aided approach for colon cancer detection in WCE images using ROI based color histogram and SVM2. Journal of medical systems, 43(2), 29.

[17] ubrata Sinha, S. (2024). A Machine Learning Approach for Detection and Classification of Colon Cancer using Convolutional Neural Network Architecture. J. Electrical Systems, 20(7s), 1065-1071.

[18] Kumar, V. R. P., Arulselvi, M., & Sastry, K. B. S. (2023). Comparative assessment of colon cancer classification using diverse deep learning approaches. Journal of Data Science and Intelligent Systems, 1(2), 128-135.

[19] Kumar, V. R. P., Arulselvi, M., & Sastry, K. B. S. (2023, March). Colon Cancer Classification using Google Net. In 2023 10th International Conference on Computing for Sustainable Global Development (INDIACom) (pp. 1657-1661). IEEE.

[20] Kumar, V. R. P., Arulselvi, M., & Sastry, K. B. S. CNN, RNN and Transfer Learning for Colon Cancer Classification.

[21] Mohamed, A. A. A., Hançerlioğullari, A., Rahebi, J., Ray, M. K., & Roy, S. (2023). Colon disease diagnosis with convolutional neural network and grasshopper optimization algorithm. Diagnostics, 13(10), 1728.

[22] DERMANE, K., & TORCH, F. (2023). Early Detection of Colon Cancer on Histopathology images (Doctoral dissertation, Ibn Khaldoun University).

[23] Hossain, M., Haque, S. S., Ahmed, H., Mahdi, H. A., & Aich, A. (2022). Early stage detection and classification of colon cancer using deep learning and explainable AI on histopathological images (Doctoral dissertation, Brac University).

[24] Hasan, M. I., Ali, M. S., Rahman, M. H., & Islam, M. K. (2022). Automated detection and characterization of colon cancer with deep convolutional neural networks. Journal of Healthcare Engineering, 2022.

[25] Albashish, D. (2022). Ensemble of adapted convolutional neural networks (CNN) methods for classifying colon histopathological images. PeerJ Computer Science, 8, e1031.

[26] Gupta, Surbhi, S. Kalaivani, Archana Rajasundaram, Gaurav Kumar Ameta, Ahmed Kareem Oleiwi, and Betty Nokobi Dugbakie. "[Retracted] Prediction Performance of Deep Learning for Colon Cancer Survival Prediction on SEER Data." BioMed Research International 2022, no. 1 (2022): 1467070.

[27] Tasnim, Z., Chakraborty, S., Shamrat, F. J. M., Chowdhury, A. N., Nuha, H. A., Karim, A., ... & Billah, M. M. (2021). Deep learning predictive model for colon cancer patient using CNN-based classification. International Journal of Advanced Computer Science and Applications, 12(8), 687-696.

[28] Babu, T., Singh, T., Gupta, D., & Hameed, S. (2021). Colon cancer prediction on histological images using deep learning features and Bayesian optimized SVM. Journal of Intelligent & Fuzzy Systems, 41(5), 5275-5286.

[29] Hamida, A. B., Devanne, M., Weber, J., Truntzer, C., Derangère, V., Ghiringhelli, F., ... & Wemmert, C. (2021). Deep learning for colon cancer histopathological images analysis. Computers in Biology and Medicine, 136, 104730.

[30] Britto, C. F., & Ali, A. R. H. (2021). Performance analysis of Colon cancer using Neural Networks. EFFLATOUNIA-Multidisciplinary Journal, 5(3), 83-90.

[31] Masud, M., Sikder, N., Nahid, A. A., Bairagi, A. K., & AlZain, M. A. (2021). A machine learning approach to diagnosing lung and colon cancer using a deep learning-based classification framework. Sensors, 21(3), 748.

[32] Qasim, Y., Al-Sameai, H., Ali, O., & Hassan, A. (2020, December). Convolutional neural networks for automatic detection of colon adenocarcinoma based on histopathological images. In International Conference of Reliable Information and Communication Technology (pp. 19-28). Cham: Springer International Publishing.

[33] Liang, M., Ren, Z., Yang, J., Feng, W., & Li, B. (2020). Identification of colon cancer using multi-scale feature fusion convolutional neural network based on shearlet transform. IEEE Access, 8, 208969-208977.

[34] Bukhari, S. U. K., Syed, A., Bokhari, S. K. A., Hussain, S. S., Armaghan, S. U., & Shah, S. S. H. (2020). The histological diagnosis of colonic adenocarcinoma by applying partial self-supervised learning. MedRxiv, 2020-08.