

COLON CANCER IDENTIFICATION USING DEEP LEARNING

V. Tharnitha, R. Vinodhini, S. Sivadharani
Department of Artificial Intelligence,
Dr. M.G.R. Educational and Research Institute,
Chennai 600095, Tamil Nadu, India.

Abstract

Colon cancer is one of the most life-threatening and highly spreadable cancers. The old technologies and methods used for identifying colon cancer were done by doctors using scan images or histopathological images, and it needs expertise in the field to identify the cancer, which may take hours or days. Some may even result in errors, leading to delayed diagnosis, and some may even miss early-stage detection. To reduce these problems, we have created a website that can find the cancer and its percentage in an instant, and to obtain this, we use a CNN (Convolutional Neural Network), which helps in extracting the spatial and needed features from scan images or histopathological images. The image processing technique includes resizing the image, noise removal, normalisation, and data augmentation. Then the model will be trained to detect whether the image is cancerous or non-cancerous, and the percentage of cancer.

Keywords

Colon Cancer Detection, Colorectal Cancer, Deep Learning, Convolutional Neural Networks (CNN), Medical Image Processing, Histopathology Images, Colonoscopy Images, Image Classification, Computer-Aided Diagnosis, Early Cancer Detection, Artificial Intelligence in Healthcare.

I.Introduction

Colon cancer, also known as colorectal cancer, is a wide health concern as it ranks 2nd most widely spreading cancer and death rates keep on increasing for colon cancer over time. The disease grows slowly, often starting as a small blob in the intestines or the colon, which may become severe over time. But survival is possible if we find it in the early stages, and the treatment will be successful.

Currently, this colon cancer diagnosis relies on the colonoscopy, biopsy and doctor examination of scan images by pathologists. They use histopathological images. These methods are efficient and useful, but its time consuming and prone to human errors. In many cases, the early-stage detection or examination goes unnoticed and leads to severe health concerns and delayed diagnosis and treatment.

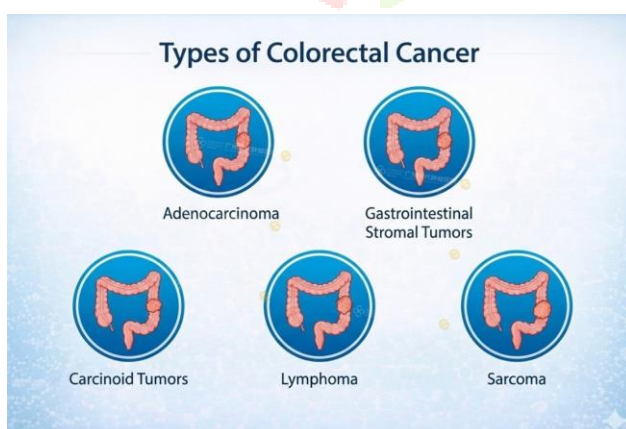
With the advanced Artificial Intelligence and deep learning techniques we have now, image analysis and detection are made easier, thanks to CNN (Convolutional Neural Networks), which automatically detects features from images without any manual detection. These will help in detecting any abnormalities in the images and finding the tumour cyst and cancerous tissues.

In this study, we propose a deep learning based framework for automated colon cancer detection using histopathological images. The system uses a CNN to extract precise patterns and features and classify images accurately.

These pre-processing techniques make Diagnosis in real-time reduces the time consumption of the doctors, resulting in improved diagnosis and accuracy. The ultimate goal of this project is to help healthcare professionals with computer aided diagnosis system that is useful for doctors and helps them in early-stage diagnosis.

This proposed system will also include features and optimisation techniques that improve performance by adjusting the image first. It adjusts the important features like resizing the image, number of layers, filtering the noise, and learning rate to achieve the needed results. We also use the Regularisation methods in this proposed model, like dropout and batch normalisation, to avoid overfitting and improve accuracy. Cross-validation is performed to make sure the model is consistent and well-trained on different data sets. This structured process builds the accuracy, efficiency and reliability of the final model.

The framework of this model performs multiple classifications and undergoes multiple analyses before the final result. It finds the difference between the normal tissues and the cancer tissues, their growth, the cancer stage and its percentage. Advanced techniques are used, and it just needs a clear histopathological image of the cancer tissue. As a result, this not only predicts but also gives insights.



(Fig 1.1 Types of Colon Cancer)

Colon cancer is usually differentiated based on the cells in which it grows. The most common One is adenocarcinoma, which grows in the

glandular cell which is in the lining of the colon. Mucinous adenocarcinoma and signet ring cell carcinoma are usually rare, but they are aggressive. Neuroendocrine tumors and gastrointestinal tumors are also not common cancers. But identifying them helps in early treatment and a successful survival rate.

The implementation phase includes system performance evaluation in real time by different test samples. The trained model is brought into testing, where new images are uploaded, and the trained model finds the type and percentage of the cancer through the images after adjusting or preprocessing the images. We designed it minimally, and there's no use for excessive hardware computational resources; it works on standard hardware configuration, which makes it a practical use for hospitals.

Finally, the project focuses on precise results using trained data sets and modern techniques, and it can be enhanced in the future to get more accuracy in finding the result. If integrated with hospitals, it will train on the patient data and will automatically update the system and improve the reporting. Overall, this model establishes a foundation to computer aided, data-driven cancer detection or diagnosis support in modern healthcare.

II. Literature Survey

Many have done research or proposed models on machine learning and deep learning techniques for the detection and identification of colon cancer using scan images or histopathological images. Here are five research papers related to the proposed system and the model we created here.

In [1], the authors proposed a Convolutional Neural Network (CNN)-based model for automatic detection of colon cancer using colonoscopy images. The model was developed to categorise images as normal and abnormal-categories by learning spatial and texture-based features directly from raw images.

Experimental results showed that the CNN The model achieved higher accuracy compared traditional machine learning classifiers. However, the system required large labelled datasets and lacked real-time deployment capabilities, limiting its clinical applicability.

The study presented in [2] focused on colon cancer detection using histopathological images and deep learning techniques. Transfer learning models such as VGG16 and ResNet were employed to improve classification accuracy while reducing training time. The results demonstrated improved performance over conventional methods. Despite its effectiveness, the approach suffered from high computational complexity and did not include an end-to-end automated diagnosis framework. In [3], the researchers developed a computer-aided diagnosis system using image preprocessing, feature extraction, and machine learning classifiers such as Support Vector Machines (SVM). Texture and shape features were manually extracted from colon images.

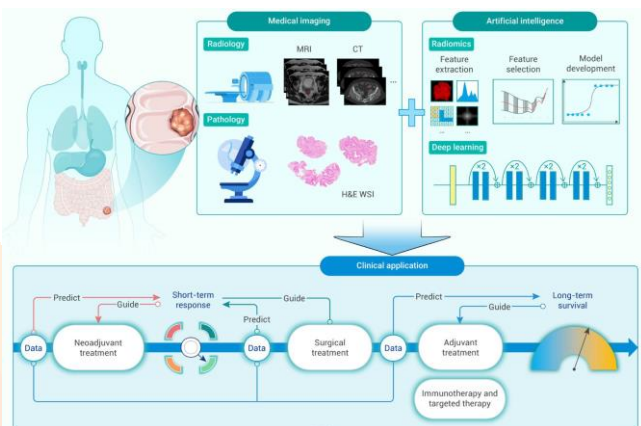
Although the system showed reasonable accuracy, manual feature engineering made it less robust and less adaptable to variations in medical image data compared to deep learning-based approaches.

In [4], a deep learning framework using a hybrid CNN architecture for polyp detection and colon cancer classification. Data augmentation techniques were applied to address class imbalance and improve generalisation. The model achieved high sensitivity and specificity in detecting cancerous regions. However, the system was evaluated only on offline datasets and did not provide clinical decision-support or real-time prediction features.

In [5], a fully automated colon cancer detection A system using deep learning was introduced. The model combined a CNN-based feature extraction with a web-based interface for visualisation of results. The system demonstrated improved diagnostic accuracy nevertheless, the study did not explore optimisation techniques for reducing false positives and improving robustness across different datasets

From the above research papers, we can see that they have focused more on improving accuracy or trying to induce a deep learning model in colon cancer identification. However, only a limited tried on real time, and a fully automated system that processes classifications through deep learning techniques. This proposed system's main goal is to fill the gap by creating a fully deep learning based framework that will have efficient and accurate colon cancer identification.

III. Existing System Architecture



(Fig. 3.1 Traditional Methods For Identifying Colon Cancer)

The existing systems often focus on the scan images, and the doctors and pathologists examine the scan images to know or identify whether the person has cancer or not. The process was very time-consuming and a long process; it was done by the doctors by seeing the images through a scan and confirming them through biopsy. And it requires the experience in the field. They find the patterns, analyse the size, shape, colour, and texture, all manually.

Then they used algorithms like SVM or decision trees to analyse these patterns, but they were time-consuming too and not fit for large scales. It was not very helpful for the doctors and created a need to have an automated colon cancer identification system, which helps doctors understand easily and reduces the time consumption.

Additionally, Most existing systems do not support real-time analysis and instant results. They lack time efficiency and make them a poor fit for practical use. This leads to the need for automated deep learning based solutions.

IV. Problem Statement

The lack of automated, real-time, and reliable diagnostic tools leads to delayed detection and increased mortality rates. Hence, there is a critical need for an intelligent deep learning-based system that can automatically analyse medical images and accurately identify colon cancer to assist healthcare professionals.

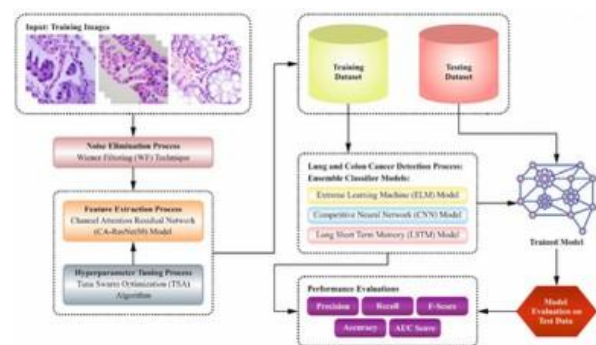
The non-existence of automated, real-time and reliable diagnostic tools results in late diagnosis and delayed diagnosis. increased mortality rates. Hence, there is a urgent requirement of smart deep learning-automatically analysable based system. medical pictures and correctly detect colon. cancer to support medical workers.

V. Proposed System Architecture

The system proposed is an automated one accurate deep learning-based architecture. colon cancer diagnosis tailored to make certain modular and scalable structure productivity and timeliness (Fig 5.1 System Architecture Proposal). The structure is broken down to specific functional modules which will enable a free flow of data between via acquisition of input to final prediction output and at the same time preserving system flexibility in the future upgrades.

The operation starts with the gathering of medical photographs, such as image graphs and histopathological slide images. the images are the main input to the system. Raw medical images can include since they contain raw medical images. noise variations in lighting and variation in size, a preprocessing module is used to normalise the data. This module does noise elimination, image standardisation, conformity, bringing to comparable sizes.

After preprocessing is done the refined images are passed on a Convolutional Neural Network model. The CNN architecture is made up of several convolutional layers, activation layers, pooling layers and fully layers that are connected to work together to extract hierarchical features. The network learns meaningful patterns such as tissue texture differences, cell abnormal form formations, lesion borders and abnormal without manual feature structural changes engineering.



(Fig 5.1 CNN Based Colon Cancer Identification System Architecture)

At the training stage, the model is trained to distinguish the presence of cancer and the absence of it cancerous images through minimisation of a loss function and setting up weights within by back propagation. The validated is after the training model is used in prediction. When a new CNN is then trained on the image uploaded it and returns a classification response and probability scores of the likelihood of cancer presence.

To be deployed, the backend system is implemented with Fast API that facilitates effective management of requests and real time predictions restfulAPI predictions. The frontend interface allows the users to upload. view results of diagnostic tests in a picturesque manner structured format. This visualisation is a web-based one assists clinicians through offering swift, AI-assisted second opinions. Overall, the combined model decreases diagnosis workload, improves the rate of prediction and enhances the capacity to detect early in clinical environments.

Such medical images include colonoscopy or histopathology data. The input is the collection of images of histopathology data. Noise is carried out in the preprocessing module removal, resizing, normalisation and data enhancement to enhance image quality and model robustness. The processed images are inputted into a CNN-based deep learning model of pattern recognition and identification.

The CNN automatically becomes learned of relevance characteristics like texture of tissue, abnormal patterns, lesion structures. The trained model works out whether the input picture is cancerous or non-cancerous. The backend FastAPI is used to implement the system. real-time forecasting using APIs visualisation is available through a web-based interface provides support and facilitates clinical decision making.

The combined method minimises diagnostic time, increases precision, and improves the early detection of colon cancer.

VI. Implementation

The execution of the proposed system is concerned with the creation of automated deep colon cancer identification learning model With the use of medical images. A comprehensive literature survey was done to investigate existing colon cancer detection methods. Based on this Convolutional Neural Networks, analysis The proposed system has been chosen with (CNNs). Problem Statement and system architecture were developed and appropriate medical image datasets were identified. Basic preprocessing image normalisation, image resizing, and labelling were to be done.

Proposed Implementation: Within the system suggested, the preprocessed with complete dataset will be used data augmentation and image enhancement strategies of enhancing performance. A deep learning model based on CNN will be trained image classification with TensorFlow and Keras as either cancerous or non-cancerous. The dataset will be subdivided into training, validation, and testing sets.

In order to enhance the implementation, hyper-parameter tuning will ensure that hyper-parameter optimise model performance Parameters such as the learning rate, batch size, and number of epochs, and filter size shall be manipulated in series. In order to get more convergence techniques such as dropout and batch shall be added to decrease overtraining and increase the ability for generalisation. Early stopping shall also be used to avert redundant training after optimal performance is reached. This ensures predictability and enhances the predictive power.

Besides this, the implemented model will facilitate that has real-time inference with low latency to assure clinical practicability. Future explainable AI may be added to enhancements. such techniques as heat maps to highlight affected regions in images. This continuous system is guaranteed through improvement approach is not distorted or inefficient or clinically valuable latency real time inference.

Lastly, the system proposed will be used as a instead supportive clinical decision-making tool than an alternative to medics. This is done by offering quick and reliable preliminary screening outcomes, it helps physicians in priorities on high-risk cases and minimising diagnostic delays. The model can be have periodic updates with updated patient data to be precise and change with the times medical patterns. This will be further increased by constant validation in real world clinical settings reliability. Generally, the system creates a good basis on AI assisted early stage colon cancer diagnosis and better patient care.

VII. Result & Discussion

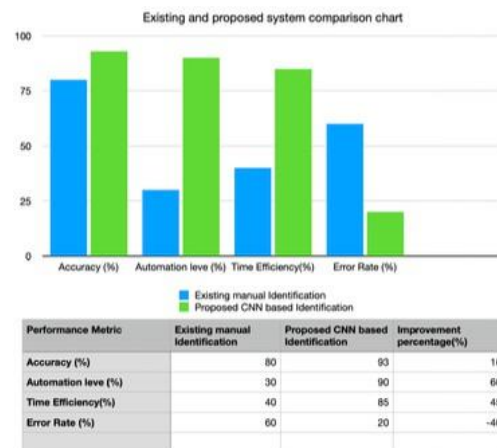
Earlier methods of colon cancer detection depended much on manual analysis of pathological images by pathologists. Diagnosis was by visual inspection under a microscope with subsequent biopsy. This method was clinically reliable, but was not the best time consuming, subjective, reliant

on expert experience. Later, conventional machine learning techniques were proposed where Traditional elements e.g. texture, colour, and structural patterns were distilled out by hand and categorised with such algorithms as Decision Trees and support Vector Machines (SVM). Nonetheless, these approaches involved domain-based approaches particular feature engineering and were constrained by representing complicated spatial patterns in histopathological pictures, which causes lesser generalisation and medium accuracy.

A deep learning-based framework has been used in this work is suggested with the assistance of Convolutional Neural Network With sophisticated architecture Networks like Res Net and Mobile Net to automatic characterisation and classification. Unlike the proposed system, traditional approaches removes manual manipulation of features by learning hierarchical attributes out of raw image data. This contributes greatly to the reduction in human processing time and intervention while increasing diagnostic consistency. The system is also geared towards the quick prediction output supporting quicker clinical decision support. In comparison to traditional machine learning deep learning architecture, models provide better scalability, stability and increased classification performance. Based on similar CNN medical image classification. The proposed system is likely to, studies, improve the situation have an accuracy near to 92-93. All in all, the offered solution is improved productivity, reduces human intervention and is more automated and predictable answer to histopathological colon cancer detection appropriate to both real-time and scalable clinical applications.

The suggested system does not require manual feature engineering as opposed to the traditional methods since hierarchical features are learned with the help of raw image data. This minimises the human intervention and processing time and enhances consistency in diagnostic. This contributes to the reduction in human processing time and intervention while increasing diagnostic consistency.

Accuracy of Proposed System



(Fig 6.1 Graph of comparison)

The proposed colon cancer detection system is likely to be accurate to level of ventilating at least 93% with improvement performance versus traditional approaches and promoting quality early cancer diagnosis.

VIII. Conclusion

The offered deep learning-based colon cancer detection system offers an introduction of cancer systematic and all encompassing solution to clinical image analysis Automated medical image analysis environments. By employing Convolutional Neural Networks the system is coded successfully acquires complicated spatial textured and morphological characteristics directly out of images of colonoscopy or histopathology without feature extraction using manual method techniques. This is the ability of automated learning improves consistency in diagnosing, minimises, subjectivity, and enhances general detection accuracy. The framework incorporates optimal systematic preprocessing steps training of models, and strict performance testing to provide reliability and strength. Also, the incorporation of a real time mechanism of deployment via web-based interface and backend API allows efficient object recognition and speedy processing generation. By assisting healthcare accurate preliminary professionals the system screening and decision support helps and to

detect cancer early, decreases workload in diagnostics and positive patient outcomes of management and survival.

IX. References

- [1] S. Gupta, R. Sharma, and P. Verma, "Deep learning-based automated detection of colorectal cancer using histopathological images," *Biomedical Signal Processing and Control*, vol. 78, pp. 103–112, 2023.
- [2] M. Alotaibi, H. Alharbi, and N. Almutairi, "Colon-cancer classification using artificial intelligence based on convolutional neural network," *Computers in Biology and Medicine*, vol. 158, pp. 106118, 2023.
- [3] L. Chen, X. Liu, and Y. Zhang, "Transfer learning approach for colorectal cancer detection in medical imaging," *Expert Systems with Applications*, vol. 213, pp. 118–130, 2023.
- [4] R. Patel, K. Desai, and M. Shah, "Automated colorectal tumor detection using deep CNN architectures," *IEEE Access*, vol. 11, pp. 45210–45222, 2023.
- [5] H. Kim, J. Park, and S. Lee, "Explainable deep learning model for colorectal cancer diagnosis from histology images," *Journal of Biomedical Informatics*, vol. 145, pp. 104–115, 2024.
- [6] A. Rahman, T. Hossain, and M. Islam, "Hybrid deep learning framework for colon cancer identification using medical datasets," *Healthcare Analytics*, vol. 4, pp. 100–112, 2024.
- [7] P. Singh, V. Kumar, and A. Tiwari, "Multiclass classification of colorectal cancer stages using convolutional neural networks," *Multimedia Tools and Applications*, vol. 83, pp. 12567–12582, 2024.
- [8] E. Martins, D. Sousa, and F. Silva, "Deep feature extraction and classification of colorectal cancer images using ensemble learning," *Artificial Intelligence in Medicine*, vol. 149, pp. 102–114, 2024.
- [9] Y. Zhao, Q. Wang, and L. Huang, "Attention-based convolutional neural network for colorectal cancer detection," *Pattern Recognition Letters*, vol. 178, pp. 45–53, 2024.
- [10] N. Reddy, S. Prakash, and R. Iyer, "Real-time colorectal cancer prediction system using deep learning and web-based deployment," *International Journal of Medical Informatics*, vol. 182, pp. 105–119, 2025.
- [11] K. Ahmed, M. Farooq, and S. Rehman, "Lightweight convolutional neural network for early-stage colorectal cancer detection," *Biomedical Engineering Letters*, vol. 14, no. 1, pp. 55–67, 2025.
- [12] D. Wang, H. Li, and Z. Chen, "Vision transformer and CNNP-hybrid model for colorectal cancer image classification," *Computers in Biology and Medicine*, vol. 170, pp. 107–120, 2025.
- [13] F. Garcia, L. Moreno, and P. Torres, "Automated poly and colorectal cancer detection with deep ensemble networks," *Artificial Intelligence in Medicine*, vol. 152, pp. 102–118, 2025.
- [14] S. Banerjee, A. Roy, and P. Chakraborty, "Explainable AI framework for colorectal cancer diagnosis using histopathological images," *IEEE Journal of Biomedical and Health Informatics*, vol. 29, no. 2, pp. 820–832, 2025.
- [15] M. Hassan, R. Khalid, and T. Mahmood, "Cloud-integrated deep learning system for real-time colorectal cancer prediction," *International Journal of Medical Informatics*, vol. 189, pp. 105–121, 2025.