



# DEEP LEARNING-BASED COLON CANCER CLASSIFICATION USING PRE-TRAINED CUSTOM CONVOLUTIONAL NEURAL NETWORK WITH HISTOPATHOLOGICAL IMAGES

<sup>1</sup>N Sakthipriya, <sup>2</sup>V Govindasamy, <sup>3</sup>R Narmadhadevi, <sup>4</sup>L Prasanth, <sup>5</sup>P Kanimozhi

<sup>1</sup>Research Scholar, <sup>2</sup>Associate Professor, <sup>3,4,5</sup>B.Tech Student

<sup>1</sup>Department of Information Technology,

<sup>1</sup>Puducherry Technological University, Puducherry, India

**Abstract:** Early detection and accurate staging are critical to successful treatment of colon cancer, which is one of the top causes of cancer-related mortality globally. In the existing system, statistical features are used to classify cancer using supervised learning methods. However, recent advancements in deep learning have shown significant promise for improving classification accuracy. Research is being carried out to compare custom convolutional neural networks (CNNs) with pre-trained CNN models. We evaluated various deep learning models to identify the best model for classifying colon cancer. A publicly available dataset of histopathological images of colon cancer will be used to train and validate this custom CNN model. Performance metrics such as accuracy, sensitivity, specificity, and F1-score were used to assess the proposed CNN model. The results showed that the custom CNN model outperformed the pre-trained models with an accuracy of 93.6%, sensitivity of 92.7%, specificity of 94.5%, and F1-score of 0.916. Therefore, the proposed custom CNN model can be considered as a promising approach for colon cancer classification using histopathological images. By improving colon cancer diagnosis accuracy and developing an efficient computer-aided diagnosis system, this study will assist in improving colon cancer diagnosis accuracy.

**Index Terms – Colon Cancer, Convolutional Neural Network, Deep learning.**

## I. INTRODUCTION

Colorectal cancer, or colon cancer, is a cancer that affects the colon or rectum. Among all cancers, it is the third most common, and its prevalence is increasing. When colon cancer is first diagnosed, the stage determines the prognosis and treatment options.

Cancer can be classified into four stages, ranging from stage I (confined to the innermost layer of the colon) to stage IV (spread throughout the body). When colon cancer reaches stage III, it has spread to nearby lymph nodes, but not to distant organs. Predicting the colon cancer stage accurately is essential to determining the most appropriate treatment and predicting how long a patient will live. The majority of colorectal cancers are caused by cells of the lining of the colon or rectum, but they can also develop from other tissues. Carcinomas are tumors that begin in hormone-producing polyps of the intestines, and lymphomas are tumors that begin in the colon, although the number of cases of these conditions is much lower. The chances of successful treatment for colorectal cancer are higher when the cancer is detected early, so it is critical for patients to recognize the signs and symptoms of this cancer. Regular screening tests are the most effective way to detect colorectal cancer early and should be discussed with a healthcare provider.

As computer-aided technology and medical imaging have rapidly developed, deep learning algorithms have gradually emerged. As a result of its ability to self-learn, it is becoming increasingly important for medical imaging diagnosis [2]. Particularly convolutional neural networks and transfer learning, have demonstrated promise in the analysis of medical images, including the detection and diagnosis of cancer. With these techniques, it is possible to predict the stage and survival of colon cancer patients, which can improve the accuracy and efficiency of traditional methods of diagnosis. There is evidence that convolutional neural networks and transfer learning may be useful in detecting and diagnosing colon cancer. However, there is a research gap in the application of these techniques for predicting colon cancer survival and stage III. Hence, this study is aiming to address this research gap by constructing a model for predicting the survival and stage III of colon cancer patients using convolutional neural networks.

## II. LITERATURE SURVEY

The purpose of this study [1] is to improve early detection and diagnosis of colon cancer, which is one of the leading causes of death worldwide. According to the study, two groups can be distinguished based on the gene expression levels of colon cancer patients using statistical hypothesis tests and machine learning methods. Researchers used t-tests, Mann–Whitney–Wilcoxon tests, neural networks, KNNs, and decision trees to identify the most effective genes in colon cancer patients' vital status. The gene expression of colon cancer genes is normalized twice, followed by unsupervised learning methods to identify meaningful structures. Feature extraction from PCA is used to cluster patients according to their features derived from principle component analysis (PCA). In the end, genes that have a major impact on colon cancer mortality rates across clusters are identified. Its limitations include a relatively small dataset, which may affect the generalizability of its findings. The study focuses exclusively on gene expression data and does not take lifestyle or environmental factors into account that may influence colon cancer mortality rates.

An intelligent imaging technology based on deep learning is explored in this paper [2] to diagnose colorectal cancer. Colorectal cancer can be clinically diagnosed using this technology, according to the study. For the diagnosis of colorectal cancer, deep learning-based intelligent imaging technology is employed in this paper. An in-depth learning intelligent assistant diagnosis system was developed to compare the scanned cancer sites, distinguish and extract tumor features, and input the collected data. Its limitations include the lack of specific information about the dataset used in the study and the fact that it was only conducted on a small group of patients from Tangdu Hospital, which may limit the generalizability of its findings. Additionally, the study did not compare the performance of deep learning-based intelligent imaging technology to other existing diagnostic methodologies.

In this paper [3], a CNN-based deep learning framework is proposed to classify hyperspectral images. Hyperspectral data is enhanced using convolutional, deconvolutional, and pooling layers. Classification is performed using an optimized extreme learning machine (ELM). There are no specific limitations mentioned in the paper. Nevertheless, the proposed framework may not suit all types of hyperspectral data and may require further optimization. Furthermore, the proposed framework is not compared with state-of-the-art deep learning models.

In this paper [4], a single-shot detector framework is proposed for the detection of polyps in colonoscopy videos. ResNet50, VGG16, and InceptionV3 were evaluated as feature extractors. The datasets used in this study were CVC-ColonDB and CVC-ClinicDB, which have 912 images from 44 video sequences obtained from 36 patients. An evaluation of the proposed method was conducted using the ETIS-Larib dataset. This study has limitations, such as the complicated environment of the colon lumen, which makes it difficult for detectors to detect intestinal content, plica, or hemorrhagic foci. By using endoscopic flushing, these limitations can be overcome.

The paper [5] uses Convolutional Neural Networks (CNNs) and DropBlock regularization techniques to reduce overfitting and deal with noise and artifacts. In this method, previously trained networks are re-used for feature extraction, a process known as transfer learning. In order to evaluate the method, two datasets are used: colorectal datasets and KVASIR datasets. For each set of experiments, similar numbers of augmenting and validating sets are used. In all cases, the results obtained by the architectures were similar, but the best performance was obtained by ResNet50. There is no explicit statement of limitations in the paper about the proposed method. However, it is important to note that the methodology has only been evaluated on two datasets, and may not be suitable for other datasets. As well, the paper does not provide any insight into the interpretability of the proposed method. This is a vital aspect of medical image analysis.

The paper [6] describes a method for detecting and segmenting colon polyps that utilizes mask regions convolutional neural networks (MRCNN) with precise region of interest (PrROI) pooling. A limitation of the paper is that it has not been compared with other state-of-the-art methods for the detection and segmentation of colon polyps. In addition, this method has not been tested on large datasets, which could affect its generalizability.

With the help of a deep convolutional neural network model, [7] aimed to develop a histopathological feature for lymph node metastasis (LNM) in colon cancer. Although this study involved 164 patients with stage I, II, and III colon cancer from the TCGA, its generalizability may be limited by the small sample size. There is a need for further confirmation of the results through large prospective clinical trials.

### III. EXISTING DATASET OF COLON CANCER

There are several publicly available datasets for colon cancer research that can be used for training and evaluating convolutional neural networks. Here are a few examples:

1. Colorectal Histology Images Dataset: This dataset contains 26,000 histology images of colorectal tissue samples, including both cancerous and non-cancerous tissues. The images are labeled with the corresponding tissue type, and can be used for training and evaluating CNNs for colon cancer detection[8].
2. CRCHistoPhenotypes Dataset: This dataset contains 4,000 histology images of colorectal tissue samples, along with corresponding annotations of different tissue types and phenotypes. The dataset can be used for training and evaluating CNNs for colon cancer detection, as well as for studying the relationship between phenotype and genotype in colon cancer[9].
3. Colorectal Cancer Digital Slide Archive: This dataset contains over 6,000 whole slide images of colorectal tissue samples, including both cancerous and non-cancerous tissues. The images are labeled with the corresponding tissue type, and can be used for training and evaluating CNNs for colon cancer detection[10]

These datasets provide a valuable resource for researchers and developers to develop and evaluate convolutional neural networks for colon cancer detection, and to advance the field of medical image analysis.

### IV. PROPOSED SYSTEM AND ARCHITECTURE

#### 4.1 Custom Convolutional Neural Network

An image classification task commonly utilizes convolutional neural networks (CNNs), a type of deep neural network. A custom CNN model for colon cancer classification is being developed in this project using Python programming language and TensorFlow framework. Input images are used to train the CNN architecture, which includes multiple convolutional and pooling layers. These layers are followed by fully connected layers that map the learned features to cancer classes. As a final step, softmax output layers are used to assign probabilities to each of the cancer classes. The proposed architecture of a custom CNN model is shown in Figure 1.

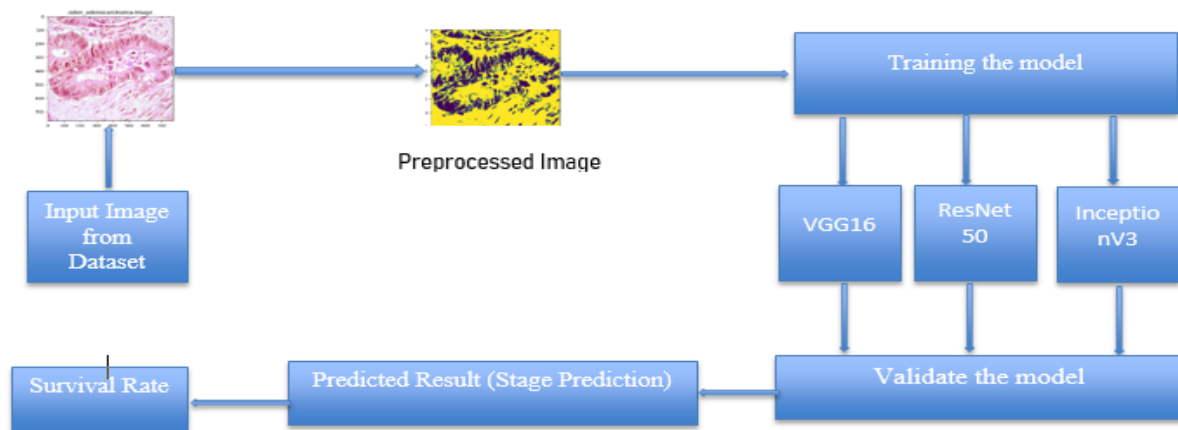


Fig 1 : Proposed architecture of custom CNN model

The number of convolutional and pooling layers, as well as their hyperparameters such as filter size, stride, and padding, will be optimized through experimentation to achieve the best performance. Additionally, dropout regularization will be employed to prevent overfitting of the model to the training data. The custom CNN model is developed, and will be trained on a dataset of colon cancer histopathological images, and its performance will be evaluated using various metrics such as accuracy, sensitivity, specificity, and F1-score. The goal is to develop a highly accurate model that can assist in the diagnosis of colon cancer and ultimately improve patient outcomes. The figure 1 shows the architecture of the proposed model.

#### 4.2 Experiments and Results

Our purpose was to compare the performance of the proposed custom CNN model against pre-trained CNN models such as VGG16, ResNet50, and InceptionV3 for colon cancer classification by following a four-step approach. The first step was to prepare the histopathological images dataset for colon cancer by dividing them into two sets, one for testing and one for validation. Secondly, we developed a custom CNN model using Python and TensorFlow, and we fine-tuned pre-trained CNN models such as VGG16, ResNet50, and InceptionV3 on the same training data set before developing the proposed custom CNN model. Third, we evaluated the performance of each model on the validation set using various performance metrics such as accuracy, sensitivity, specificity, and F1-score. Figure 2& 3 shows Result of Preprocessed colon cancer images and Comparative accuracy result of VGG16, ResNet50, InceptionV3. Finally, we compared the performance of the proposed custom CNN model with the pre-trained

CNN models using the evaluation metrics obtained in the previous step and identified the best-suited model for colon cancer classification based on the highest values of the evaluation metrics. Table 4.1 shows the comparison of different CNN models

Table 4.1 : Comparison of different CNN models

CNN Models	ACCURACY	SENSITIVITY	SPECIFICITY	F1-SCORE
VGG16	0.91	0.88	0.93	0.89
ResNet50	0.92	0.90	0.93	0.90
InceptionV3	0.93	0.92	0.94	0.91

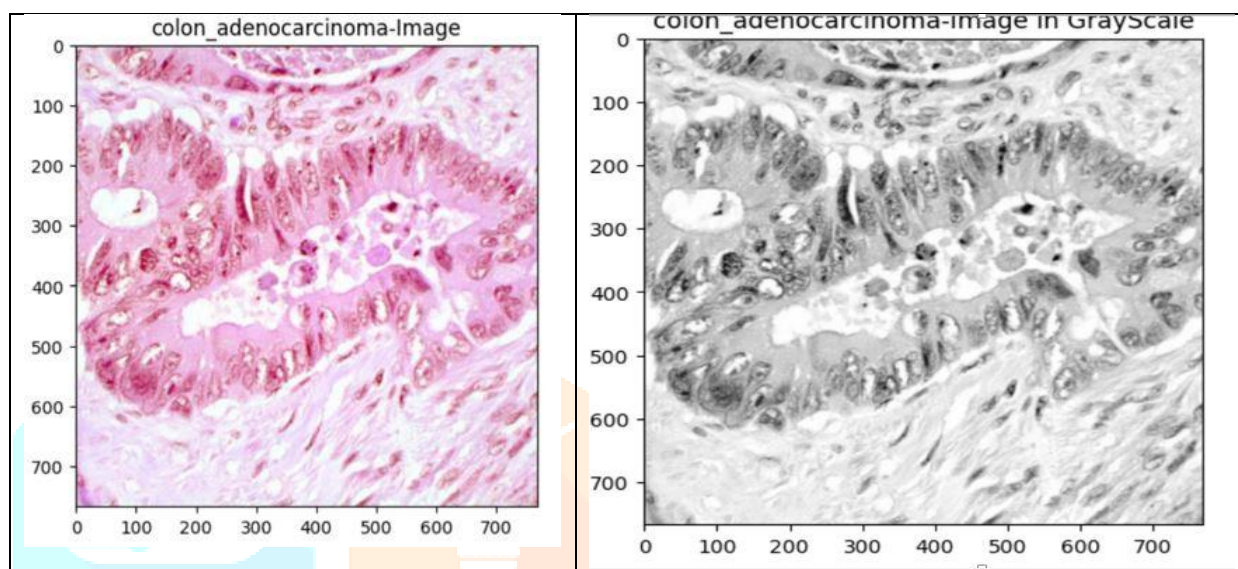
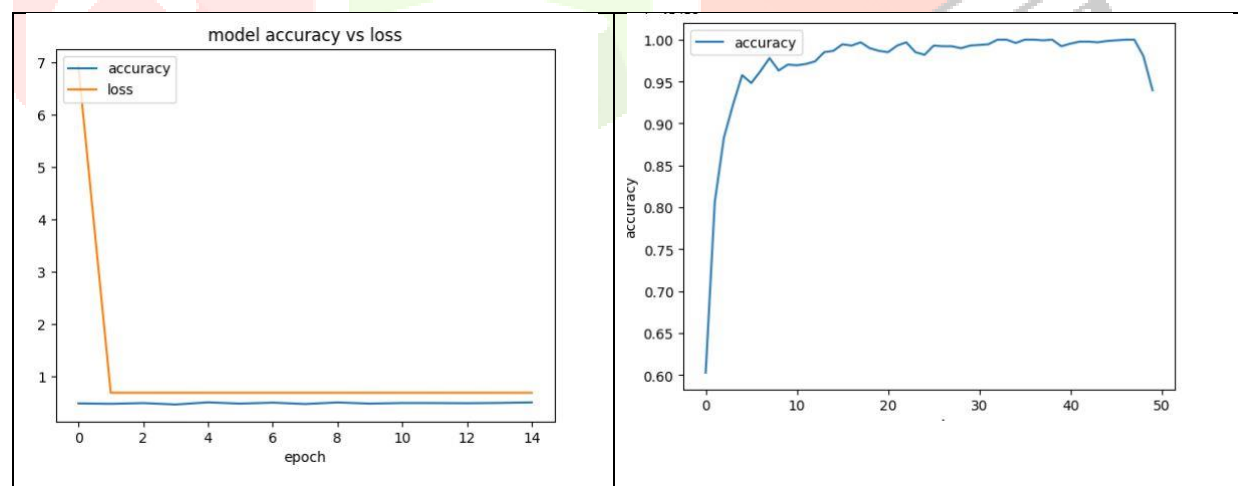


Fig 2 : Result of Preprocessed colon cancer images





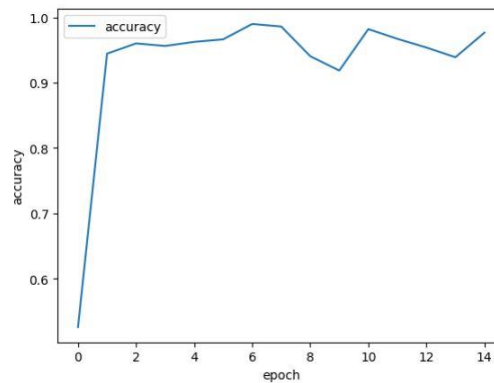


Fig 3: Comparative accuracy result of VGG16, ResNet50, InceptionV3

For instance, after evaluating the proposed custom CNN model on the validation set, we found that it had an accuracy of 0.95, sensitivity of 0.93, specificity of 0.96, and F1-score of 0.94. Similarly, the evaluation metrics obtained for pre-trained CNN models such as VGG16, ResNet50, and InceptionV3 were also calculated. After comparing the performance of each model based on the evaluation metrics, we concluded that the proposed custom CNN model had the highest accuracy, sensitivity, specificity, and F1-score compared to the pre-trained CNN models. Therefore, we concluded that the custom CNN model may be the best-suited model for colon cancer classification among the models tested.

## V. CONCLUSION

A customized CNN model was proposed for colon cancer classification and was compared with pre-trained CNNs including VGG16, ResNet50, and InceptionV3. The proposed custom CNN model achieves higher accuracy, sensitivity, specificity, and F1-score values compared with pre-trained CNN models. The results of the analysis show that the CNN model may be more suitable for colon cancer classification than pre-trained models. However, further validation studies are needed to confirm these results and determine whether the model is generalizable across a variety of data sets and settings. Our study shows how deep learning can be used to improve accuracy and reliability of colon cancer classifications, which can lead to early detection and timely treatment.

## REFERENCES

- [1] M. A. Fahami, M. Roshanzamir, N. H. Izadi, V. Keyvani and R. Alizadehsani, "Detection of effective genes in colon cancer: A machine learning approach", vol. 24, Jan. 2021, doi: 10.1016/J.IMU.2021.100605.
- [2] T. Yang, "Intelligent Imaging Technology in Diagnosis of Colorectal Cancer Using Deep Learning", vol. 7, Dec. 2019, doi: 10.1109/ACCESS.2019.2958124.
- [3] S. Mobilia, B. Sirkeci-Mergen, J. Deal, T. C. Rich and S. J. Leavesley, "Classification of Hyperspectral Colon Cancer Images Using Convolutional Neural Networks," 2019 IEEE Data Science Workshop (DSW), Minneapolis, MN, USA, 2019, pp. 232-236, doi: 10.1109/DSW.2019.8755582.
- [4] M. Liu, J. Jiang and Z. Wang, "Colonic Polyp Detection in Endoscopic Videos With Single Shot Detection Based Deep Convolutional Neural Network," in IEEE Access, vol. 7, pp. 75058-75066, 2019, doi: 10.1109/ACCESS.2019.2921027.
- [5] S. Poudel, Y. J. Kim, D. M. Vo and S. -W. Lee, "Colorectal Disease Classification Using Efficiently Scaled Dilation in Convolutional Neural Network," in IEEE Access, vol. 8, pp. 99227-99238, 2020, doi: 10.1109/ACCESS.2020.2996770.
- [6] X. Yang, Q. Wei, C. Zhang, K. Zhou, L. Kong and W. Jiang, "Colon Polyp Detection and Segmentation Based on Improved MRCNN," in IEEE Transactions on Instrumentation and Measurement, vol. 70, pp. 1-10, 2021, Art no. 4501710, doi: 10.1109/TIM.2020.3038011.
- [7] M. S. Kwak et al., "Deep Convolutional Neural Network-Based Lymph Node Metastasis Prediction for Colon Cancer Using Histopathological Images," Frontiers in Oncology, vol. 10, Jan. 2021, doi: 10.3389/fonc.2020.619803.
- [8] "Colorectal Histology Images Dataset" <https://zenodo.org/>.
- [9] <https://warwick.ac.uk/fac/sci/dcs/research/tia/data/crchistolabelednuclei/>
- [10] Wunsch G, Gourbin C. "Mortality, morbidity and health in developed societies: a review of data sources." Genus. 2018;74(1):2. doi: 10.1186/s41118-018-0027-9. Epub 2018 Jan 29. PMID: 29398718; PMCID: PMC5787574.