Early Detection of Colon Cancer using Fine Tuned Light Weight CNN Model

Vanmathi C
School of Computer Science Engineering and InformationSystems
*Vellore Institute of Technology,*Vellore,
Tamilnadu, India
vanmathi.c@vit.ac.in

Nanda Kishore K V
School of Computer Science Engineering and InformationSystems
Vellore Institute of Technology,
Vellore, Tamilnadu, India
nandakishore.kv2020@vistudent.ac.in

***Abstract* - There are around 1.8 million new cases of colorectal cancer diagnosed every year, making it one of the most prevalent forms of the disease. Therefore, the infection and mortality rates can be reduced with early identification of this cancer. Most diagnostic colonoscopy systems are currently incorporating artificial intelligence approaches validated for predicting advanced cancers. It is common practice to combine convolutional neural network-based patterns with image patches and preprocesses. The purpose of this paper was to develop a novel lightweight deep learning Convolutional Neural Network (CNN) that uses Adam optimizer for reliable colon cancer detection. Applying a database of publicly available histopathology images, the proposed approach is analyzed in the context of the existing best practices for colon cancer detection analysis. Based on the results, it is clear that the proposed deep neural network for colon cancer diagnosis achieves an accuracy of 98.40 percent, which is the highest accuracy achieved by any of the existing deep learning methods.**

Keywords - Colon cancer, Cancer detection, Adam optimizer, CNN, histopathological images

I. INTRODUCTION

Cancer is one of the most serious diseases that affects people's health and has a high death rate. Malignant tumors have a common source: cancers. Benign tumors are not cancer; they are frequently removed and provide only a remote risk because they rarely return after surgery. However, malignant tumors and malignancies are harmful because of their erratic and unchecked growth. In 2018, colorectal cancer was the second leading cause of cancer-related death worldwide [1]. Certain dietary factors, such asa low-fiber and high-fat diet, as well as family history, enhance the likelihood of developing colorectal cancer. Constipation, tiredness, and weakness are common symptoms. So, early cancer detection increases the likelihood of a successful treatment. There are several distinct forms of colorectal cancer, with adenocarcinoma being the most common and developing in the mucus-producing glands of the large intestine; gastrointestinal carcinoid tumor developing in the hormone-producing cells of the large intestine; gastrointestinal stromal tumor developing in the Cajal interstitial cells of the colon; primary colorectal lymphoma developing in the colon; and sarcoma developing in the blood vessels. The best strategy to avoid colorectal cancer is to have any precancerous lesions removed via endoscopy. As early detection of malignant lesions improves the prognosis of patients with colorectal cancer, there is a need for dependable, early, and precise endoscopic diagnosis [2].

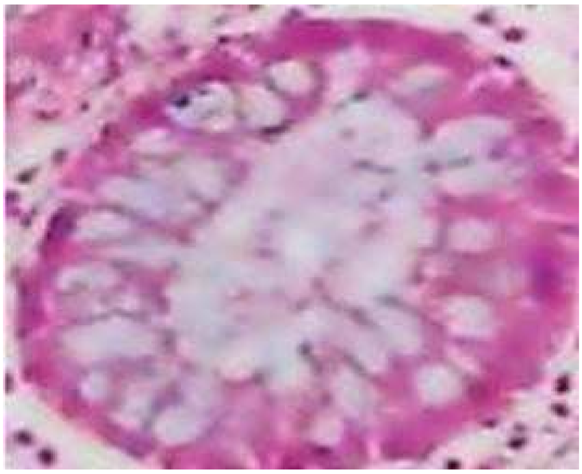
Pathologists rely most heavily on histopathological image analysis when classifying colorectal tumors with varying degrees of differentiation. Figure 1 shows some examples of images. There have been recent advancements in computer-aided diagnosis (CAD) systems that can scan for colon tumors or cancers automatically. Cancer detection, diagnosis, and tumor segmentation are all areas where deep learning algorithms shine due to their ability to automatically extract rich information from raw images. Furthermore, the paper is structured as follows: In Part II, discussed about some of the related research that has been done. Section III explains the suggested method, a CNN model employing Adam Optimizer. is explained. The experimental findings and discussion are presented in Section IV, and the conclusion is provided in Section V. In digital imaging, ever image is composed of pixels which are the smallest addressable units of an image. Every pixel stores information that gives the color of the image at that particular point. Each pixel has a specific color that is represented by the combination Red, Green and Blue components. Each of these three components can take a value ranging from 0 to 255. A value 0 means that specific color component is totally absent while a value 255 means that the color component is fully present. Therefore, we can produce any color as a combination of these three basic colors.

II. RELATED WORK

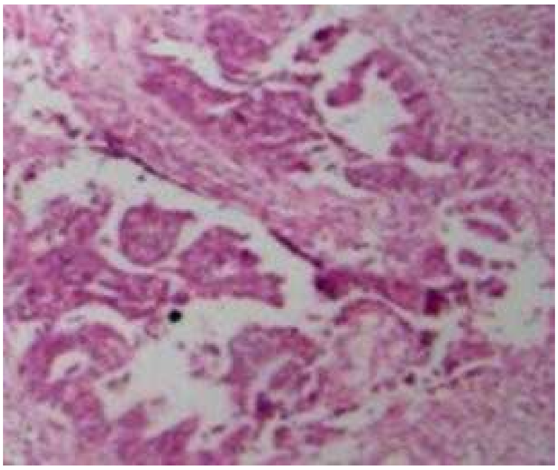
The most reliable method for detecting colorectal cancer is colonoscopy. The rate at which polyps are overlooked during colonoscopy, however, rises in direct proportion to the extent to which the expert is versed in endoscopy. Therefore, AI technologies may aid in filling in the gaps in clinical expertise, resulting in a lower incidence of missed lesions during colonoscopies [3]. A major one [4] is played by medical imaging, which is an effective application employed in the early diagnosis of cancer. Even though there are more medical imaging data, it is hard and takes a lot of time to figure out how the data relate to how quickly the disease is getting worse. Furthermore, the accuracy rate

979-8-3503-2518-8/23/$31.00 © 2023 IEEE

drops dramatically, and the duration of early detection is extended [5] if physicians' misinterpretation of data is considered in the diagnosis of diseases. Cancer detection, categorization, and tumor segmentation diagnosis are just some of the many applications of machine learning, a subfield of artificial intelligence, in medical image processing [6].



1(a)



1(b)

Fig. 1 Histopathological images of (a) normal and (b) malignant colon tissue

Non-specialists can get help from a computer-aided diagnostic system for endocytoscopic imaging in identifying lesions. Because it simply takes one click to get a real-time diagnostic report [7], such a system is more helpful to trainees than the opinions of experts. When comparing different models for detecting colorectal cancer early on, it was discovered that the one combining ANN and Gaussian expectation-maximization performedthe best [8].

Success with image analysis and other biomedical activities prompted the initial push to use Deep Learning (DL) to histology. Ciresan et al. [9] conducted one of the earliest rigorous experiments with deploying DL architectures in a Whole Slide Images processing setting. It is possible to identify projecting, flat, and recessed lesions in colon endoscopic cases by employing CNNs like AlexNet and Caffe. It was shown by Ito et al.[10] that this method produces reliable diagnostics with high areas under the receiver operating characteristic curve (AUC). The survival rate of patients with colon cancer was predicted by Ahmed et al. [11] using an artificial neural network. The survival rate is analyzed using a feed forward neural network with three layers. Using deep convolution networks, Kather et al. [12] analyzed the survival rate of people with colorectal cancer. One hundred thousand histological pictures from patients' tissue slides are acquired during this procedure. Colorectal data from 409 patients from various German hospitals is used to test the proposed system. Iizuka et al.

[13] used a convolution and recurrent neural network to develop a method for detecting stomach and colonic epithelial carcinoma. In this study, we collect colorectal cancer histology photos from patients and then extract useful information from those images. The new network classifies tumors as adenomas, adenocarcinomas, or non- neoplastic by analyzing the entire slide. The automatic computer-aided technique based on backpropagation neural networks was created by Daniel et al. [14] to make breathomics gastric cancer predictions. The processing of medical pictures using convolutional neural networks (CNNs) has proven widely successful. To identify mitosis in breast WSI, researchers implemented a deep max-pooling Convolutional Neural Network (CNN). In reality, convolutional neural networks (CNNs) are currently the gold standard for analysing medical images. Through the classification and segmentation of histopathological pictures, they have made significant contributions to therevolution in the identification of cancers such as breast, lung, colon, and brain [15–17].

III. PROPOSED WORK

The proposed method for predicting colorectal cancer from the database employs a Convolution Neural Network (CNN)trained using the Adam optimizer. Figure 2 shows the comprehensive colorectal cancer detection framework employing the new method. CNN is an artificial neural network with a specialized ability to recognize patterns. The ability to recognize patterns is what makes CNNs so effective in the field of image analysis. Convolutional neuralnetworks (CNNs) consist of a series of hidden layers called convolutional layers, each of which accepts input, performs some transformation on it, and then sends the result to the next layer. Input channels are what feed information into convolutional layers, while output channels are what send processed data out. The change performed by a convolutional layer is known as a convolution operation or cross correlations.The CNN based model is designed using a database of 10,000 histopathological colon images [18]. The images of colon tissue split evenly between adenocarcinomas and benign polyps. Each JPEG picture is 768 pixels on the wide side. After these images have been normalized, they are sent into our CNN model, which is involved in detecting colon cancer. There are ten convolutional blocks in the model, and a max pool layer is included in each one. The architecture consists of a flatten layer at the base and a completely connected 256-unit ReLU-activated layer at the top. After the flatten layer, a 0.5-probability dropout layer is added, followed by a SoftMax layer for classification. Adamis now the most popular and effective optimizer available. Adam usually functions best with a lower learning rate, beginning with 0.001 and adjusting up or down from there. In this case, a value of 0.005 is appropriate. Figure 3 depicts some sample image from the data set.

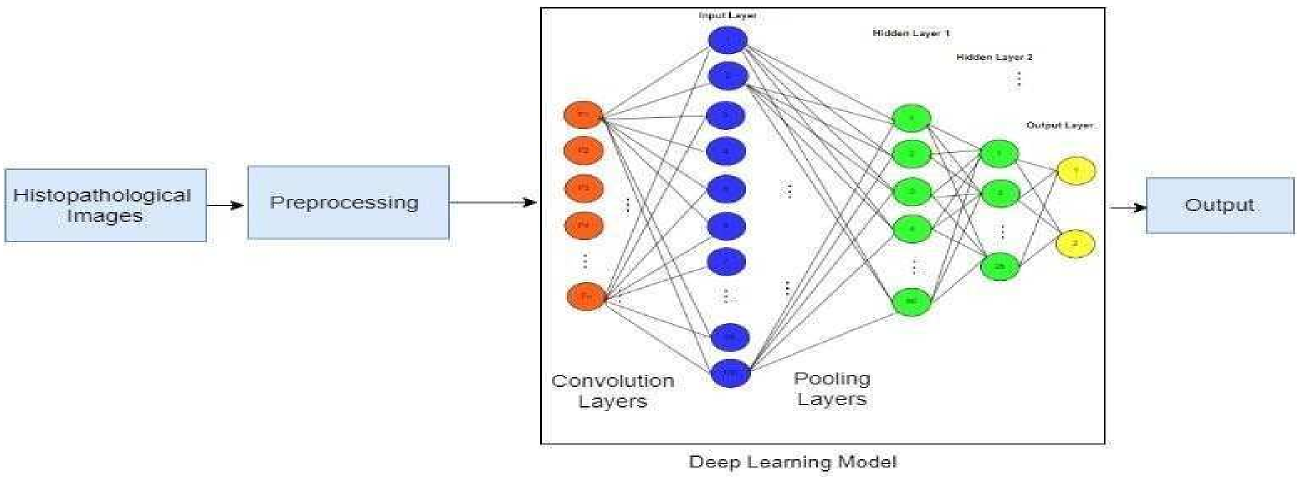


Fig. 2 The framework of the light weight CNN

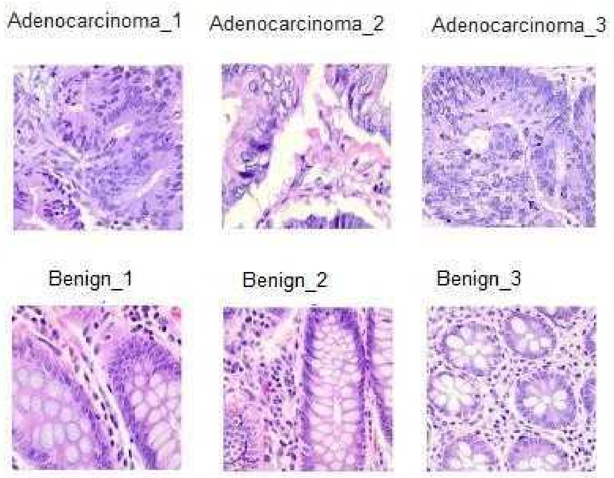


Fig. 3 Sample Images

IV. RESULTS AND DISCUSSION

To evaluate the efficacy of the proposed technique, histopathology images from the Kaggle dataset [18] is used. The architecture is simulated using Python 3.7.2. The model was trained for 35 iterations with a total of 1,023,112 parameters, with a minimum batch size of 40. The system used 6,000 of the records for training, 2,000 for validation, and 2,000 for testing. The suggested model has better accuracy at detecting colon cancer than the state-of-the-art CNN method. Figure 4 provides an epoch versus accuracy graph to illustrate the performance of the deep learning model. This visual depiction demonstrates how a model acquires information from a training sample over time. Here, the blue line represents the evolution of training sample accuracy, while the orange line represents validation sample accuracy. Both start off with somewhat differingaccuracy values, and as time progresses, the differences become more pronounced. After 20 iterations, the accuracy gap between training and validation nearly vanished. The model has been successfully trained if the validation accuracy is greater than the training accuracy.

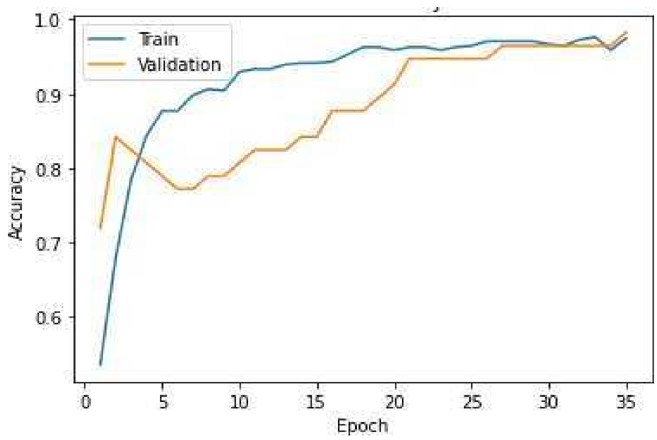


Fig. 4 Model Accuracy

The figure 5 shows the loss value incurred by the training and validation samples. Initially both the curves starts at distinct value of loss value. As number of epoch increases both the curves heading towards the minimal value. The figure says the validation curve reaches the minimal earlier than the training curve that proves that model has been trained efficiently.

In order to show how well the suggested method works, it is compared to the state-of-the-art techniques [11–14], which use CNN to predict colon cancer. The metrics Precision, Recall, F1score, and Accuracy are used to measure how well the suggested method works. The formulas for calculating the metrics are provided in equation 1 to equation 4. The classification of the Adenocarcinoma and Benign are classified very efficiently and it is given in table1. The experimental results show that the suggested method of predicting colon cancer using CNN is 98.4% more accurate than the methods that are already in use. Table 2 shows the results of the analysis.

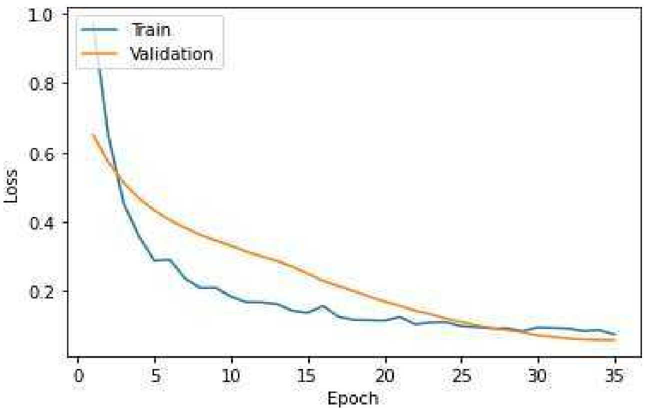


Fig. 5 Model loss

Precision =TP/ (TP+FP) (1)
Recall=TP/ (TP+FN) (2)
F1Score=TP/ (TP+1/2(FP+FN)) (3)
Accuracy= (TP+FN) / (TP+TN+FP+FN) (4)

Where

TP = The number of times where the model successfullypredicted the positive class.
TN = The number of times the negative class was properlypredicted by the model
FP = The number of times the model wrongly predicted thepositive class.
FN = The number of times the negative class was wronglypredicted by the model.

TABLE 1 MODEL PREDICTION PERFORMANCE METRICS, P–PRECISION, R-RECALL, FS-F1SCORE AND A-ACCRACY

Class	P	R	FS	A
Adenocar cinoma	0.988	0.983	0.985	0.992
Benign	0.989	0.978	0.983	0.975

TABLE 2 COMPARATIVE ANALYSIS OF THE EXISTING METHODS OVER CURRENT METHOD

Authors / Metrics	Recall	Precision	F1score
Ahmed et al. [11]	89.7	86.9	88.3
Kather et al. [12]	94.7	95.7	95.2
Izuka et al. [13]	96.89	96.39	96.64
Daniel et al. [14]	95.98	96.29	96.135
Proposed	98.8	98.0	98.4

V. CONCLUSION

Colon cancer is in the top three of the most serious and dangerous cancers in the world. Early detection is crucial fortreating any type of cancer. The positive impacts and results of using deep learning to medical image analysis in early cancer detection and screening have led to its recent rise in popularity. The accuracy of the proposed method for predicting colon cancer cells is higher (98.40%) than that of previous research work. Better results were achieved while using the proposed lightweight CNN with Adam Optimizer for classifying images of colon cancer. There is still anopportunity for improvement in the classification accuracy needed for the precise prediction of colon cancer in thefuture.

REFERENCES

[1] Bray, Freddie, Jacques Ferlay, Isabelle Soerjomataram, Rebecca L. Siegel, Lindsey A. Torre, and Ahmedin Jemal. "Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries." CA: a cancer journal for clinicians 68, no. 6 (September, 2018): 394-424.

[2] Bejnordi, Babak Ehteshami, Guido Zuidhof, Maschenka Balkenhol, Meyke Hermsen, Peter Bult, Bram van Ginneken, Nico Karssemeijer, Geert Litjens, and Jeroen van der Laak. "Context-aware stacked convolutional neural networks for classification of breast carcinomas in whole-slide histopathology images." Journal of Medical Imaging 4, no. 4 (December, 2017): 044504-044504.

[3] Silva, Juan, Aymeric Histace, Olivier Romain, Xavier Dray, and Bertrand Granado. "Toward embedded detection of polyps in wce images for early diagnosis of colorectal cancer." International journal of computer assisted radiology and surgery 9 (September, 2014): 283-293.

[4] Fu, Jachih JC, Ya-Wen Yu, Hong-Mau Lin, Jyh- Wen Chai, and Clayton Chi-Chang Chen. "Feature extraction and pattern classification of colorectal polyps in colonoscopic imaging." Computerized medical imaging and graphics 38, no. 4 (June, 2014):267-275.

[5] Bernal, Jorge, F. Javier Sánchez, Gloria Fernández- Esparrach, Debora Gil, Cristina Rodríguez, and Fernando Vilariño. "WM-DOVA maps for accuratepolyp highlighting in colonoscopy: Validation vs. saliency maps from physicians." Computerized medical imaging and graphics 43(July, 2015): 99-111.

[6] Bernal, Jorge, Javier Sánchez, and Fernando Vilarino. "Towards automatic polyp detection with a polyp appearance model." Pattern Recognition 45, no. 9 (September, 2012): 3166-3182.

[7] Mori, Yuichi, Shin-ei Kudo, Kunihiko Wakamura, Masashi Misawa, Yushi Ogawa, Makoto Kutsukawa, Toyoki Kudo et al. "Novel computer-aided diagnostic system for colorectal lesions by using endocytoscopy (with videos)." Gastrointestinal endoscopy 81, no. 3 (May, 2015): 621-629.

[8] Wan, Jing-Jing, Bo-Lun Chen, Yi-Xius Kong, Xing-Gang Ma, and Yong-Tao Yu. "An early intestinal cancer prediction algorithm based on deep belief network." Scientific reports 9, no. 1(November, 2019): 17418.

[9] Cireşan, Dan C., Alessandro Giusti, Luca M. Gambardella, and Jürgen Schmidhuber. "Mitosisdetection in breast cancer histology images with deep neural networks." In Medical Image Computing and Computer-Assisted Intervention– MICCAI 2013: 16th International Conference, Nagoya, Japan, September 22-26, 2013, Proceedings, Part II 16, pp. 411-418. Springer Berlin Heidelberg, 2013.

[10] Ito, Nao, Hiroshi Kawahira, Hirotaka Nakashima,Masaya Uesato, Hideaki Miyauchi, and Hisahiro Matsubara. "Endoscopic diagnostic support system for cT1b colorectal cancer using deep learning." Oncology 96, no. 1 (December, 2018): 44-50.

[11] Ahmed, Farid E. "Artificial neural networks for diagnosis and survival prediction in colon cancer."Molecular cancer 4, no. 1 (August, 2005): 1-12.

[12] Kather, Jakob Nikolas, Johannes Krisam, Pornpimol Charoentong, Tom Luedde, Esther Herpel, Cleo-Aron Weis, Timo Gaiser et al.

"Predicting survival from colorectal cancer histology slides using deep learning: Aretrospective multicenter study." PLoS medicine 16, no. 1 (January, 2019): e1002730.

[13] Iizuka, Osamu, Fahdi Kanavati, Kei Kato, Michael Rambeau, Koji Arihiro, and Masayuki Tsuneki. "Deep learning models for histopathological classification of gastric and colonic epithelialtumours." Scientific reports 10, no. 1 (January, 2020): 1504.

[14] Daniel, D. Arul Pon, and K. Thangavel. "Breathomics for gastric cancer classification using back-propagation neural network." Journal of medical signals and sensors 6, no. 3 (September, 2016): 172.

[15] Mohite, Aashka. "Application of transfer learning technique for detection and classification of lung cancer using CT images." Int J Sci Res Manag 9, no. 11 (November, 2021): 621-34.

[16] Spanhol, Fabio Alexandre, Luiz S. Oliveira, Caroline Petitjean, and Laurent Heutte. "Breast cancer histopathological image classification using convolutional neural networks." In 2016 international joint conference on neural networks (IJCNN), pp. 2560-2567. IEEE, July, 2016.

[17] Ker, Justin, Yeqi Bai, Hwei Yee Lee, Jai Rao, and Lipo Wang. "Automated brain histology classificati on using machine learning." Journal of Clinical Ne uroscience 66 (August, 2019): 239-245.

[18] Borkowski AA, Bui MM, Thomas LB, Wilson C P, DeLand LA, Mastorides SM. Lung and Colon Cancer Histopathological Image Dataset (LC2500 0). arXiv:1912.12142v1 [eess.IV], December, 2019.