ISSN: 0975-833X

Available online at http://www. journalcra.com

International Journal of Current Research
Vol. 16, Issue, 03, pp. 27579-27585, March, 2024
DOI: https://doi.org/10.24941/ijcr.46630.03.2024

INTERNATIONAL JOURNAL OF CURRENT RESEARCH

RESEARCH ARTICLE

NOVEL APPROACH IN DETECTING COLON CANCER USING DCNN: A SYSTEMATIC STUDY

***Malavika, C. and Jeyashanthi, N.**

Velalar College of Engineering and Technology, Thindal, Erode, Tamil Nadu

ARTICLE INFO

Article History:
Received 14th December, 2023
Received in revised form
20th January, 2024
Accepted 24th February, 2024
Published online 30th March, 2024

Key words:
Adenocarcinoma, Benign Tissue, Colon Cancer, Histopathology, DCNN, LC25000.

***Corresponding author:**
Malavika Chandran

ABSTRACT

Colon cancer remains a significant global health concern, and early detection is crucial for effective treatment and improved patient outcomes. Histopathological examination of tissue samples is a standard diagnostic procedure, but the manual analysis of large datasets is time-consuming and prone to human error. This study proposes a novel approach for colon cancer detection by Deep Convolutional Neural Networks (DCNN). This research focuses on developing an automated system that utilizes DCNN to analyze histopathological images of colon tissues. The DCNN model is trained on a comprehensive dataset (LC25000) comprising both adenocarcinoma and benign tissue samples. The model's ability to extract high-level representations from complex image data allows for accurate and efficient classification. The evaluation of the proposed system involves assessing its performance in terms of sensitivity, specificity, and overall accuracy. In this DCNN, the model get 98.7% accuracy, 99% precision, 99% recall, 99% F1 score. The results demonstrate the model's capability to accurately identify cancerous regions within histopathological images, providing a valuable tool for pathologists and clinicians.

Citation: Malavika, C. and Jeyashanthi, N., 2024. *"Novel approach in detecting colon cancer using dcnn: a systematic study". International Journal of Current Research, 16, (03), 27579-27585.*

INTRODUCTION

According to the World Health Organization (WHO), cancer is the largest purpose of morality loss in the world (WHO, "Cancer", 2022). Colon cancer is a growth of cells that begins in a part of the large intestine called the colon. The colon is the first and longest part of the large intestine. Colon cancer is sometimes called colorectal cancer. This term combines colon cancer and rectal cancer, which begins in the rectum (Mayoclinic, 2023). Colon and rectal cancers are uncommon in underdeveloped countries but are the second-highest common type of cancer in rich societies. Each year, more than 940,000 instances appear worldwide, and approximately 500,000 people die due to colon cancer (WHO, "colon cancer", 2021).

Colon cancer begins from forming small, benign collections of cells called polyps on the colon's inner wall. Some of these polyps may eventually grow into colon malignancies. A tumor originates in the majority of cases of colon cancer when healthy cells in the colon or rectum multiply uncontrollably. Adenocarcinoma of the colon or rectum begins in the epithelial cells of the large intestine and subsequently spreads to the other layers (Sanower Hossain *et al*., 2022). According to GLOBOCAN data, the incidence and mortality of colon cancer vary considerably between countries among world regions. They are also associated with the socioeconomic status of the country. According to the world bank, the new cases and death are more remarkable in areas with higher income levels and lesser in areas with lower income level is shown in table 1.1 (GCO, 2021).

The etiology of CRC remains unclear, but it may be related to the Genetic factors. About 20% of Colon cancer cases are related to genetic factors, and investigations have shown a three-fold increased risk of cancer in the first-generation relatives of colon cancer patients. Familial Adenomatous Polyposis (FAP) has been identified as a genetic syndrome that predisposes to colon cancer, and the Mismatch Repair Gene (MMR) has also been linked to inherited colon cancer (Maida *et al*., 2017). It is currently believed that high fat, high animal protein, and low cellulose diet are related to the incidence of colon cancer. Excessive fat intake will promote bile secretion, bile acid decomposition, increased intestinal carcinogens, and the activity of intestinal anaerobic bacteria (murphy *et al*., 2019). Non-cancerous diseases such as colorectal polyps, colorectal adenomas, ulcerative colitis and Crohn's disease, etc. can contribute to colon cancer (Roncucci *et al*., 2015). Carcinogenic exposure and lifestyle, such as sedentary and overweight, are risk factors for CRC, and the incidence of sigmoid and rectal cancer is higher in patients undergoing pelvic radiation therapy (Sun, 2014). The current way of detecting cancer is extremely time-consuming and labor-expensive. Pathologists must get extensive knowledge by studying labeled histopathological images to identify colon images. As a result, a significant amount of resources and manual work are squandered. As a result, increased diagnostic accuracy and diagnostic speed are necessary.

Computer technologies have garnered considerable attention due to their inherent advantages, including computational power, speed, and storage capacity (Mu *et al*., 2007).

Table 1. World colon cancer estimated incidence and mortality rates in 2020

Population	Incidence	Mortality
Upper middle income	887,025	461,511
High income	819,143	340,272
Low middle income	194,954	112,556
Low income	29,542	20,392
Total	1,930,664	934,731

ML is one of the most exciting uses of computer-aided technology, owing to its capacity for human-like learning, which automatically improves the predictive performance of its models by learning from data (Whitehall and Lu, 2001). There have been numerous research studies conducted to date on colon cancer analysis helped by computer technologies (Sertel *et al*., 2009). In a large bound of applications, deep learning algorithms for image identification have proven to be incredibly effective, frequently outperforming human ability. The key concept is that an adaptable software network may be trained, parameters assigned values, to identify images through many tagged images. Once trained, the network can be used to classify the appropriate label for unlabeled images (Sena *et al*., 2019). The current study's primary objective is to evaluate the use of deep learning for the histological investigation of colon cancer by analyzing digitized pathology images by DCNN model. So, our proposed model gives the contribution such as,

(i) This project proposes a fine-tuned DL model that yields promising results in the classification of colon cancer.
(ii) Our proposed DL model achieves much better accuracy when compared to other existing related works within a short time.

The proposed model can save both time and space by utilizing effective data processing techniques.

Related Work: Various attempts have been made by researchers to detect cancer from histopathological images using image processing techniques, machine learning, and deep learning techniques. The research of cancer detection based on histopathological image analysis has advanced significantly over the years. Some of the important related works based on histopathological image analysis of cancer. Shahid Mehmood *et al*.,(2022) suggested that a pretrained neural network (AlexNet) for detecting the lung and colon cancer of histopathological images by modifying four of its layers before training it on the dataset. The initial results were promising, with an overall accuracy of 89% for all image classes. To improve the accuracy and efficiency of the model, they focused on enhancing the quality of images in the under performing class using a simple and efficient contrast enhancement technique, instead of applying image enhancement techniques to the entire dataset. By implementing this methodology, the overall accuracy from 89% to 98.4%, but also ensured computational efficiency.

Panyun Zhou *et al.* (2022) recommended a technique called gradient-weighted class activation map (Grad-CAM) to identify important areas in the HCCANet model for better understanding. The results of the experiment showed that the HCCANet model performed better than four advanced deep learning techniques (ResNet50, MobileNetV2, Xception, and DenseNet121) and four classical machine learning techniques (KNN, NB, RF, and SVM). The model achieved 90.2%, 85%, and 86.7% accuracy in classifying colorectal cancers with high, medium, and low differentiation levels respectively. The overall accuracy was 87.3% with an average AUC value of 0.9 and this study shows promise for postoperative adjuvant diagnosis and grading of colorectal cancer. Ben Hamida *et al*., (2021) planned the DL architectures to classify and identify colon cancer regions in sparsely annotated histopathological data. A review and comparison of state-of-the-art CNN models is conducted, including AlexNet, VGG, ResNet, DenseNet, and Inception. Transfer learning techniques are used to overcome the lack of rich WSI datasets, training the network on ImageNet to generate diverse learned features. Patch-level classification accuracy is evaluated on the AiCOLO colon cancer dataset, achieving up to 96.98% accuracy with the ResNet model. The CNN models are also tested on the CRC-5000, nct-crc-he-100k, and merged datasets, with ResNet achieving accuracy rates of 96.77%, 99.76%, and 99.98% respectively. Additionally, a pixel-wise segmentation strategy for colon cancer WSIs is presented using UNet and SegNet models. To address sparse annotation, a multi-step training strategy is introduced. UNet and SegNet are tested in different scenarios, including data augmentation and transfer learning, achieving accuracy rates of up to 76.18% and 81.22% respectively. The training strategy and models are evaluated on the CRC-5000, nct-crc-he-100k, and Warwick datasets, with SegNet achieving accuracy rates of 98.66%, 99.12%, and 78.39% respectively. Meiyan Liang *et al*., (2020) proposed that the identification of colon cancer using multi-scale feature fusion convolutional neural network (MFF-CNN) based on shearlet transform in histopathological images. This framework extracts shearlet coefficients at multiple scales and combines them with the original image. The MFF-CNN achieves a 96% identification accuracy and an average F-1 score of 0.9594. It also reduces false negative and false positive rates. This network offers a new perspective in real-time, accurate cancer diagnosis. Bukhari *et al*., (2020) stated that the three variants of CNN (ResNet-18, ResNet-34 and ResNet-50) have been employed to evaluate the images. Three CNN architectures (ResNet-18, ResNet-30, and ResNet50) were applied for the classification of digitized images of colonic tissue. The accuracy (93.91%) of ResNet-50 was the highest which is followed by ResNet-30 and ResNet-18 with the accuracy of 93.04% each. Sameen Aziz *et al*., (2023) suggested a new approach called ImageNet-VGG16 (IVNet) is introduced for the real-time diagnosis of breast cancer within a hospital setting. The main objective of this research was to accurately classify breast cancer images into Grade-1, Grade-2, and Grade-3. Through extensive experimental research, the researchers achieved an impressive 97% correct classification rate by combining VGG16 and ImageNet as the proposed feature engineering method, IVNet.

Jyothi Peta *et al*.,(2023) projected that it involves five steps: image acquisition, encryption, key generation, secure data storage, and disease classification. Medical images are collected and encrypted using the E-EIE method. Optimal keys are generated using the I-SCSO algorithm. The FLF framework ensures secure storage of the encrypted images. The images are then decrypted and classified using the C2T2Net model. The proposed classifier's loss is reduced using the CTSO algorithm. The simulation and experimental analysis are conducted using Python and the BreakHis Database. The simulation results demonstrate that the proposed study achieves superior performance in terms of accuracy (95.68%), recall (95.6%), precision (95.66%), F-measure (95.63%), specificity (95.6%), and kappa coefficient (95.26%). Saeed Iqbal *et al*.,(2022) proposed a introduces the BreastUNet, a novel Deep Convolutional Neural Network (CNN) with a feature grafting approach specifically designed for analyzing mitotic nuclei in breast histopathology images. The results demonstrate that the proposed model outperforms the others on the test set, achieving an impressive F1 score of 0.95. Additionally, the model exhibits high sensitivity and specificity, both at 0.95, and an area under the precision curve of 0.95. Irum Hirra *et al*.,(2021) preferred that novel patch-based deep learning approach called Pa-DBN-BC is introduced for the detection and classification of breast cancer in histopathology images, utilizing the Deep Belief Network (DBN). The proposed method involves two phases: unsupervised pre-training and supervised fine-tuning. The model achieved an accuracy of 86% on this dataset, demonstrating its effectiveness in breast cancer detection and classification. Kausik Das *et al*.,(2020) suggested a Deep Multiple Instance Learning (MIL) based CNN framework. The effectiveness of this method has been evaluated using three breast cancer datasets: Break His, IUPHL, and UCSB. Impressively, the proposed framework achieved accuracies of 93.06%, 96.63%, and 95.83% on these respective datasets. This demonstrates the potential of the Deep MIL-based CNN framework in accurately classifying breast cancer cases based on histopathological images.

Experimental Analysis

Our proposed methodology involves utilizing a diverse dataset of histopathology images for training deep convolutional neural networks (DCNNs). These networks will be fine-tuned to automatically analyze images, enabling the precise identification of cancerous tissues. Image preprocessing, including segmentation techniques using Open CV, will be applied to enhance feature extraction. The trained model's performance will be evaluated using accuracy metrics and confusion matrix analysis, ensuring its effectiveness in histopathology image analysis. Once validated, the model will be deployed to classify test images, providing rapid and accurate cancer identification and severity analysis for clinical applications. The flowchart of our process is shown in Figure 3.1.

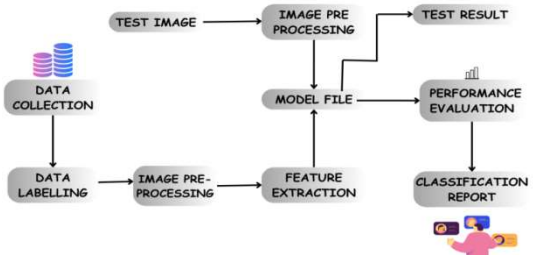

Figure 3.1 Proposed methodology's flowchart

Details about the Dataset: From the LC25000 datasets, a total of 10 thousand digital photographs of histopathology slides were available. The collection contains histological images of osteosarcoma stained with hematoxylin and eosin (H&E). A team of clinical scientists at the University of Texas Southwestern Medical Center in Dallas gathered the data. This dataset was created from archival samples from 50 patients treated at Children's Medical Center, Dallas, from 1995 to 2015 (Borkowski *et al*., 2019).

Table 3.1. The dataset's content.

CANCER TYPE	CLASS NAME	NO OF SAMPLES
Colon adenocarcinoma	colon_aca	5000
Colon benign tissue	colon_n	5000

It contains 500 images of colon tissue in total (250 images of benign colonic tissue and 250 images of colon adenocarcinomas), which have been augmented to 10,000 images using the Augmentor program (Borkowski *et al*., 2019). Detailed contents of the dataset are shown in Table 3.1. Colon adenocarcinoma is the most frequent colon cancer, accounting for almost 95% of all cases. Adenocarcinoma develops when a specific type of polyp (tissue growth) called adenoma forms inside the large intestine and eventually transforms into cancer. All images in the dataset are HIPAA-compliant, verified, and royalty-free. Some of the sample images of colon adenocarcinoma and colon benign tissue is shown in Figure 3.2 and Figure 3.3.

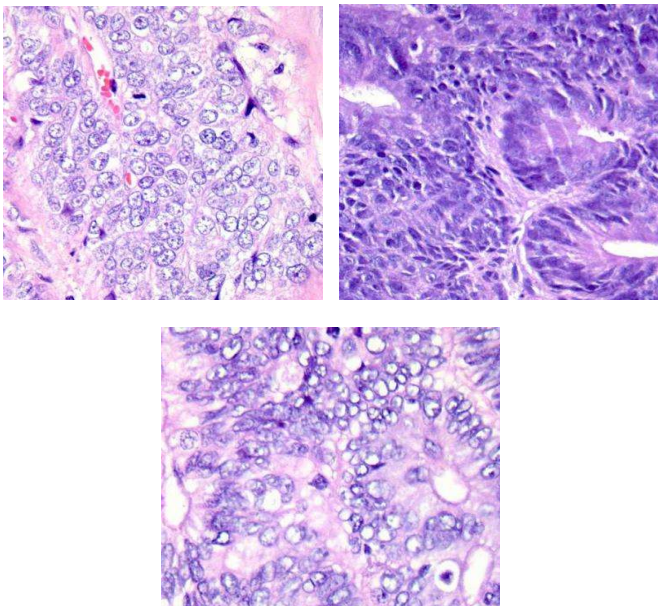

Figure 2. Colon adenocarcinoma

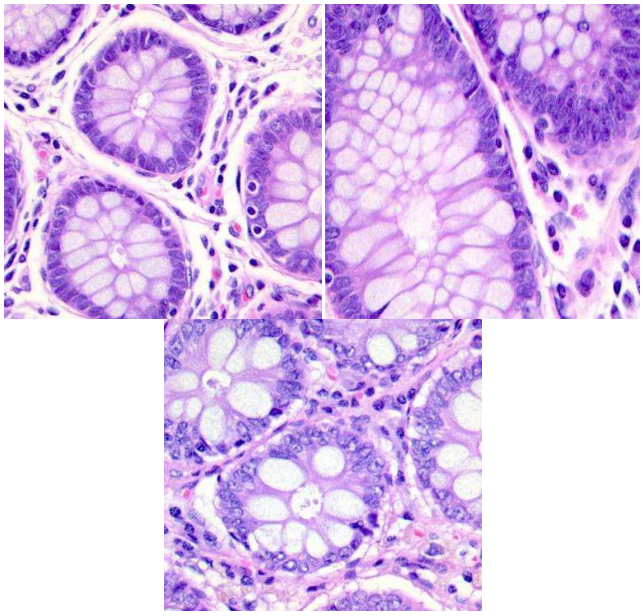

Figure 3. Colon benign

Preprocessing: The goal in preprocessing is to create images that are suitable for the following step of the detection system. Preprocessing data are the first and most critical step in preparing data for use with a deep learning model. To get a high classification rate in this study, have to eliminate noise and artifacts from the images. Using the libraries matplotlib and OpenCV, visualize and resize the images to the specified format suitable based on the computational speeds and capacity of the system.

In this preprocessing, the following filters are applied:

1. Resize
2. CLAHE
3. Median blur
4. Thresholding

***Image resizing*:** Image resizing refers to the scaling of images. Scaling comes in handy in many image processing as well as machine learning applications. It helps in reducing the number of pixels from an image and that has several advantages e.g. It can reduce the time of training of a neural network as the more the number of pixels in an image more is the number of input nodes that in turn increases the complexity of the model. It also helps in zooming in on images. Many time, it needed to resize the image i.e. either shrink it or scale it up to meet the size requirements. OpenCV provides us several interpolation methods for resizing an image.

```
img_array = cv2.resize(img_array, (250, 250))
```

***Clahe*:** Contrast Limited Adaptive Histogram Equalization (CLAHE) is used to equalize images. CLAHE is a variant of Adaptive histogram equalization (AHE) which takes care of over-amplification of the contrast. CLAHE operates on small regions in the image, called tiles, rather than the entire image. The neighboring tiles are then combined using bilinear interpolation to remove the artificial boundaries (Tomasz kryjak *et al*., 2022).

```
clahe = cv2.createCLAHE(clipLimit=100.0, tile GridSize=(8, 8))
        img_array = clahe.apply(img_array)
```

Clahe has one additional step over Adaptive Histogram Equalization and that is clipping of the histogram. The 5 steps in CLAHE are mentioned below:

i. Divide the image into tiny regions.
ii. Decide the mapping functions of local histogram.
iii. Choose the clipping point of histogram.

iv. Apply the function to every region.
v. Reduce the noise by the background subtraction method.

Median blur

Image: The image we need to apply the smoothening
Kernelsize: the size of the kernel as it always takes a square matrix the value must be a positive integer more than 2.

Syntax: cv. Median Blur (image, kernel size): In this method of smoothing, it will simply take the median of all the pixels inside the kernel window and replace the center value with this value. The one positive of this method over the gaussian and box blur is in these two cases the replaced center value may contain a pixel value that is not even present in the image which will make the image's color different and weird to look, but in case of a median blur though it takes the median of the values that are already present in the image it will look a lot more natural (Pythongeeks, 2021).

Thresholding: Image Thresholding is an intensity transformation function in which the values of pixels below a particular threshold are reduced, and the values above that threshold are boosted. This generally results in a bi-level image at the end, where the image is composed of black and white pixels. Thresholding belongs to the family of point-processing techniques. There are various ways of performing thresholding (Adaptive, Inverse, etc.), but the primary focus of this article will be on binary thresholding and would touch upon other thresholding methods in the end.

Binary Thresholding: The function takes in argument a source image, threshold at which the cutoff has to take place, maximum intensity value represented by the color space, the mode of thresholding and returns an integer value (denoting result of the operation) and an image object containing the resultant image after the processing.

Binary – Inverse Thresholding: In this, the output will be the inverse of above output i.e. white pixel will be black and vice-versa.

Proposed Dcnn Model: Convolutional Neural Network (CNN), also called ConvNet, is a type of Artificial Neural Net-work(ANN), which has deep feed-forward architecture and has amazing generalizing ability as compared to other networks with FC layers, it can learn highly abstracted features of objects especially spatial data and can identify them more efficiently. A deep CNN model consists of a finite set of processing layers that can learn various features of input data (e.g. image) with multiple level of abstraction . The initiatory layers learn and extract the high level features (with lower abstraction), and the deeper layers learns and ex-tracts the low level features (Anirudha Ghosh *et al*., 2020). CNN is composed of multiple building blocks (known as layers of the architecture). The structure of proposed DCNN model is depicted in Figure 3.4.

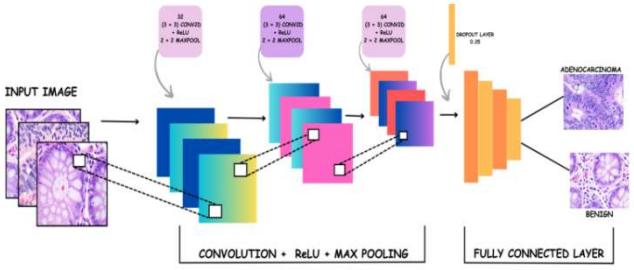

Figure 4. Proposed dcnn model's architecture

Convolutional layer: Convolutional layer is the most important component of any CNN architecture. It contains a set of convolutional kernels (also called filters), which gets convolved with the input image(N-dimensional metrics) to generate an output feature map. Mathematically, the convolution function is defined as follows (Junaid N. Z., 2021):

$$(f * g)(t) = \int_{-\infty}^{\infty} f(\tau)g(t-\tau)d\tau$$

Pooling layer: The pooling layers are used to sub-sample the feature maps (produced after convolution operations), i.e. it takes the larger size feature maps and shrinks them to lower sized feature maps.While shrinking the feature maps it always preserve the most dominant features (or information) in each pool steps. There are different types of pooling techniques are used in different pooling layers such as max pooling, min pooling,average pooling, gated pooling, tree pooling, etc. Max Pooling is the most popular and mostly used pooling technique (Anirudha Ghosh *et al*., 2020). Max Pooling selects the maximum element from each of the windows of the feature map. Thus, after the max-pooling layer, the output would be a feature map containing the most dominant features of the previous feature map (wikidocs, 2023).

Activation function (ReLU): The Rectifier Linear Unit (ReLU) is the most commonly used activation function in Convolutional Neural Networks. It is used to convert all the input values to positive numbers. The advantage of ReLU is that it requires very minimal computation load compared to others. The mathematical representation of ReLU is

$$f(x) = \max(0, x)$$

The main advantage of using the ReLU function over other activation functions is that it does not activate all the neurons at the same time. In ReLU function, if the input is negative it will convert it to zero and the neuron does not get activated (Abu Sufian, 2021).

Optimizer(Adam): Adaptive Moment Estimation (Adam) is among the top-most optimization techniques used today. In this method, the adaptive learning rate for each parameter is calculated. This method combines advantages of both RMSprop and momentum .i.e. stores decaying average of previous gradients and previously squared gradients (Anirudha Ghosh *et al*., 2020). Our proposed model chooses 0.1 learning rate for investigation.

Flatten layer: The Flatten layer in a neural network serves a crucial role in transforming the multi-dimensional output from preceding convolutional or pooling layers into a one-dimensional array, often referred to as a vector. As neural networks progress through various layers, convolutional layers are effective at capturing spatial hierarchies and patterns in input data. However, when transitioning to fully connected layers, which require one-dimensional input, a Flatten layer becomes necessary (Opengenus, 2021).

Dropout layer: Another typical characteristic of CNNs is a Dropout layer. The Dropout layer is a mask that nullifies the contribution of some neurons towards the next layer and leaves unmodified all others. Apply a Dropout layer to the input vector, in which case it nullifies some of its features. Dropout layers are important in training CNNs because they prevent overfitting on the training data. If they aren't present, the first batch of training samples influences the learning in a disproportionately high manner.

***Fully connected layer*:** Usually the last part (or layers) of every CNN architecture (used for classification) is consist of fully-connected layers, where each neuron inside a layer is connected with each neuron from it's previous layer. The last layer of Fully-Connected layers is used as the output layer (classifier) of the CNN architecture. The Fully-Connected Layers are type of feed-forward artificial neural network (ANN) and it follows the principle of traditional multi-layer perceptron neural network (MLP). The FC layer stake input from the final convolutional or pooling layer, which is in the form of a set of metrics (feature maps) and those metrics are flattened to create a vector and this vector is then fed into the FC layer to generate the final output of CNN.

Training and Validating the model: In the proposed model of DCNN, the dataset is divided into two sections: one for training and

another for validation. For training, 80% of dataset is used and for validation, 20% of dataset is used. During training, the CNN makes predictions for each input, and the difference between the predicted output and the actual label is calculated using a loss function. 'Sparse Categorical Cross-Entropy' is used as loss function in this model. The gradient of the loss with respect to the model parameters is computed, and the model's weights are updated using optimization algorithms (ADAM). The training process is repeated for multiple iterations which is known as epochs (15) to refine the model's parameters and reduce the loss. In this model DCNN is passed over 15 times the dataset is used. During validation, a portion of the labeled dataset, distinct from the training set, is set aside for validation (20%). This data is not used during training and helps assess the model's performance on unseen examples. Hyperparameters and model architecture are adjusted based on the performance on the validation set. This helps prevent overfitting (when the model memorizes the training data but fails to generalize to new data) by tuning the model to perform well on both training and validation data. Performance metrics like accuracy, precision, recall, and F1 score are calculated to quantify how well the model generalizes to data. Testing helps to ensure that the model has learned meaningful patterns from the data and can make accurate predictions on new, unseen samples.

RESULTS AND DISCUSSION

The primary objective of our proposed model is to classify benign and adenocarcinoma colon cancer tissue using DCNN. In training phase, the images are trained for 15 epochs with batch size 32. The proposed model achieved a training accuracy of 98.7% and validation accuracy of 94.6%. In figure (a) and (b) depict the accuracy and loss between the training and testing phases of the proposed DCNN, respectively. The following graphs show the epochs and the respective accuracy and loss.

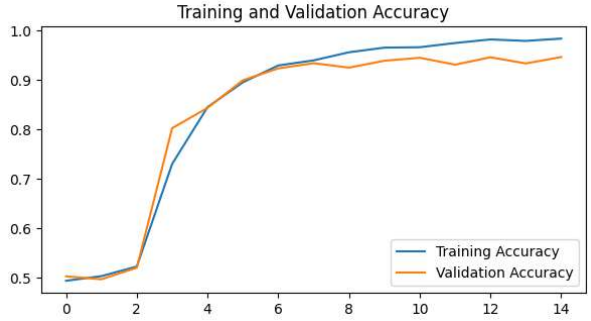

Figure (a)

Figure (b)

Figure 5. Proposed DCNN model's performance curve (a) accuracy curve and (b) loss curve

For each Epochs, the accuracy is increased and simultaneously loss is decreased. Table 4.1 shows the epochs range from 1 to 15 and how it reduces the loss and rapidly increases the accuracy for training and validation. These below table is plotted as curve Figure 4.1 (a) and (b).

Table 4.1. Epochs with training accuracy & loss and validation accuracy & loss

EPOCHS	TRAIN_ LOSS	TRAIN_ACC	VAL_LOSS	VAL_ACC
1/15	0.7136	0.4979	0.7012	0.4970
2/15	0.6977	0.5042	0.6971	0.5030
3/15	0.6779	0.552	0.5507	0.7400
4/15	0.4243	0.8048	0.3010	0.8785
5/15	0.2878	0.8809	0.2645	0.8890
6/15	0.2262	0.9122	0.2386	0.9050
7/15	0.1748	0.9349	0.2568	0.8915
8/15	0.1334	0.9488	0.2635	0.9030
9/15	0.1200	0.9528	0.2023	0.9255
10/15	0.0713	0.9751	0.1726	0.9475
11/15	0.0604	0.9779	0.1829	0.9455
12/15	0.0379	0.9862	0.1882	0.9445
13/15	0.0529	0.9814	0.1695	0.9455
14/15	0.0418	0.9840	0.2029	0.9455
15/15	0.0384	0.9879	0.1852	0.9430

In Figure 4.2 The confusion matrix is shown. The confusion matrix has 2 rows and 2 columns which consist of adenocarcinoma and benign tissues. This confusion matrix shows the TP, TN, FP, FN and performance metrics are manually calculated from these values for the comparison. Performance metrics such as accuracy, precision, recall, sensitivity, specificity, error rate, false positive rate, negative predicted value.

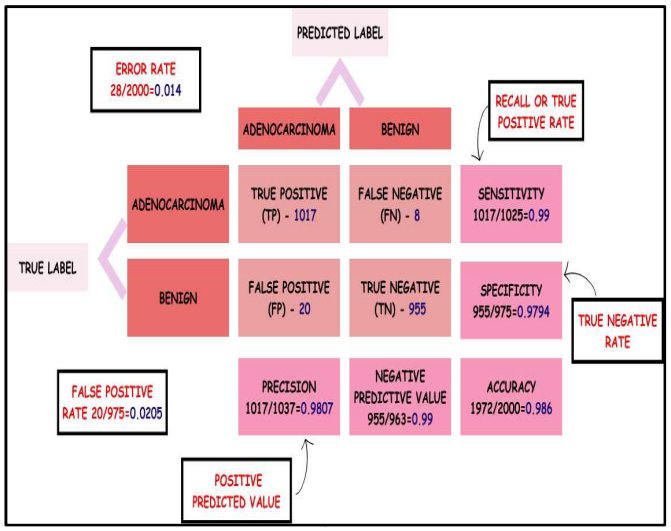

Figure 4.2. Confusion matrix for proposed model

In table 4.2, the classification report is shown. A classification report is a summary of key metrics that evaluate the performance of a classification model, especially in the context of multi-class classification problems. It provides detailed information on precision, recall, F1 score, and support for each class. The model obtains sensitivity: 0.980713596141755 and specificity: 0.9916926272 066459.

Table 4.2 Classification report

Classes	Precision	Recall	F1-Score	Support
Class 1 (aca)	0.98	0.99	0.99	1025
Class 2 (benign)	0.99	0.98	0.99	975
Accuracy			0.99	2000
Macro avg	0.99	0.99	0.99	2000
Weighted avg	0.99	0.99	0.99	2000

In table 4.3, the comparison between the performance metrics obtained from the model and calculated from formula is shown below:

Table 4.3 Comparison between the value obt. from model and formula

Performance metrics	Value obt. from model %	Value obt from formula%
ACCURACY	98.7	98.6
PRECISION	99	98.7
RECALL	99	99
F1 SCORE	99	98
SENSITIVITY	98.07	99
SPECIFICITY	99.16	97.94

Table 4.4 Comparisons with previous work

Authors	Year	Cancer t/p	Image type	Classifier	Accuracy%
Bukhari et al.	2020	Colon	histopathological	RESNET 50	93.91
Mangal s. *et al*.	2020	Colon	histopathological	CNN	96.61
Liang M. *et al*.	2020	Colon	histopathological	MFF CNN	96
Mehmood *et al*.	2022	Colon	histopathological	ALEXNET	89
Zhou P. *et al*.	2022	Colon	histopathological	HCCANET	87.3
Sameen A. *et al*.	2023	Breast	histopathological	IV-NET	97
Jyothi P. *et al*.	2023	Breast	histopathological	C2T2Net	95.68
Saeed I. *et al*.	2022	Breast	histopathological	BreastUNet	95
Irum H. *et al*.	2021	Breast	histopathological	DBN	86
Dan xue *et al*.	2020	cervical	histopathological	ETL	98.37
Chen li *et al*.	2019	cervical	histopathological	MHCRF	93
Mangal s. *et al*.	2020	Lung	histopathological	CNN	97.89
Hatuwal *et al*.	2020	Lung	histopathological	CNN	97.2
Masud M. *et al*.	2021	Lung	histopathological	CNN	96.33
Proposed model	2023	Colon	histopathological	DCNN	**98.7**

The DCNN model was trained using tensorflow framework on thonny platform. The proposed model's result is compared with some existing works, which is shown in Table 4.4.

CONCLUSION

In conclusion, this project successfully employed a comprehensive approach to colon cancer detection from histopathological images using Deep Convolutional Neural Networks (DCNN). The preprocessing phase played a crucial role in enhancing image quality and ensuring the robustness of the model. Employing a CNN model with rectified linear unit (ReLU) activation function and the ADAM optimization algorithm, our system demonstrated better accuracy in classifying histopathological images for colon cancer diagnosis. The incorporation of softmax activation for classification further refined our model's ability to distinguish between adenocarcinoma and benign tissue with high precision. The DCNN model was evaluated on the dataset named LC25000 in which obtained better training and testing accuracy respectively than other existing models. In this DCNN, the model get 98.7% accuracy, 99% precision, 99% recall, 99% F1 score. The achieved results underscore the potential of DCNNs in advancing early detection methods for colon cancer, contributing to improved patient outcomes and facilitating the integration of automated diagnostic tools into clinical practice. While our approach shows promise, ongoing research is essential to address challenges, and pave the way for practical implementation in real-world healthcare scenarios**.** The development of computer-supported technology for diagnosing malignant (adenocarcinoma) tumors will give pathologists a substantial amount of support.

Future scope: Most of the research related to colon cancer detection focuses on whether a given histopathological image is cancerous or non cancerous. So, the future research can plan to work on a larger dataset with more labeled colon tissue and build up a successful model with preprocessing processes to provide the best accuracy in prediction and classification. In the future research, focuses on detection the various stages of colon cancer from histopathological images. Collaborating with medical professionals, particularly pathologists, will likely increase. This interdisciplinary approach can lead to the development of more clinically relevant models and solutions and compare the results obtained from future model with real time data. Automation of the detection and classification of colon cancer from histopathological images has the potential to assist pathologists and improve the efficiency in crucial areas of cancer screening.

REFERENCES

1. Araghi M., Soerjomataram I., Jenkins M., "Global trends in colorectal cancer mortality: projections to the year 2035," International Journal of Cancer, vol. 144, no. 12, pp. 2992–3000, 2019.
2. Atlamazoglou V., Yova D., Kavantzas N., and Loukas S., "Texture analysis of fluorescence microscopic images of colonic tissue sections," Medical, & Biological Engineering & Computing, vol. 39, no. 2, pp. 145–151, 2001.
3. Aleksandrova K, Pischon T, Jenab M, Bueno-de-Mesquita HB, Fedirko V, Norat T, *et al*. Combined impact of healthy lifestyle factors on colorectal cancer: a large European cohort study. BMC Med. 2014;12:168.
4. Babu T., ''Colon cancer prediction on different magnified colon biopsy images,'' in Proc. 10th Int. Conf. Adv.Comput. (ICoAC), Chennai, India, Dec. 2018, pp. 277–280.
5. Borkowski A.A., Bui M. M., Thomas L. B., Wilson C. P., DeLand L. A., and Mastorides S. M., "Lung and colon Cancer Histopathological Image Dataset (Lc25000)," 2019.
6. Bukhari S. U. K., Asmara S., Bokhari S. K. A., Hussain S. S. , Armaghan S. U., and Shah S. S. H. , "The Histological Diagnosis of Colonic Adenocarcinoma by Applying Partial Self Supervised Learning," medRxiv, 2020.
7. Benard F, Barkun AN, Martel M, von Renteln D. Systematic review of colorectal cancer screening guidelines for average-risk adults: Summarizing the current global recommendations. World J Gastroenterol. 2018;24(1):124–38. (PMC free article) (PubMed)
8. Ben Hamida A, Devanne M, Weber J, Truntzer C, Derangere V, Ghiringhelli F, Forestier G, Wemmert C. Deep learning for colon cancer histopathological images

analysis. Comput Biol Med. 2021 Sep;136:104730. Epub 2021 Aug 4. PMID: 34375901.
9. Bray, F.; Ferlay, J.; Soerjomataram, I.; Siegel, R.L.; Torre, L.A.; Jemal, A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J. Clin. 2018, 68, 394–424. (CrossRef) (PubMed).
10. Clark K.,Vendt B., Smith K. *et al*., "The cancer imaging archive (tcia): maintaining and operating a public information repository," Journal of Digital Imaging, vol. 26, no. 6, pp. 1045–1057, 2013.
11. Das K, Conjeti S, Chatterjee J and Sheet D, "Detection of Breast Cancer From Whole Slide Histopathological Images Using Deep Multiple Instance CNN," in IEEE Access, vol. 8, pp. 213502-213511, 2020.
12. Erainnovator.com, 'Recurrent Neural Network', 2022.
13. Ghosh, Anirudha & Sufian, A. & Sultana, Farhana & Chakrabarti, Amlan & De, Debashis. (2020). Fundamental Concepts of Convolutional Neural Network. 10.1007/978-3-030-32644-9_36.
14. GLOBOCAN. Global Cancer Observatory (GCO): Cancer Today (accessed on 15 August 2021).
15. Hatuwal B. K. and Thapa H. C., ''Lung cancer detection using convolutional neural network on histopathological images,''Int. J. Comput.Trends Technol., vol. 68, no. 10, pp. 21–24, Oct. 2020.
16. Hirra I *et al*., "Breast Cancer Classification From Histopathological Images Using Patch-Based Deep Learning Modeling," in IEEE Access, vol. 9, pp. 24273-24287, 2021
17. Iqbal S. and Qureshi A. N., "A Heteromorphous Deep CNN Framework for Medical Image Segmentation Using Local Binary Pattern," in IEEE Access, vol. 10, pp. 63466-63480, 2022.
18. Kuipers EJ, Grady WM, Lieberman D, Seufferlein T, Sung JJ, Boelens PG, *et al*. Colorectal cancer. Nat Rev Dis Primers. 2015;1:15065.
19. Li C., "Cervical Histopathology Image Classification Using Multilayer Hidden Conditional Random Fields and Weakly Supervised Learning," in IEEE Access, vol. 7, pp. 90378-90397, 2019.
20. Liang M., Ren Z., Yang J., Feng W., and Li B., "Identification of Colon Cancer Using Multi-Scale Feature Fusion Convolutional Neural Network Based on Shearlet Transform," in IEEE Access, vol. 8, pp. 208969-208977, 2020.
21. Masud M. *et al*., ''A machine learning approach to diagnosing lung and colon cancer using a deep learning-based classification framework,'' Sensors, vol. 21, no. 3,p. 748, Jan. 2021.
22. Mangal S., ''Convolution neural networks for diagnosing colon and lung cancer histopathological images,'' 2020.
23. Mu T., and Nandi A. K., "Breast cancer detection from fna using svm with different parameter tuning systems and som–rbf classifier," Journal of the Franklin Institute, vol. 344, no. 3-4, pp. 285–311, 2007.
24. Mehmood S., *et al*., "Malignancy Detection in Lung and Colon Histopathology Images Using Transfer Learning With Class Selective Image Processing," in IEEE Access, vol. 10, pp. 25657-25668, 2022.
25. Maida M, Macaluso FS, Ianiro G, Mangiola F, Sinagra E, Hold G, *et al*. Screening of colorectal cancer: present and future. Expert Rev Anticancer Ther. 2017;17(12):1131–46.(PubMed).
26. Murphy N, Moreno V, Hughes DJ, Vodicka L, Vodicka P, Aglago EK, *et al*. Lifestyle and dietary environmental factors in colorectal cancer susceptibility, 2019.
27. NIH. Advances in Colorectal Cancer Research: Stages of Colorectal Cancer. Available online: https://www.nih.gov/researchtraining/advances-colorectal-cancer-research (accessed on 30 December 2021).
28. Opengenus.org, ' Convolutional neural network', 2021.
29. Peta J and Koppu S, "Enhancing Breast Cancer Classification in Histopathological Images through Federated Learning Framework," in IEEE Access, vol. 11, pp. 61866-61880, 2023.
30. Recio-Boiles A, Cagir B. Colon Cancer. StatPearls; Treasure Island (FL): 2022. (PubMed).
31. Sameen Aziz, Munir K, Raza K, Almutairi M.S. and Nawaz S, "IVNet: Transfer Learning Based Diagnosis of Breast Cancer Grading Using Histopathological Images of Infected Cells," in IEEE Access, vol. 11, pp. 127880-127894, 2023.
32. Sena P., Fioresi R., Faglioni, Losi, Faglioni, and Roncucci L., "Deep learning techniques for detecting preneoplastic and neoplastic lesions in human colorectal histological images," Oncology Letters, vol. 18, no. 6, pp. 6101–6107, 2019.
33. Sertel O., Kong J., Shimada H., Catalyurek U.V., Saltz J.H., and Gurcan M.N., "Computer-aided prognosis of neuroblastoma on whole-slide images: classification of stromal development," Pattern Recognition, vol. 42, no. 6, pp. 1093–1103, 2009.
34. Sandhu J, Lavingia V, Fakih M. Systemic treatment for metastatic colorectal cancer in the era of precision medicine. J Surg Oncol. 2019;119(5):564–82. (PubMed)
35. Sun Y. Advanced course in clinical oncology.1. Beijing: People's Military Medical Press; 2014. p. 889.
36. Simon K. Colorectal cancer development and advances in screening. Clin Interv Aging. 2016;11:967–76. (PMC free article) (PubMed).
37. Sung, H.; Ferlay, J.; Siegel, R.L.; Laversanne, M.; Soerjomataram, I.; Jemal, A.; Bray, F. Global Cancer Statistics 2020: GLOBOCAN. Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA Cancer J. Clin. 2021, 71, 209–249. (CrossRef).
38. Python Geeks, 'Averaging, Gaussian Blur and Median Blur in OpenCV', 2021.
39. Roncucci L, Mariani F. Prevention of colorectal cancer, Eur J Intern Med. 2015;26(10):752–6.2015 (PubMed).
40. Wikidocs.net, 'Convolutional Neural Networks - EN - Deep Learning', 2023.

****