J. Electrical Systems 20-01 (2024): 537-546

Kalpana K.[1*]

Dr. G. N. K. Suresh Babu[2]

AI-Based Deep Learning Model for Classification and Segmentation of Colorectal Cancer

Abstract: Colorectal cancer is the most common malignancy in the world, accounting for almost 35% of all cancer-related deaths. Colorectal cancer is third in terms of cancer diagnoses, behind only lung and breast cancers. In particular, AI-guided clinical treatment can aid in the reduction of health inequities. In modern times, digital pathology is crucial in the evaluation of tumors. Despite the abundance of extensively annotated datasets, current approaches struggle to handle the large size and high resolution of Whole Slide Images (WSIs). When it comes to histopathology image segmentation and tissue classification, these models appear like a promising option, especially considering the scalability of Deep Learning (DL) methods. Using DL architectures, this research focuses on colon cancer location classification and highlighting in a sparsely annotated histopathology data setting. To begin, we will examine and contrast many cutting-edge Convolutional Neural Networks (CNN). Due to the scarcity of high-quality WSI datasets, we have turned to transfer learning approaches. The technique's defining characteristic is the extensive collection of learnt features that is produced by training the network on a massive computer vision dataset. With VGG, we were able to achieve an accurate patch-level classification rate of up to 94.56% during testing and evaluation on our colon cancer dataset. This paper's proposed method outperformed the state-of-the-art algorithms for histopathology image classification, which had the lowest error rate. A simple, effective, and efficient method for histopathology image categorization. Through effective utilization of the dataset, we achieved state-of-the-art results.

Keywords: Colorectal Cancer, Diagnosis , Deep Learning, Classification, CNN

1. INTRODUCTION

Germany is one of the countries where colorectal (bowel) cancer ranks high in terms of incidence. Cancer of the colon (the large intestine) or the rectum is what the word "colorectal cancer" describes. Cancer of the small intestine is extremely uncommon. Colorectal polyps, which develop in the colon or rectum lining, are the most common cause of colorectal cancerHowever, not all polyps develop into cancer, and even if they do, the process takes many years. By removing polyps, a colonoscopy can help reduce the risk of colorectal cancer. The rectum or colon is the origin of colorectal cancer.Cancers of the colon or rectal region are other names for these diseases.

The similarities between colon and rectal cancer lead to their possible classification as a single disease. A part of the digestive system, frequently called the GI system, the large intestine (or big gut) is made up of the colon and rectum (see image below).The bulk of the large intestine is composed of the colon, a muscular tube that is about 5 feet (1.5 meters) long. The direction in which food travels through each segment of the colon is known as its "direction." The initial portion is called the ascending colon.

Undigested food from the small intestine enters the process in a pouch called the cecum. The transverse colon refers to the second section of the colon that travels along the right side of the belly. As it travels through the body, it sways from side to side. The movement to the left of the third segment is described by the descending colon.The sigmoid colon is the name given to this fourth portion because of its "S" shape. The function of the sigmoid colon is to connect the rectum to the anus. The proximal colon includes both the ascending and transverse parts. Part of the distal colon are the descending colon and the sigmoid colon.

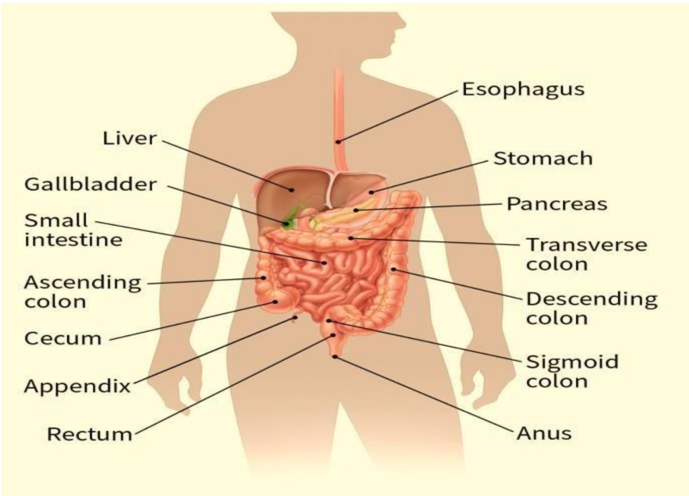


Figure 1. Colorectal Cancer

[1*]Research Scholar, Sristhi College of Science and Management, Bangalore
[2]Professor-CSE, Research Supervisor(University of Mysore), Sristhi College of Science and Management, Bangalore

1.1 Malignancies of the Rectum and Colon

The majority of cancers that affect the colon are known as adenocarcinomas. This type of cancer starts in the cells that normally make mucus, which helps keep the colon and rectum moist. This subtype of colorectal cancer is often brought up when doctors talk about the disease. Compared to other subtypes of adenocarcinoma, the prognosis (outlook) for mucinous and signet ring subtypes may be less favorable.
The rectum and colon can also be the sites of other, far less common tumor types.
Here are a few examples:

• Cancerous growths. At first, they are made in the intestine by special cells that create hormones. Gastrointestinal carcinoid tumors are to be seen.

• Nerve cells found in the lining of the gastrointestinal tract are the original source of gastrointestinal stromal tumors (GISTs). A few malignant cancers are actually quite harmless. The majority of these cancers are found in the stomach and small intestine. Their presence in the colon or rectum is unusual. See GIST, which stands for gastrointestinal stromal tumor.

• Lymphomas are cancers that affect cells in the immune system. Although lymph nodes are the most typical site of origin, other organs such as the colon or rectum might also serve as a source.

• Sarcomas can grow in a variety of connective tissues, including the colon and rectum walls, blood vessels, layers of muscle, and others. Information about lymphomas of the digestive system may be found in Non-Hodgkin Lymphoma. The diagnosis of rectum or colon sarcomas is quite rare. Refer to Sarcoma of Soft Tissue.

2. LITERATURE REVIEW

The objective of this investigation [1] is to enhance the early detection and diagnosis of colon cancer, a condition that is among the most prevalent causes of deaths worldwide. The study has determined that the gene expression levels of colon cancer patients can be used to differentiate between two groups using statistical hypothesis tests and machine learning methods. Researchers used t-tests, decision trees, neural networks, KNNs, and Mann-Whitney-Wilcoxon tests to identify the genes that were most important for colon cancer patients' vital state.
The gene expression of colon cancer genes is normalized twice, and an unsupervised learning method is employed to identify meaningful structures. Clustering patients according to their features derived from principle component analysis (PCA) is achieved through feature extraction from PCA.Genes that significantly influence colon cancer mortality rates across clusters are ultimately identified. The generalizability of its findings may be impacted by the relatively tiny dataset that it operates with. Gene expression data are the sole focus of the investigation, and lifestyle or environmental factors that may impact colon cancer mortality rates are not considered. This paper [2] investigates an intelligent imaging technology to diagnose colorectal cancer that is based on deep learning. In accordance with the investigation, this technology is capable of diagnosing colorectal cancer clinically. To diagnose colorectal cancer, this paper implements intelligent imaging technology that is based on deep learning. To compare the examined cancer sites, extract and distinguish tumor features, and input the collected data, an intelligent assistant diagnosis system with in-depth learning capacity was created.
The generalizability of the findings may be affected by the fact that the study was conducted on a small group of patients from Tangdu Hospital and lacks explicit information about the dataset used in the study. Its limitations are as follows. Additionally, the investigation did not evaluate the efficacy of intelligent imaging technology that is based on deep learning in comparison to other established diagnostic methodologies. A deep learning framework that is based on CNNsis proposed in this paper [3] for the classification of hyperspectral images. By employing convolutional, deconvolutional, and aggregating layers, hyperspectral data is optimized. An optimised extreme learning machine (ELM) is employed to conduct classification. No particular constraints are specified in the paper. Although the framework that has been proposed may not be suitable for all forms of hyperspectral data, it may necessitate additional optimization. Additionally, there is no comparison between the proposed framework and the most recent deep learning models. This article [4] suggests a single-shot detector framework for the detection of lesions in colonoscopy videos. As feature extractors, ResNet50, VGG16, and InceptionV3 were assessed. In this investigation, the datasets utilized were CVC-ColonDB and CVC-ClinicDB, which comprise 912 images from 44 video sequences from 36 patients. Using the ETIS-Larib dataset, an assessment of the proposed methodology was applied. This investigation is subject to constraints, including the intricate environment of the colon lumen, which complicates the detection of intestinal content, plica, or hemorrhagic foci by detectors. These restrictions may be surmounted through endoscopic cleansing.

To mitigate overfitting and mitigate noise and anomalies, the paper [5] implements DropBlock regularization techniques and Convolutional Neural Networks (CNNs). Transfer learning, a process that involves the reuse of previously trained networks for feature extraction, is employed in this approach. This method is assessed using two datasets: colorectal datasets and KVASIR datasets. Identical quantities of augmenting and validating sets are implemented for each set of experiments. In each instance, the architectures yielded comparable outcomes; however, ResNet50 demonstrated noteworthy performance. The proposed technique does not include any explicit limitations in the paper. At the same time, it is crucial to acknowledge that the methodology has only been assessed on two datasets and may not be applicable to other datasets. In addition, the paper fails to offer any assistance regarding the interpretability of the proposed methodology. This is an indispensable piece of medical image analysis. A method for detecting and segmenting colon lesions is described in the article [6] using mask regions convolutional neural networks (MRCNN) with precise region of interest (PrROI) pooling. Among the paper's shortcomings is the lack of comparisons with other cutting-edge approaches to colon polyp detection and segmentation. The lack of validation on large datasets also raises concerns about the method's generalizability. [7] used deep convolutional neural network modeling to try to build a histopathological characteristic for colon cancer lymph node metastasis (LNM). Although the study included 164 patients with stage I, II, or III colon cancer from the TCGA, the small sample size may have an impact on the study's generalizability.
There is a necessity for additional confirmation of the results through large prospective clinical trials.

3. PROBLEM STATEMENT

Based on the aforementioned literature review on optical colorectal imaging and colonic pathology, deep learning algorithms for colorectal polyp classification and segmentation are created. A data imbalance problem arises when there is a discrepancy in the quantity of samples in each class. The high volume of photos across several categories raises the possibility of overfitting the model. Starting from scratch with segmentation architecture training and definition is no picnic. Colonoscopy image retrieval is notoriously difficult due to the training procedure's requirement of an enormous number of images. Diagnosis may be challenging with colorectal optical images. Poor colorectal imaging quality is commonly caused by liquid interference and inadequate light. The borders of typical tissue polyps are also not very distinct. This may lead to a decrease in the accuracy of normal and polyp tissue classification and further confusion regarding the types of polyps and normal tissue borders. It leads to poor polyp and normal tissue classification accuracy. In order to get over these problems and improve the research's performance, the suggested strategy is applied.

4. PROPOSED MODEL

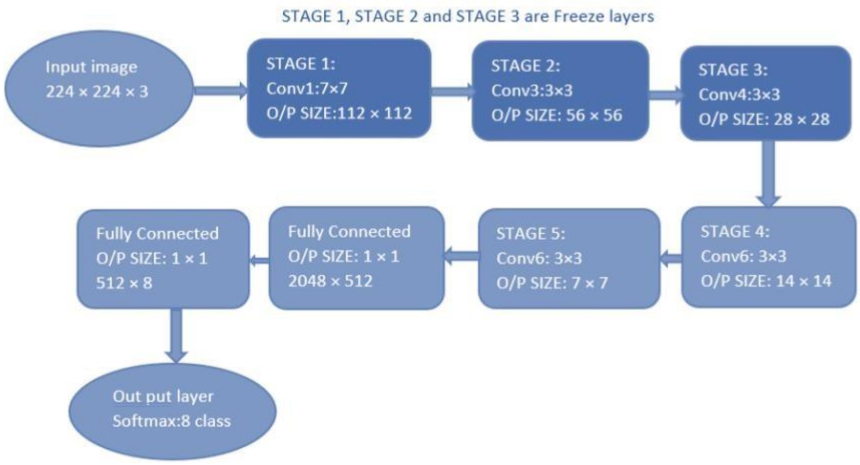


Figure 2. Proposed Model

Using the same dataset, we conducted experiments on eight different models. Nonetheless, we were able to recreate and show how each model performed. This can be shown in Figure 4 for the eight categories of colorectal cancer pictures. Classification into eight groups: Tumor, Stroma, Complex, Lympho, Debris, Mucosa, Adipose, and Empty

4.1 Data Set

A dataset of colorectal histology photographs is accessible on Kaggle.com. The following information is provided by Kather et al., 2019 [26]: the source is open source and includes compressed files. The dataset consists of 5,000 images, with each image measuring 224×224 pixels. After compression, each image is reduced to 64×64 pixels. The dataset is organized into eight non-overlapping classes: tumor, stroma, complex, lympho, debris, mucosa, adipose, an empty set, and empty. Utilizing an image augmentation technique, we had uniformly enhanced all eight categories of produced photos. There are multi-class tissue characteristics in this dataset, which has 20,000 distinct photos. At this point, one is using 60% of the 20,000 colorectal histopathology photos for training purposes, 10% for validation, and 30% for testing.

The dataset contains 5000 150x150 images of human colorectal cancer, with 8 different classes:

- Tumor
- Stroma
- Complex
- Lympho
- Debris
- Mucosa
- Adipose
- Empty (background)

4.2 Data preprocess

Making 2 numpy arrays for storing images and their labels. X=images, y=labels

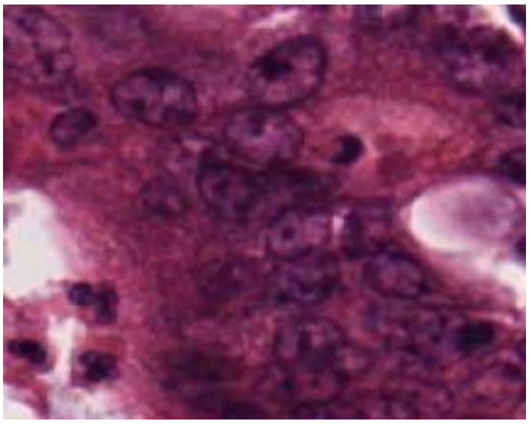


Figure 3 Colorectal Cancer

Splitting data into test and train

90% of dataset will be used for training, and the rest 10% will be used for testing.

Checking that train and test dataset both have all the 8 classes

making sure that each dataset has all classes, by using histogram for each dataset, before we begin to train the model

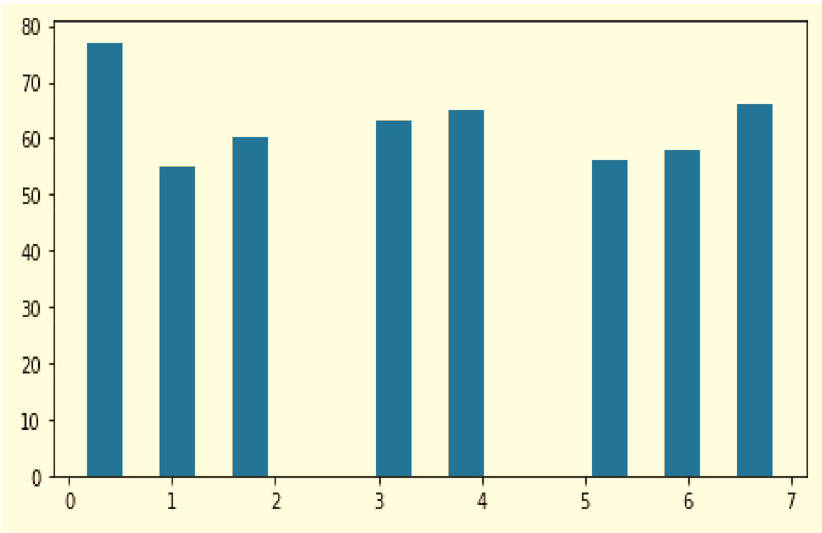


Figure 4. Test Dataset

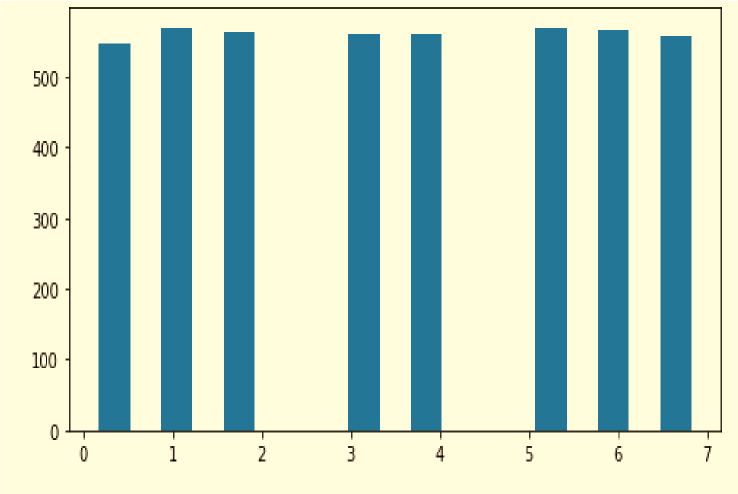


Figure 5 Train Dataset

4.3 Building Models

Building the first model (very simple custom model)

the first model will be a very simple model, just to see the difference between different models performance. this simple model wont get any good results, its input is connected straight to 8 dense output.

5. RESULTS AND DISCUSSION

VGG16

Transfer learning

We will use the VGG16 model, but we need to change its top layers to be able to feed images of different size to its network. we need to change the output of the convolutional stack accordingly to our data. In this way, we can apply the VGG16 architecture to images of our size. paste our own densely connected classifier on top of it: 256 dense layer -> dropout 0.5 -> 8 dense softmax outputfreezing the first layer of the network to prevent the weights in this layer from being updated during training. train accuracy = 0.9068889021873474
test accuracy = 0.8740000128746033

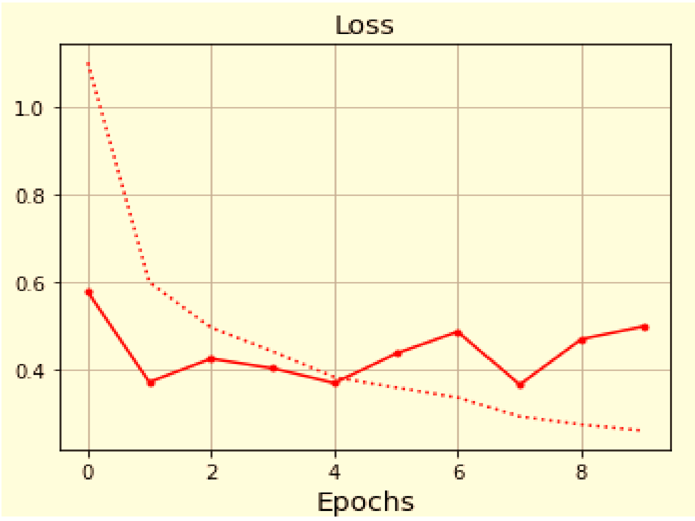


Figure 6 VGG16 Loss

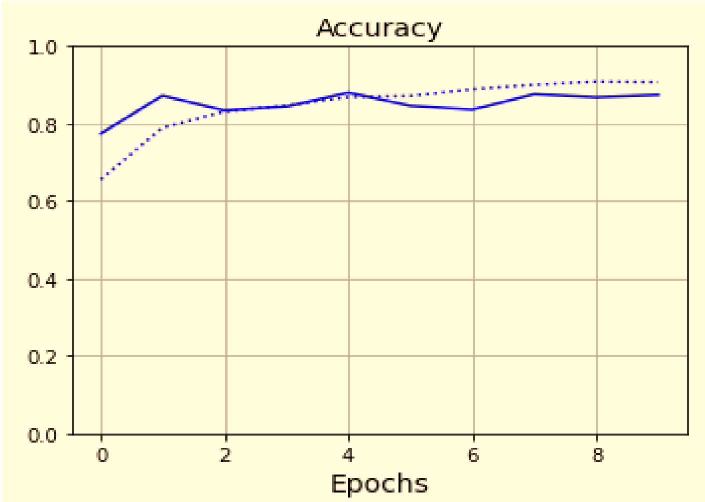


Figure 7 VGG16 Accuracy

VGG16 with Data Augmentation

train accuracy = 0.8903312683105469

test accuracy = 0.8759999871253967

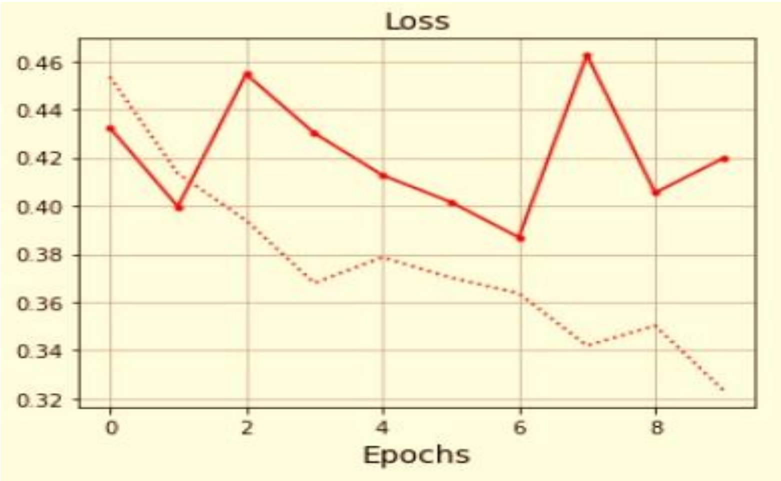


Figure 8 VGG16 with Augmentation Loss

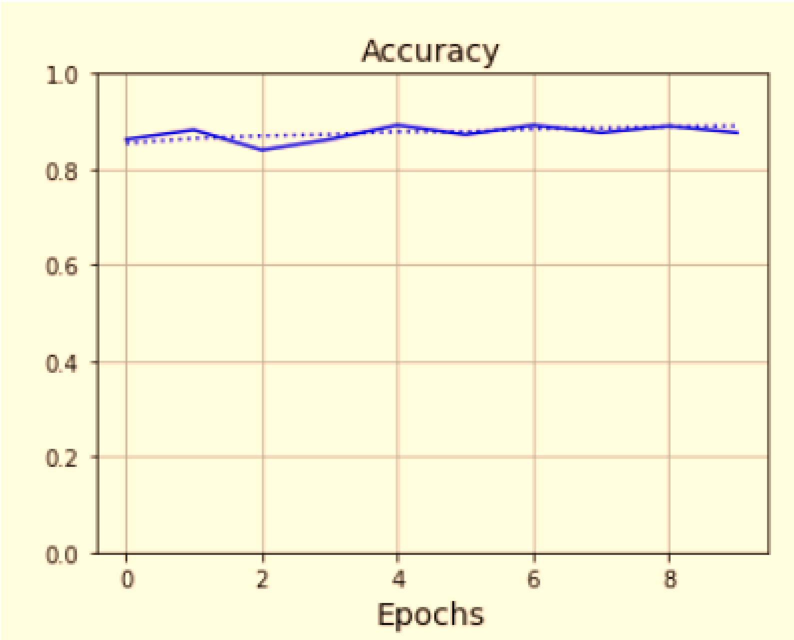


Figure 9. VGG16 with Augmentation Accuracy

Model Fine Tuning
train accuracy = 0.9456132650375366
test accuracy = 0.9240000247955322

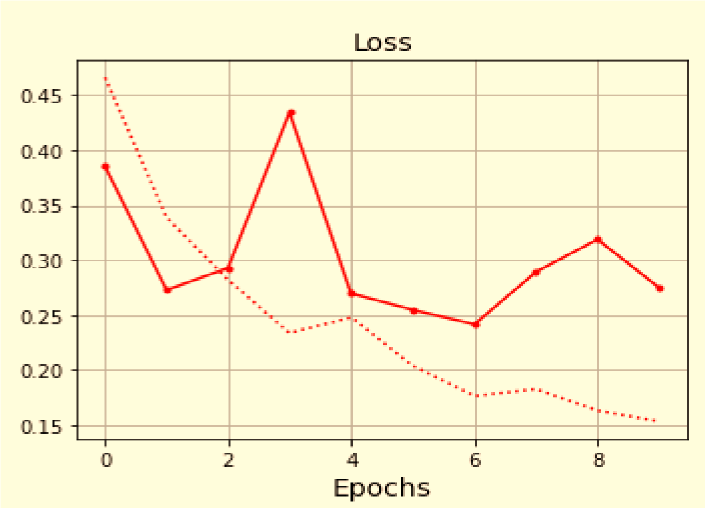


Figure 10 Loss

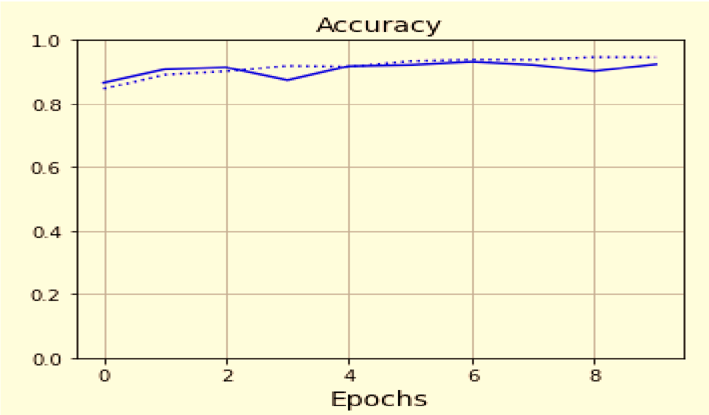


Figure 11 Accuracy

Overall models accuracy results:

Model	Accuracy Train DS	Accuracy Test DS
VGG16	90.68	87.40
VGG16 + Data Augmentation	89.03	87.59
Proposed Approach	94.56	92.40

Predicting
Using 'VGG16 like' model to predict

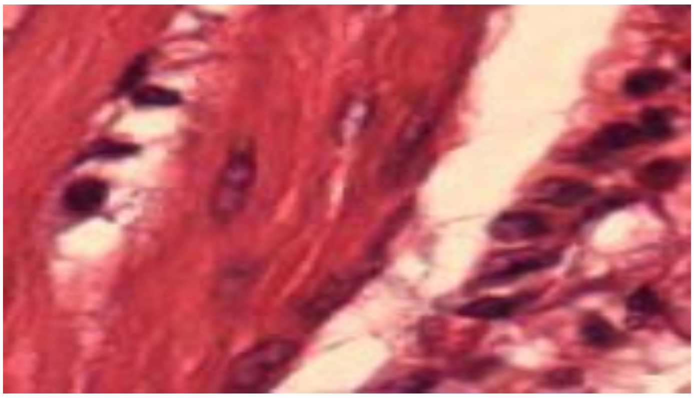


Figure 12 VGG16

Confusion Matrix

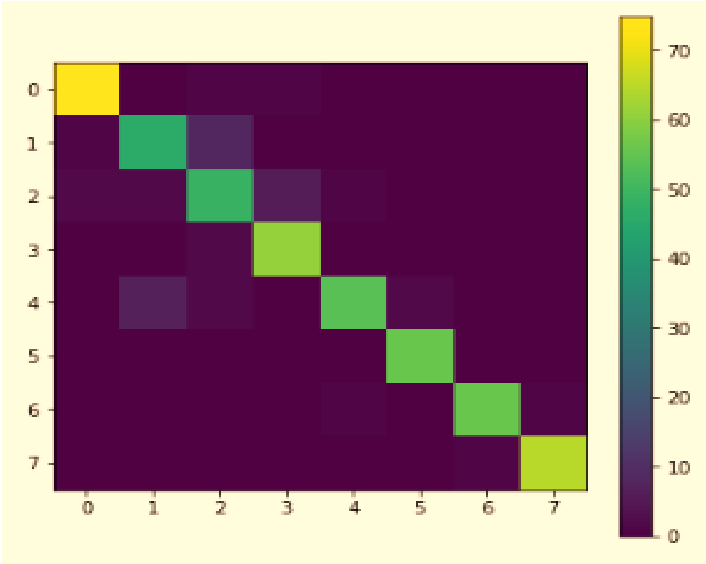
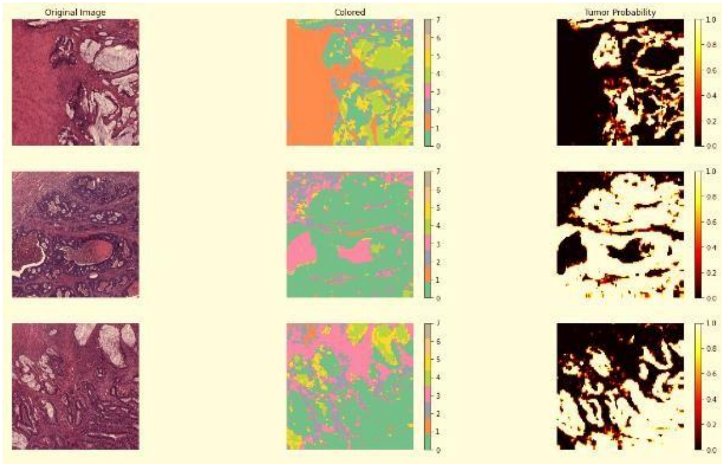
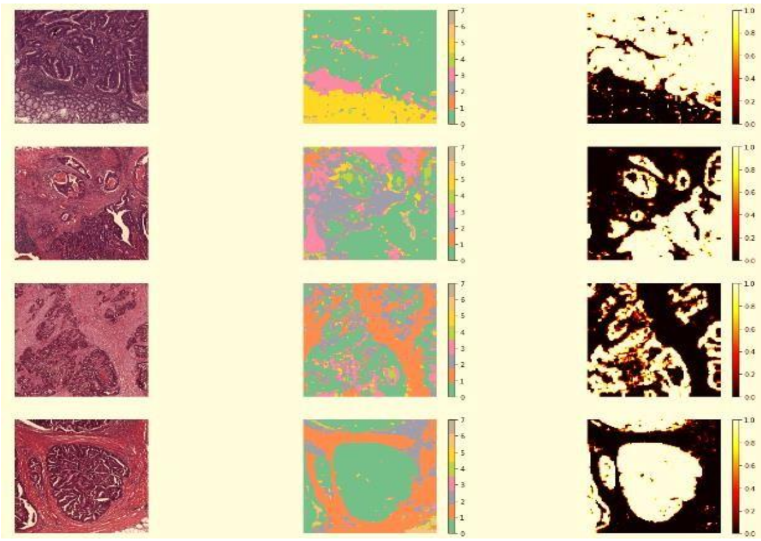


Figure 13 Confusion Matrix

Load 'colorectal_histology_large' dataset

The dataset contains 10 5000x5000 images of human colorectal cancer





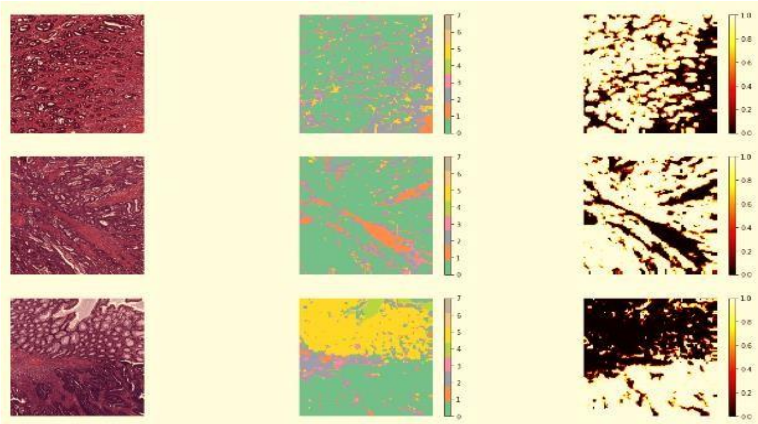


Figure 14 Proposed Approach Prediction of Colorectal Cancer

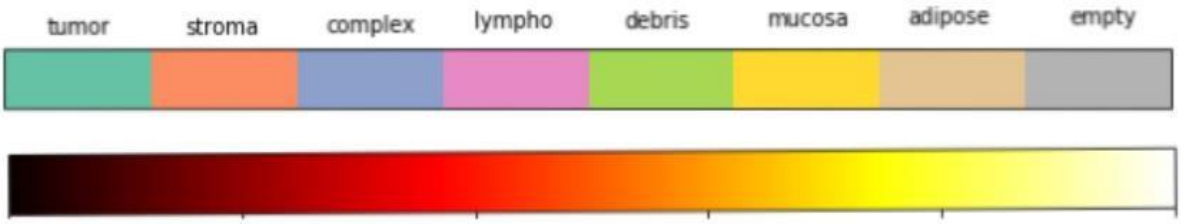


Figure 15 Prediction of Colorectal Cancer

5.CONCLUSION

These results show that colorectal histopathology scans can consistently detect colon cancers. It may be said that the outcomes were better than the previous models. Using the transfer learning approach, we want to enhance our future work. This indicates that there is a pressing need for further evidence to back up the outcomes of colorectal cancer AI diagnosis in order to optimize the performance of the models for validation at the practice level.

REFERENCES

[1] Egeblad, M.; Nakasone, E.S.; Werb, Z. Tumors as Organs: Complex Tissues that Interface with the Entire Organism. Dev. Cell 2010, 18, 884–901.

[2] Gurcan, M.N.; Boucheron, L.E.; Can, A.; Madabhushi, A.; Rajpoot, N.M.; Yener, B. Histopathological Image Analysis: A Review. IEEE Rev. Biomed. Eng. 2009, 2, 147–171.

[3] Janowczyk, A.; Madabhushi, A. Deep learning for digital pathology image analysis: A comprehensive tutorial with selected use cases. J. Pathol. Inform. 2016, 7, 29.

[4] Kather, J.N.; Krisam, J.; Charoentong, P.; Luedde, T.; Herpel, E.; Weis, C.-A.; Gaiser, T.; Marx, A.; Valous, N.A.; Ferber, D.; et al. Predicting survival from colorectal cancer histology slides using deep learning: A retrospective multicenter study. PLoS Med. 2019, 16, e1002730.

[5] Kather, J.N.; Weis, C.-A.; Bianconi, F.; Melchers, S.M.; Schad, L.R.; Gaiser, T.; Marx, A.; Zöllner, F. Multi-class texture analysis in colorectal cancer histology. Sci. Rep. 2016, 6, 27988.

[6] Korbar, B.; Olofson, A.M.; Miraflor, A.P.; Nicka, C.M.; Suriawinata, M.A.; Torresani, L.; Suriawinata, A.A.; Hassanpour, S. Deep learning for classification of colorectal polyps on whole-slide images. J. Pathol. Inform. 2017,

[7] M. A. Fahami, M. Roshanzamir, N. H. Izadi, V. Keyvaniand R. Alizadehsani, “Detection of effective genes in colon cancer: A machine learning approach”, vol. 24, Jan. 2021.

[8] 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.

[9] 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.

[10] 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.

[11] 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,

[12] T. Yang, "Intelligent Imaging Technology in Diagnosis of Colorectal Cancer Using Deep Learning", vol. 7, Dec. 2019.
[13] Tsai, M.J.; Tao, Y.H. Machine Learning Based Common Radiologist-Level Pneumonia Detection on Chest X-rays. In Proceedings of the 2019 13th International Conference on Signal Processing and Communication Systems (ICSPCS), Gold Coast, Australia, 16–18 December 2019.
[14] Wunsch G, Gourbin C. "Mortality, morbidity and health in developed societies: a review of data sources." Genus. 2018;74(1):2.
[15] 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.
[16] Xu, J.; Luo, X.; Wang, G.; Gilmore, H.; Madabhushi, A. A Deep Convolutional Neural Network for segmenting and classifying epithelial and stromal regions in histopathological images. Neurocomputing 2016, 191, 214–223.
[17] Zhang, X.; Su, H.; Yang, L.; Zhang, S. Fine-grained histopathological image analysis via robust segmentation and large-scale retrieval. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 5361–5368.
[18] Zhang,S.; Metaxas, D. Large- scale medical image analytics: Recent methodologies, applications and future directions. Med. Image Anal. 2016, 33, 98–101.