Al-Farabi Journal of Engineering Sciences Volume (4), Issue (2) December (2025)
ISSN: 2617-5517 (issn.org)
Al-Farabi Journal of Engineering Sciences
https://iasj.rdd.edu.iq/journals/journal/view/97
مجلة الفارابي للعلوم الهندسية تصدرها جامعة الفارابي

COLON CANCER DETECTION FROM HISTOLOGICAL IMAGE DATASET USING DEEP LEARNING ALGORITHMS

Mohaimen oday hazim alrawi1
1 Department of Biomedical Engineering, Near East University, Nicosia, Mersin 10, Turkey

ABSTRACT

Cancer is one of the most dangerous and widespread diseases in the world, including colon cancer, notably impacting those over the age of fifty compared to younger populations. Prompt identification and detection are crucial for mitigating the progression of the disease and facilitating effective therapy. This research employs an image-data technique, microscopic imagery pertinent to colon cancer. The research employs three separate methods for analysis. For the analysis of microscopic imaging data related to colon cancer. It utilized the ResNet18, EfficientNet-82, and MobileNet-55 algorithms. The ResNet algorithm attained an accuracy of 99.65% on the test data and 98.80% on the training data, whereas the Efficient Net algorithm achieved 100% accuracy on the test data and 99.31% on the training data. Lastly, the Mobile Net algorithm recorded 100% accuracy on the test data and 98.96% on the training data.

The main purpose of the study was to employ artificial intelligence in the process of classifying colon cancer image data in order to improve accuracy and overall results, speed up the screening process, reduce costs, and facilitate disease recognition. The primary objective of the investigation was to implement artificial intelligence in the classification of colon cancer image data to enhance accuracy and overall outcomes, expedite the screening process, reduce expenses, and simplify disease identification.

1. INTRODUCTION

Colon cancer is among the most common cancers worldwide, primarily affecting people over the age of 50[1]. Most cases have been discovered in North America, Australia, Northern Europe, and Western Europe. In poorer countries, rates are typically lower. The colon consists of four anatomical segments: the rectum, the descending colon, the transverse colon, and the sigmoid colon. Polyps are an early warning sign of colon and rectal cancer and should be removed to lower the risk of infection. [2] [3]. The accuracy, sensitivity, and other parameters used in evaluating colon cancer outcomes have been improved using artificial intelligence techniques and methods, which interpreted the image and digital data in an excellent and remarkable way [4].

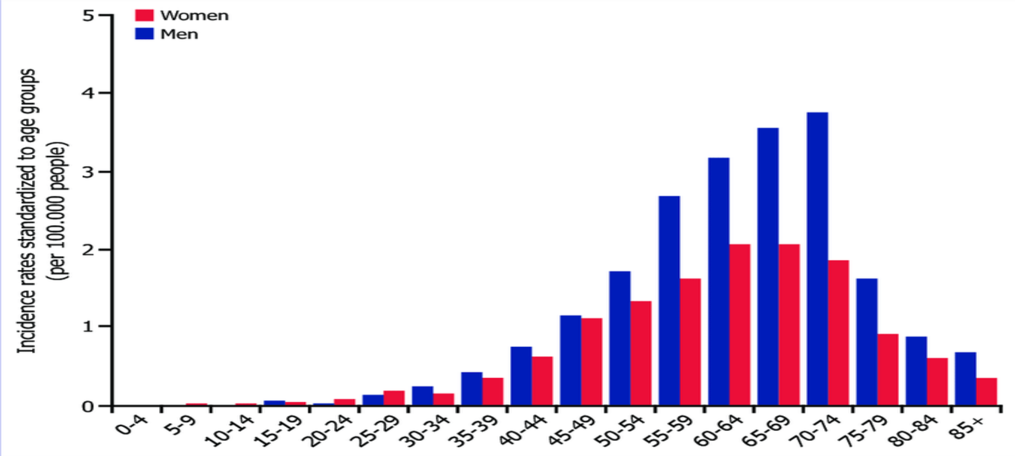

Figure 1-1 illustrates the infection rate among different age groups in Türkiye [5].

The advent of AI has catalyzed a profound scientific revolution in medicine and human health-related research. This has improved detection accuracy and enabled professionals to achieve a comprehensive understanding of an individual's health and disease risk through the study of imaging data [6]. Artificial intelligence is commonly used to recognize features in tissue images, as well as to identify diseases using digital and genetic data, such as sequences, numbers, or reports on a dedicated lifestyle. All of this enables artificial intelligence to recognize data and train on it

[7]. Artificial Intelligence (AI) has proven to be highly beneficial in improving image quality, enhancing diagnostic accuracy, and minimizing noise through the utilization of data obtained from CT scans, X-rays, and microscopic images [8] [9]. Transfer learning is a machine learning technique wherein a model created for one task is employed as the foundation for a subsequent task. This approach is beneficial when the secondary task is relevant to the primary one or when data for the secondary task is limited, hence minimizing training duration. By utilizing features obtained from prior work, the model may adapt more efficiently to the new task, hence improving learning speed and performance. Transfer learning alleviates the risk of overfitting, as the model possesses pre-existing generalizable attributes advantageous for the ensuing task [12]. The data in the research paper comes from a Kaggle dataset containing images of colon adenocarcinoma tissue. The dataset contains 5,000 images of this malignant tumor and 5,000 images of benign tumors.

2. METHOD

2.1 DATA COLLECTION

The data in the research paper comes from a Kaggle dataset containing images of colon adenocarcinoma tissue. The dataset contains 5,000 images of this malignant tumor and 5,000 images of benign tumors.

2.2 PREPROCESSING

Before inputting the images into the model, the RESIZE technique, which was set to 224 × 224 pixels, was employed to ensure that all of the images were of a uniform size. The mean and standard deviation of the color channels of the microscope images were standardized, meaning that they were transformed so that they had a mean of [0.485, 0.456, 0.406] and a standard deviation of [0.229, 0.224, 0.225]. In order to guarantee that both the training dataset and the test dataset were appropriate for analysis, this normalization technique was applied in a consistent manner to each of them. [10].

2.3 DATA AUGMENTATION

Various strategies are employed employing Torch vision for the transformation of training and testing datasets in microscope image data. Transformations, such as Random Horizontal Flip, are employed in the training set to invert images, thereby enhancing data diversity randomly. Random Resized Crop is employed to crop photographs to the specified dimensions, while Random Rotation is utilised to rotate images by 10 degrees (Cook, 2021). All picture data were normalised via the standard ImageNet, and the standard deviation was incorporated to maintain uniformity.

2.4 DATA SPLIT

A total of twenty percent of the microscopic images related to colon cancer were designated for testing, while seventy percent were allocated for training and ten percent for validation purposes.

2.5 DATA TRAINING

There were five thousand microscopic images of adenocarcinoma that is associated with colon cancer and five thousand photographs of non-malignant illnesses in the collection. The dataset was divided into three sections: 70 percent was allocated for model training, 20 percent was allocated for testing, and 10 percent was allocated for validation. The data was analysed using three deep learning algorithms: an 18-layer ResNet, an 82-layer Efficient Nett, and a 55-layer Mobile Nett. For the data, we utilised a batch size of 32 and trained for a total of five epochs. Each one of these models achieved a high level of accuracy in terms of classification accuracy, precision, recall, and F1 score.

2.6 DATA TESTING

We used three different methods to test the model's performance on the microscopic picture data: ResNet with 18 layers, EfficientNet with 82 layers, and MobileNet with 55 layers. We used 20% of the dataset for this. The results showed that the model did a good job overall.

2.7 DATA VALIDATION

The model's efficacy was assessed using the following methods: ResNet with 18 layers, EfficientNet with 82 layers, and MobileNet with 55 layers, with ten percent of the microscopic image data being used. The model's overall performance was commendable, as indicated by the results.

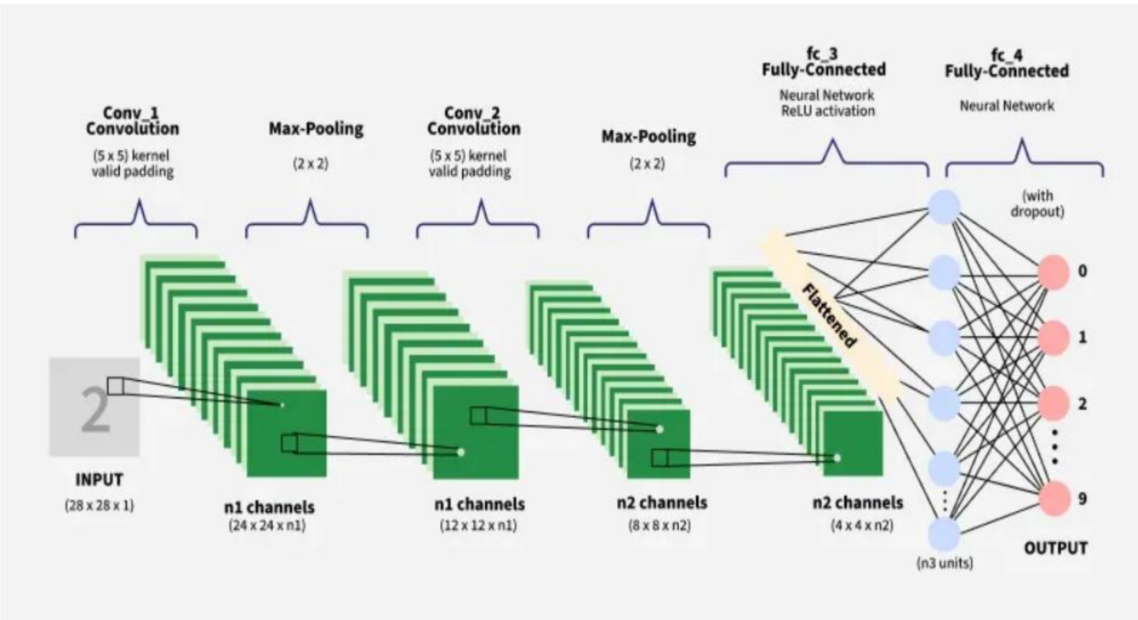

Figure 2-1. CNN Algorithm Architecture[11].

3. RESULTS AND DISCUSSION

This chapter examines the empirical outcomes derived from training three deep learning algorithms: ResNet-18, EfficientNet-82, and MobileNet-52, on picture data pertaining to adenocarcinoma colon cancer, comprising 5000 infected photos and 5000 non-infected images. The model was assessed using multiple metrics, specifically accuracy, sensitivity, F1 score, and recall.

Test Accuracy	Train Accuracy	Validation Accuracy
99.65%	98.80%	99.20%

Figure 3.1 result for ResNet-18

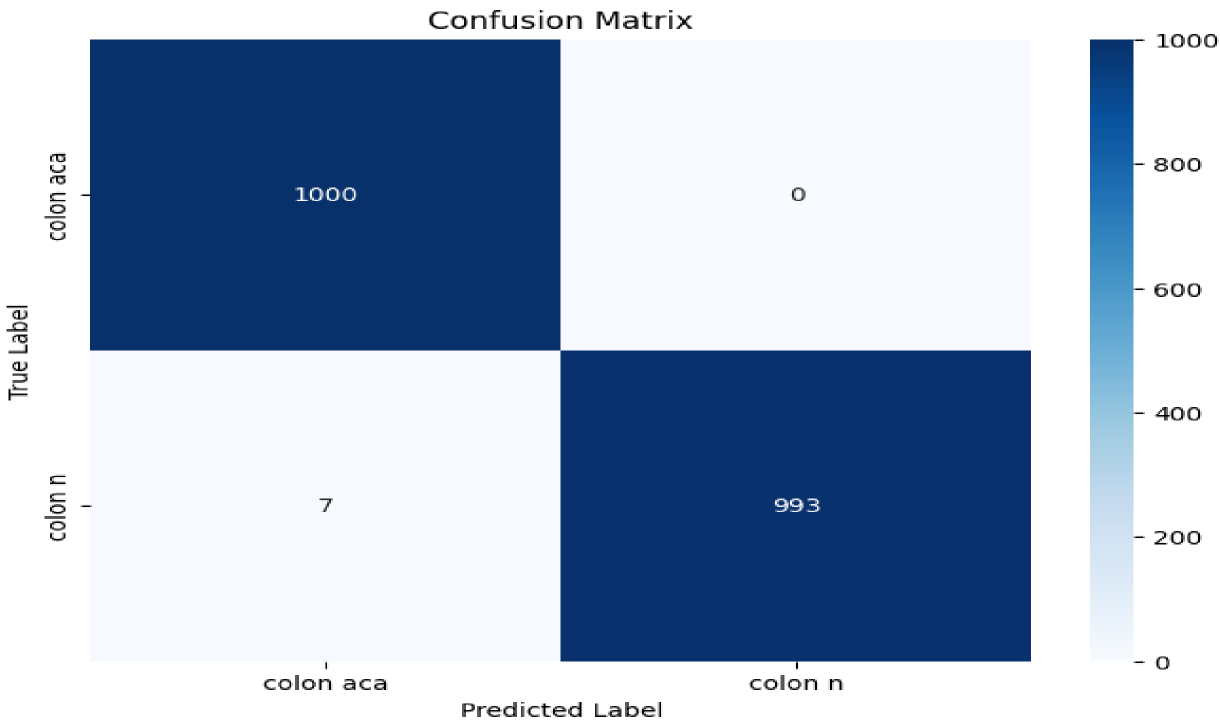

Figure 3.2 confusion matrix of ResNet-18 for testing data

Test Accuracy	Train Accuracy	Validation Accuracy
100%	99.31%	100%

Figure 3.3 result for Efficient Net-52

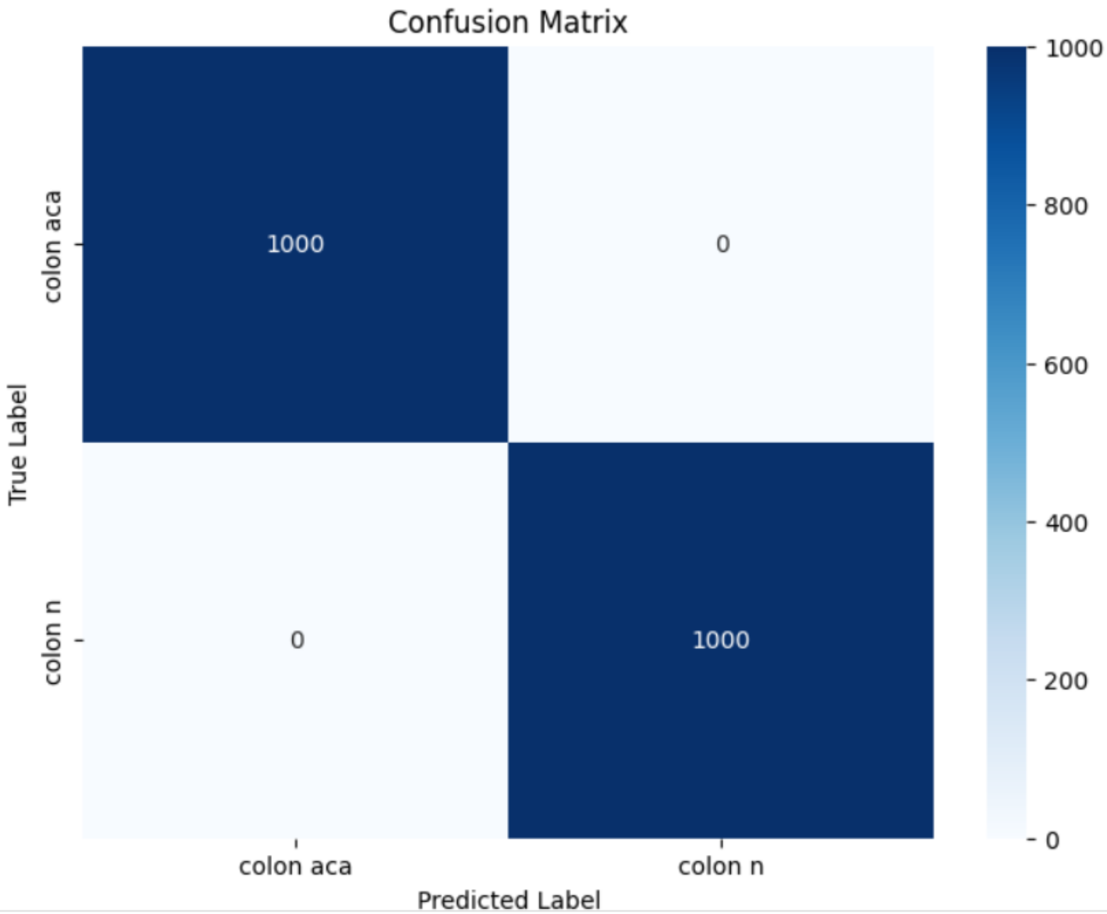

Figure 3.4 confusion matrix of Efficient Net-52 for testing data

Test Accuracy	Train Accuracy	Validation Accuracy
100%	98.96%	99.90%

Figure 3.5 result for Mobile Net- 82

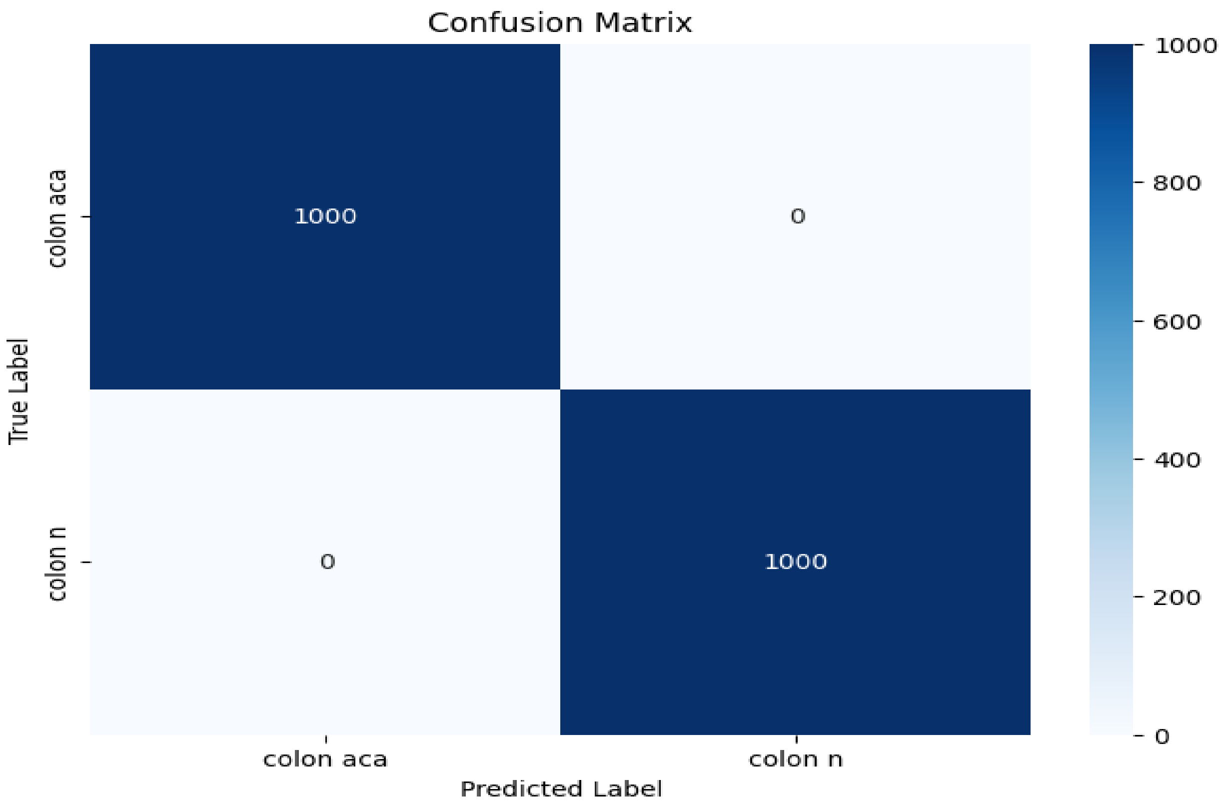

Figure 3.4 confusion matrix of Mobile Net-82 for testing data

CONCLUSION

This paper shows how well microscopic imaging can classify colorectal cancer, getting very good results across a wide range of data processing methods and giving a full analysis of the outcomes. Microscopic picture data was put into groups using the models ResNet, MobileNet, and EfficientNet. All three showed high levels of accuracy across all methods used. These investigations have a lot of potential and benefits for healthcare workers. They could speed up diagnoses, lower costs, and give doctors very reliable data. Furthermore, they stress the huge potential of artificial intelligence (AI) in the classification of histological imaging, which would allow doctors and patients to make smart choices based on the data's insights.

REFERENCES

[1] Brown JS, Amend SR, Austin RH, Gatenby RA, Hammarlund EU, Pienta KJ. Updating the definition of cancer. Molecular Cancer Research [Internet]. 2023b Jul 6;21(11):1142–7. Available from: https://doi.org/10.1158/1541-7786.mcr-23-0411.

[2] Engstrom PF, Arnoletti JP, Benson AB, Chen YJ, Choti MA, Cooper HS, et al. Colon cancer. Journal of the National Comprehensive Cancer Network [Internet]. 2009d Sep 1;7(8):778–831. Available from: https://doi.org/10.6004/jnccn.2009.0056.

[3] Haq I, Mazhar T, Asif RN, Ghadi YY, Saleem R, Mallek F, et al. A deep learning approach for the detection and counting of colon cancer cells (HT-29 cells) bunches and impurities. PeerJ Computer Science [Internet]. 2023c Dec 5;9:e1651. Available from: https://doi.org/10.7717/peerj-cs.1651.

[4] Patel SK, George B, Rai V. Artificial intelligence to decode cancer mechanism: Beyond patient stratification for precision oncology. Frontiers in Pharmacology [Internet]. 2020 Aug 12;11. Available from: https://doi.org/10.3389/fphar.2020.01177.
[5] Cinar D, Tas D. Cancer in the elderly. Northern Clinics of Istanbul [Internet]. 2015c Jan 1; Available from: https://doi.org/10.14744/nci.2015.72691.
[6] Sarumi UA, Ameen ZS, Al-Turjman F, Altrjman C, Mubarak AS. A novel attendance system via integrated Wifi and blockchain technologies. 2022 International Conference on Artificial Intelligence in Everything (AIE) [Internet]. 2022 Aug 1;209–15. Available from: https://doi.org/10.1109/aie57029.2022.00046.
[7] Sancar N, Okonkwo MOO, Onakpojeruo EP, Selma JS, Fannoh JJ, Ibrahim I, et al. A comparative study of statistical models for forecasting Covid-19 cases in Greece. AIP Conference Proceedings [Internet]. 2024 Jan 1;3085:020029. Available from: https://doi.org/10.1063/5.0194723.
[8] Lee D, Yoon SN. Application of Artificial Intelligence-Based Technologies in the Healthcare Industry: Opportunities and challenges. International Journal of Environmental Research and Public Health [Internet]. 2021c Jan 1;18(1):271. Available from: https://doi.org/10.3390/ijerph18010271.
[9] Alfadhel M. Artificial Intelligence in genomic Analysis: Revolutionizing healthcare. Journal of Biochemical and Clinical Genetics [Internet]. 2023c Jan 1;089–90. Available from: https://doi.org/10.24911/jbcgenetics/183-1709450443.
[10] Cook B. TorchVision Transforms: Image preprocessing in PyTorch [Internet]. Sparrow Computing. 2021b. Available from: https://sparrow.dev/blog/torchvision-transforms.
[11] GeeksforGeeks. Convolutional Neural Network (CNN) in machine learning [Internet]. GeeksforGeeks. 2025a. Available from: https://www.geeksforgeeks.org/deep-learning/convolutional-neural-network-cnn-in-machine-learning/.
[12] Al-Timemy AH, Alzubaidi L, Mosa ZM, Abdelmotaal H, Ghaeb NH, Lavric A, et al. A Deep Feature Fusion of Improved Suspected Keratoconus Detection with Deep Learning. Diagnostics [Internet]. 2023 May 10;13(10):1689. Available from: https://doi.org/10.3390/diagnostics13101689.

BIOGRAPHIES OF AUTHORS (10 PT)

The recommended number of authors is at least 2. One of them as a corresponding author.
Please attach clear photo (3x4 cm) and vita. Example of biographies of authors:

Author 1 	Mohaimen Alrawi Fields of Study: Bachelor's in Bioengineering (2016-2021), Master's in Biomedical Engineering (2023-2025) Role: Computer Engineer, Embedded Engineer. 20226969@ std.neu.edu.tr