Lung and colon cancer recognition using Histopathological Image classification: CNN based approach

Tamanna Ferdaus
Dept. of Computer Science & Engineering
University of Barishal
Barishal, Bangladesh

Dr. Tania Islam
Dept. of Computer Science & Engineering
University of Barishal
Barishal, Bangladesh

***Abstract*—Cancer is a dreadful disease that refers to abnormal growth of cell tissue. Every year, millions of people die due to delayed recognition or late-stage detection of the disease. Therefore, accurate identification of cancer cells becomes most important. Considering the number of people affected by cancer, there is an urgent need to find better ways of diagnosis and treatment methods. Neural networks are a burning research area in medical science. A variety of these techniques as Decision Trees (DTs), Artificial Neural Network (ANN) and Support Vector Machines (SVM) have been widely applied in cancer detection. Medical image processing relies heavily on image segmentation due to the diverse nature of medical images. This paper introduces a classification method for distinguishing cancer cells, employing deep learning Convolutional Neural Networks (CNNs).**

***Index Terms*—Image Segmentation, CNN, Histopathological Image, Cancer**

I. Introduction

Cancer is a leading cause of death worldwide, and previous statistics prediction says that there will be nearly 2 million deaths in 2023 approximately [1]. Cancer has the potential to impact various organs within the human body. Nevertheless, the most frequently affected regions include the colon, lungs, liver, breasts, rectum, brain, prostate, stomach, and skin. Among both males and females, lung and colon cancers stand out as the most prevalent causes of death. In medical treatment, knowing the type and location of cancer is crucial. Detecting cancer early can save lives, so having an automated tool is important. However, accurately predicting cancer has been a challenge. This study suggests using artificial neural networks to find abnormal lung tissue growth early. To improve accuracy, a tool with a higher chance of detection is considered. Manual interpretation of results can lead to mistakes. The research analyzed lung images from healthy and cancer patients, creating databases for different CT scan views. A neural network, focusing on image texture, distinguishes normal from malignant images. To tackle this, CNN deep learning algorithm is proposed for Cancer detection. The study includes a quantitative analysis of the proposed network using confusion matrix computation and classification accuracy results.

II. Related Work

Cancer is a major health problem worldwide. Early detection and diagnosis of cancer are essential for successful treatment. Researchers around the world are working on this field to get the best possible result using different methods like machine learning, deep learning etc.

A. Machine Learning

Anji Reddy Vakaa, Badal Sonia, and Sudheer Reddy K. [5] introduced a machine learning-based approach for breast cancer detection. The paper explores innovative techniques to enhance accuracy in identifying breast cancer conditions. By leveraging machine learning algorithms, the authors showcase the potential for improved diagnostic capabilities in breast cancer detection.

In [6] authors Shubham Sharma, Archit Aggarwal, and Tanupriya present a study focused on employing machine learning algorithms for breast cancer detection. The paper explores the application of these algorithms to enhance accuracy in identifying breast cancer. The authors highlight the potential of machine learning techniques in advancing breast cancer diagnosis through their research.

Sri Hari Nallamala, Pragnyaban Mishra, and Suvarna Vani Koneru [7] investigate the application of machine learning techniques for breast cancer detection. The study delves into various approaches within machine learning to enhance the accuracy of breast cancer identification. The authors emphasize the significance of machine learning methods in advancing the capabilities of breast cancer detection through their research.

B. Deep Learning

One of the most challenging as well as demanding task is to segment the region of interest from an object and segmenting the cancer from an lung or colon images is an ambitious one. Tonmoy Hossain, Fairuz Shadmani Shishir [3] present a novel approach to detecting brain tumors. Leveraging Convolutional Neural Networks (CNNs), the study focuses on advanced

image processing techniques. The authors demonstrate the effectiveness of their model in accurate and efficient brain tumor detection. In this work, CNN gained an accuracy of 97.87%, which is very compelling. The main aim of this paper is to distinguish between normal and abnormal pixels, based on texture based and statistical based features.

In this paper [4] Maheep Singh and Krishan Kumar propose an innovative method for cancer detection using histopathological images. Their approach employs Convolutional Neural Networks (CNNs) to enhance accuracy in identifying cancerous conditions. They highlighted the potential of CNNs in advancing cancer detection through advanced image processing techniques. The results show that the proposed method is able to achieve high accuracy in detecting cancer in both breast and skin cancer images.

A. Asuntha and Andy Srinivasan [9] explore the application of deep learning techniques for the detection and classification of lung cancer. The paper delves into the utilization of advanced neural networks to enhance accuracy in identifying and categorizing lung cancer cases. The authors underscore the potential of deep learning methodologies in advancing the field of lung cancer detection through their research. Shivangi Jaina, Vandana Jagtapb, and Nitin Pise [10] present a study that employs image processing techniques for the computer-aided detection of melanoma skin cancer. The paper explores innovative approaches to enhance accuracy in identifying melanoma through advanced image analysis. The authors highlight the significance of image processing in the context of melanoma detection, showcasing its potential impact on diagnostic capabilities.

III. Dataset

In this study, LC25000 dataset is used, introduced by A. Borkowski and team [2] in 2020 . The dataset comprises 25,000 images depicting lung and colon tissues, categorized into five groups. Lung tissue images encompass adenocarcinoma, squamous cell carcinoma, and benign categories, while colon tissue images fall into adenocarcinoma and benign categories.

The LC25000 dataset originally included 1,250 pathology slide images of lung and colon tissues. Through augmentation techniques such as flipping and rotating the images under various conditions, the dataset was expanded to 25,000 images across five categories, with 5,000 images in each category.

Histopathological images from dataset:

Image Type	Class ID	Class Title	Total Image
Lung Adenocarcinoma	0	Lung_aca	5000
Lung squamous cell	1	Lung_scc	5000
Lung benign	2	Lung_n	5000
Colon adenocarcinoma	3	Colon_aca	5000
Colon benign	4	Colon_n	5000

TABLE I
DESCRIPTION OF DATASET

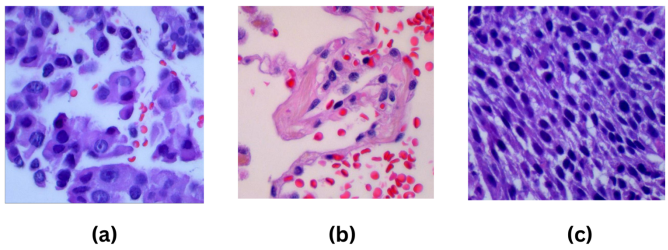

Fig. 1. (a) Lung Adenocarcinoma (b) Lung squamous cell (c) Lung benign

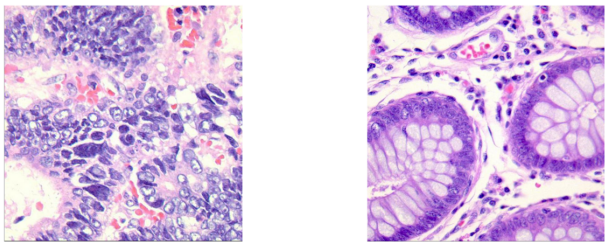

Fig. 2. (a) Colon adenocarcinoma (b) Colon benign

Training options	Parameters
Image Size	224*224
No of Epoch	10

TABLE II
TRAINING PARAMETERS

IV. Methodology

In our prospective model, lung and colon cancer segmentation and detection using machine learning algorithm had been done, and a comparison of the classifiers for our model is delineated. Our proposed lung and colon cancer

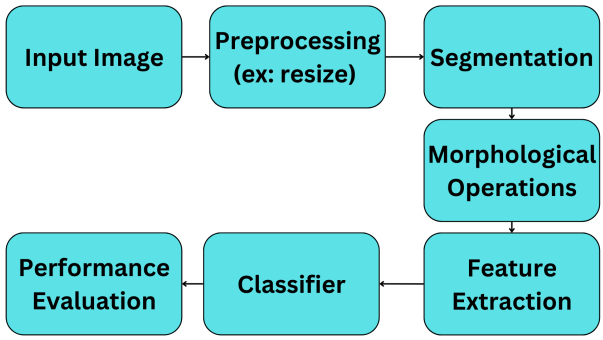

Fig. 3. Fig: Proposed methodology for classification using Traditional Classifiers

image segmentation system consists of several stages: resize, segmentation by efficient algorithm, morphological

operations, , feature extraction and classification by traditional classifiers. The results of our work accomplished satisfactory results. Convolutional Neural Network is broadly used in the field of Medical image processing. Over the years lots of researchers tried to build a model which can detect the tumor more efficiently. We tried to come up with an exemplary which can accurately classify the tumor from lung cancer images. A fully-connected neural network can detect the cancer, but because of parameter sharing and sparsity of connection, we adopted CNN for our model. A Layered Convolutional Neural Network is introduced and implemented for cancer detection. The aggregated model consisting of these stages provides us with the most prominent result for the apprehension of the cancer.

V. Results & Discussions

Evaluating machine learning models is important for classification tasks. There are many different evaluation metrics, but accuracy, precision, recall and F-score are among the most common.

True Positive (TP): The number of positive predictions that are correctly predicted.

True Negative (TN): The number of negative predictions that are correctly predicted.

False Positive (FP): The number of negative predictions that are incorrectly predicted as positive.

False Negative (FN): The number of positive predictions that are incorrectly predicted as negative.

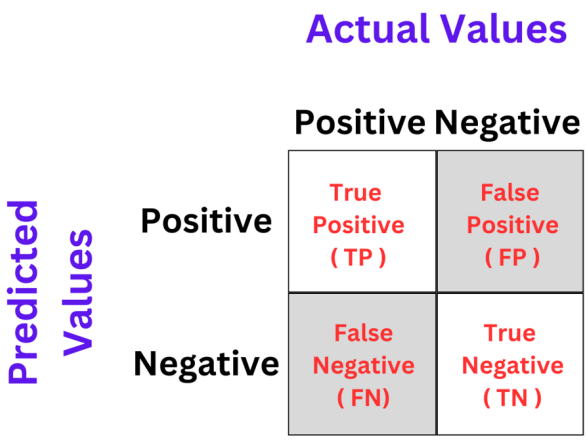

Fig. 4. Confusion Matrix

Precision is the percentage of all positive predictions that are actually positive.

$$Precision = \frac{TP}{TP + FP}$$

Recall is the percentage of true positive predictions that are correctly predicted.

$$Recall = \frac{TP}{TP + FN}$$

F-score is a measure of both precision and recall. It is calculated as the harmonic mean of precision and recall.

$$F\_score = \frac{Precision.Recall}{Precision + Recall}2$$

And finally Accuracy tells you how often the model makes correct predictions overall.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}$$

Data Resizing and Data Augmentation was done on the dataset images to ensure proper training efficiency.

Fig. 5. Experimental Result

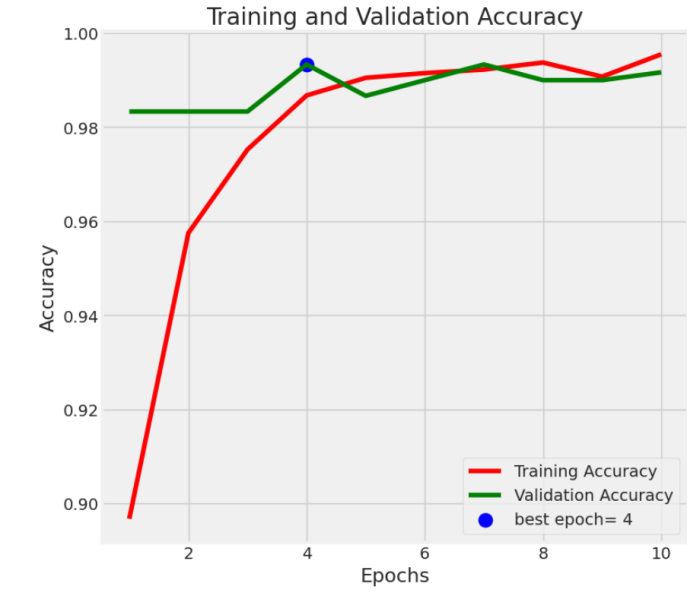

Fig. 6. Experimental Result

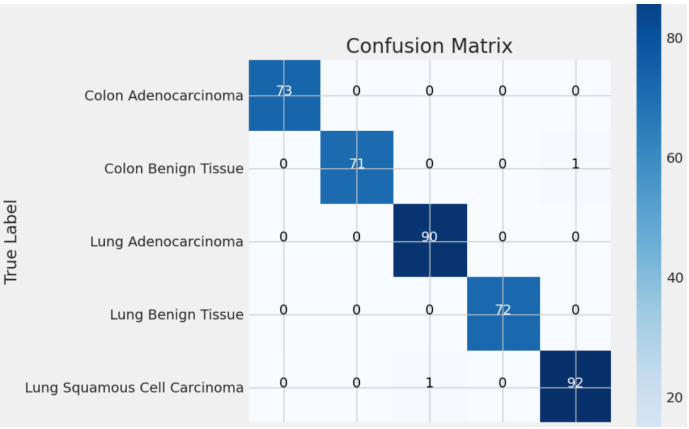

Fig. 7. Experimental Result

VI. Future Work

As there's no precise dataset of Bangladesh cancer patients is available, that can create a huge impact if we can collect that and predict our local issues with more accuracy.

VII. Conclusion

Cancer, a devastating illness characterized by abnormal cell growth, claims millions of lives annually, often due to delayed detection. Accurate identification of cancer cells is crucial, prompting a pressing need for improved diagnostic and treatment approaches. In the realm of medical research, neural networks, including Decision Trees, Artificial Neural Networks, and Support Vector Machines, are actively explored for cancer detection. This paper introduces a classification method that utilizes Convolutional Neural Networks in medical image processing to distinguish cancer cells and proved its efficiency with accuracy.

	precision	recall	f1-score	support
Colon Adenocarcinoma	1.00	1.00	1.00	73
Colon Benign Tissue	1.00	1.00	1.00	72
Lung Adenocarcinoma	1.00	0.98	0.99	90
Lung Benign Tissue	1.00	1.00	1.00	72
Lung Squamous Cell Carcinoma	0.98	1.00	0.99	93
accuracy			0.99	400
macro avg	1.00	1.00	1.00	400
weighted avg	1.00	0.99	0.99	400

Fig. 8. Experimental Result

References

[1] Rebecca L. Siegel MPH, Kimberly D. Miller MPH, "CA: A Cancer Journal for Clinicians," ACL Journals, vol. 73, pp. 17-48, January 2023.

[2] Andrew A. Borkowski and Marilyn M. Bui and L. Brannon Thomas and Catherine P. Wilson and Lauren A. DeLand and Stephen M. Mastorides "Lung and Colon Cancer Histopathological Image Dataset (LC25000).

[3] T. Hossain, F. S. Shishir, M. Ashraf, M. A. Al Nasim and F. Muhammad Shah, "Brain Tumor Detection Using Convolutional Neural Network," 2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT), Dhaka, Bangladesh, 2019, pp. 1-6, doi: 10.1109/ICASERT.2019.8934561.

[4] Dabral, Ishani ,Singh, Maheep, Kumar, Krishan. (2021). Cancer Detection Using Convolutional Neural Network. 10.1007/978-3-030-67187-730.

[5] Reddy, Anji,Soni, Badal,K., Sudheer. (2020). Breast cancer detection by leveraging Machine Learning. ICT Express. 6. 10.1016/j.icte.2020.04.009.

[6] Sharma, Shubham , Aggarwal, Archit, Choudhury, Tanupriya. (2018). Breast Cancer Detection Using Machine Learning Algorithms. 114-118. 10.1109/CTEMS.2018.8769187.

[7] Breast Cancer Detection using Machine Learning Approaches. Sri Hari Nallamala, Pragnyaban Mishra, Suvarna Vani Koneru.International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-7, Issue-5S4, February 2019

[8] Asuntha, A., Srinivasan, Andy. (2020). Deep learning for lung Cancer detection and classification. Multimedia Tools and Applications. 79. 10.1007/s11042-019-08394-3.

[9] Asuntha, A. , Srinivasan, Andy. (2020). Deep learning for lung Cancer detection and classification. Multimedia Tools and Applications. 79. 10.1007/s11042-019-08394-3.

[10] Jain, Shivangi Jagtap, Vandana Pise, Nitin. (2015). Computer Aided Melanoma Skin Cancer Detection Using Image Processing. Procedia Computer Science. 48. 736-741. 10.1016/j.procs.2015.04.209.