Detection of Subtypes of Lung and Colon Cancer Using CNN

P. Narasimha[1], J. Likhith Chowdary[2], J. Vijay Kumar[3], K. Trivenu[4], SK. Mulla Almas Khan[5]
[1,2,3,4] B.Tech. Students, Department of IT, Vasireddy Venkatadri Institute of Technology, Guntur
[5] Assistant Professor, Department of IT, Vasireddy Venkatadri Institute of Technology, Guntur

Abstract- A combination of many metabolic abnormalities and inherited illnesses can lead to the deadly disease known as cancer. Lung and colon cancer are two of the most prevalent causes of death and dysfunction among people in today's world. The Histological Diagnosis of these tumors is usually the most important element in determining the best course of treatment. This research proposes a Deep Learning approach to diagnose Lung Cancer and Colon Cancer from medical pictures using the Convolutional Neural Network (CNN) algorithm. CNN is trained on a large dataset of lung imaging data in order to recognize the features of malignancy. The trained model is evaluated to determine how effectively it can identify cancerous regions using an alternative set of images. The recommended technique successfully identifies lung cancer with high sensitivity, specificity, and accuracy, indicating that radiologists may find it useful for Early Diagnosis and treatment planning. In essence, the suggested CNN algorithm more accurately identifies the subtypes of cancer in the colon and lung. in order to increase the likelihood of an early diagnosis, which can lower the total death rate.

Keywords: CNN, Histological Diagnosis, Lung Cancer, Colon Cancer, Deep Learning, Early Diagnosis

I. INTRODUCTION

The growth and spread of the body's abnormal cells out of control is a characteristic shared by a group of illnesses collectively referred to as cancer. The ability of these cells to spread to other areas of the body through the lymphatic or circulatory systems is known as metastasis.

Lung and colon cancer come in a variety of forms, each with special traits and methods of therapy. Under a microscope, lung cancer may be generally classified into two separate types based on how the cancer cells appear: Non-small cell lung cancer (NSCLC) accounts for the greater percentage of lung cancer cases (85%). Adenocarcinoma, squamous cell carcinoma, and giant cell carcinoma are among the subtypes of NSCLC that can be further subdivided.

Additionally, colon cancer can be categorized according to the kind of cell that gives rise to it: 1. Adenocarcinoma: This is the most common kind of colon cancer, accounting for around 95% of cases. Adenocarcinomas arise from the glandular cells lining the inner surface of the colon.

2. Carcinoid tumors: These are an uncommon kind of colon cancer that start in the colon's hormone-producing cells.

A patient may get chemotherapy, radiation therapy, immunotherapy, targeted therapy, or other treatments for cancer, depending on the kind and stage of the disease. As early identification and treatment can improve the chance of recovery and survival, routine screenings and check-ups are essential to the prevention and management of cancer.

II. LITERATURE SURVEY

Characterizing histopathological imaging data of many cancer kinds, such as skin, breast, lung, colon, and colorectal cancer, has garnered significant attention.

There are several applications for ML, DL, and TL in the detection of lung and colon cancer.

1. Using the ML technique to detect lung and colon cancer: Masood et al. (2021) proposed an ML method based on DL that analyzed pathological pictures of lung and colon cancers to identify five distinct tissue types. Following feature extraction using the 2D Fourier and 2D Wavelet (2D FW) methods, they merged the features and used them to train a CNN model. The findings showed that the recommended design was the most accurate in identifying cancerous tissues, with a 96.33 percent rate.

2. Using CAD to detect lung and colon cancer: (Nishiio et al., 2021) described an automated CAD system for classifying pictures of lung tissue histopathology. They evaluated eight machine learning methods on two datasets, including both traditional texture analysis (TA) and homology-based image processing (HI) to extract visual characteristics. The CAD system with HI outperformed the TA system in both datasets. They came to the conclusion that HI was far more effective for CAD systems than TA and that this may lead to the development of an accurate CAD system for lung tissues. Moreover, (Shandilya and Nayak) created a CAD technique in 2022 for categorizing lung tissue histology pictures. They used a collection of histological pictures of lung tissue that was made accessible to the public for the development and validation of CAD. To extract features from the picture, multiscale processing was applied.

III. METHODOLOGY

1. Data Collection: This stage involves gathering from several sources the pertinent medical pictures of the colon and lung. These pictures may be from a colonoscopy, an MRI, or a CT scan.
2. Data Pre-processing: To eliminate any noise or artifacts, the gathered photos are pre-processed.
After that, the pictures are standardized and scaled to a standard size in order to feed them into the CNN algorithm.
3. CNN Model Training: The pre-processed pictures are used to train a CNN model.
Multiple convolutional layers in the model learn attributes from the photos, and then fully connected layers categorize the images as either malignant or non-cancerous.
4. Model Testing: The Trained CNN model is then examined using another set of pictures that weren't utilized to teach the example. The model's performance is assessed using many measures, such as ROC, specificity, accuracy, and sensitivity bend.
5. Interpretation: After obtaining the model tested and taught, the outcomes are analyzed to determine the effectiveness of the suggested approach. Moreover, Modifications or advancements can be modified the model in light of the outcomes. It's critical to remember that the particulars of each stage may vary. Based on the specific dataset and research objectives. As an example, the CNN model construction and data enhancement techniques, and choosing the hyperparameter can all have a substantial impact on the model's execution. To generate precise and reliable outcomes, it is essential to appropriately schedule and execute each stage of the method.

IV. SYSTEM IMPLEMENTATION

There are many processes involved in implementing a Convolutional Neural Network (CNN) algorithm-based system for the diagnosis of lung and colon cancer. This is a high-level summary of the procedure:

1. Gather a sizable collection of medical photos showing the colon and lungs, including samples that are malignant and those that are not. The pictures are accessible to the general public through databases or medical facilities.
2. Pre-process the photos to make sure they are of the same size and quality and to get rid of any noise or artifacts that might skew the model's results.
3. Create more training pictures by transforming the original photos in different ways, such rotating, resizing, and flipping. This enhances the training data's variety and strengthens the model's capacity for generalization.
4. Make use of a CNN algorithm that is appropriate for identifying lung and colon cancer. To determine the ideal setup, this may entail testing with various layers and hyperparameters.
5. Use the pre-processed and supplemented dataset to train the CNN. To reduce the prediction error, this entails feeding the CNN the pictures and modifying the model's parameters.
6. Test the trained CNN's sensitivity, specificity, and accuracy in identifying malignant areas using an independent set of test pictures.
7. Include the CNN model that has been trained into a software program that can receive fresh medical pictures and produce a forecast of whether the picture shows a colon or lung in good condition or has cancerous regions.
8. Install the system in a clinical environment and track its effectiveness over time to make sure it is accurate and dependable in supporting the identification and management of colon and lung cancer.

V. PREREQUISITES

The following requirements must be met in order to use Saturn Cloud's CNN for the detection of lung and colon cancer:

1. A fundamental knowledge of programming principles, including variables, functions, loops, and conditional expressions, is necessary. It's also advised to have some programming experience using Python.
2. It also requires knowledge of deep learning, neural networks, and supervised learning, among other machine learning principles.
3. You would require access to medical picture datasets of the colon and lungs, which are available from public databases or medical facilities. In order to obtain authorization to download the dataset for execution, we use the Kaggle website in this.
4. In order to use the platform and execute the code, we must first register for a Saturn Cloud account. A cloud-based platform that offers free access to hardware for machine learning algorithms, including GPUs and TPUs.
5. In order to apply the CNN method and carry out data pre-processing and visualization activities, you would need to install and import the necessary libraries and frameworks, such as TensorFlow, Kera's, NumPy, and Pandas.

VI. LIMITATIONS

Large volumes of excellent medical pictures are needed for CNN testing, validation, and training. The process of gathering this kind of data, however, may be difficult as it calls for access to big medical picture datasets, raises ethical questions, and is prone to overfitting—a phenomenon in which a model gets overly tuned to its training set and underperforms when exposed to fresh, untested data. Although overfitting can be lessened by regularization strategies, CNNs are still susceptible to it.

CNNs sometimes provide false negative results, failing to identify cancer when it is present, or false positive results, detecting cancer when none exists. Missed diagnosis and needless biopsies may result from this. The fact that CNNs can only analyze medical pictures may hinder their ability to identify diseases or tumors that do not have visible signs.

VII. FUTURE SCOPE

CNN-based lung and colon cancer detection systems should become more accurate as long as medical imaging databases keep expanding and improvements are made to CNN architectures and training methods. Researchers are looking on ways to make CNN models easier to understand, such employing visualization strategies that clarify the model's predictions. To deliver more precise and thorough diagnoses, CNN-based lung and colon cancer detection systems might be combined with additional diagnostic instruments like blood tests or genetic screenings.

Early cancer identification might greatly enhance patient outcomes and boost treatment success rates. CNN-based systems for lung and colon cancer detection could be employed for this purpose.

CNN-based lung and colon cancer detection systems might be utilized to create individualized treatment regimens that are catered to each patient and their unique circumstances by evaluating medical imaging and other patient data.

Nevertheless, there is hope for the use of CNN algorithms in lung and colon cancer diagnosis as these technologies advance and become more widely used in clinical settings. They have the potential to significantly improve cancer diagnosis and treatment outcomes and might be essential in the battle against the disease.

VIII. RESULT

i. CNN Model Summary:

```
Model: "sequential"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 conv2d (Conv2D)             (None, 70, 70, 32)        896

 activation (Activation)     (None, 70, 70, 32)        0

 max_pooling2d (MaxPooling2  (None, 35, 35, 32)        0
 D)

 conv2d_1 (Conv2D)           (None, 35, 35, 64)        18496

 activation_1 (Activation)   (None, 35, 35, 64)        0

 max_pooling2d_1 (MaxPoolin  (None, 17, 17, 64)        0
 g2D)

 conv2d_2 (Conv2D)           (None, 17, 17, 128)       73856

 activation_2 (Activation)   (None, 17, 17, 128)       0

 max_pooling2d_2 (MaxPoolin  (None, 8, 8, 128)         0
 g2D)

 conv2d_3 (Conv2D)           (None, 8, 8, 256)         295168

 activation_3 (Activation)   (None, 8, 8, 256)         0

 max_pooling2d_3 (MaxPoolin  (None, 4, 4, 256)         0
 g2D)

 conv2d_4 (Conv2D)           (None, 4, 4, 512)         1180160

 activation_4 (Activation)   (None, 4, 4, 512)         0

 max_pooling2d_4 (MaxPoolin  (None, 2, 2, 512)         0
 g2D)

 conv2d_5 (Conv2D)           (None, 2, 2, 1024)        4719616

 activation_5 (Activation)   (None, 2, 2, 1024)        0

 max_pooling2d_5 (MaxPoolin  (None, 1, 1, 1024)        0
 g2D)

 flatten (Flatten)           (None, 1024)              0

 dense (Dense)               (None, 512)               524800

 dense_1 (Dense)             (None, 5)                 2565

=================================================================
Total params: 6815557 (26.00 MB)
Trainable params: 6815557 (26.00 MB)
Non-trainable params: 0 (0.00 Byte)
```

ii. Confusion Matrix for Train Data:

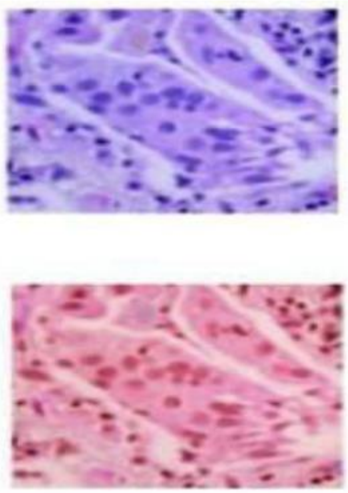
```
[[3585,    0,  197,    0,    0],
 [   0, 3711,   32,    0,    0],
 [ 238,   27, 3425,    3,   32],
 [   0,    0,    0, 3756,   14],
 [   0,    3,    1,  235, 3491]]
```

iii. Confusion Matrix for Test data:

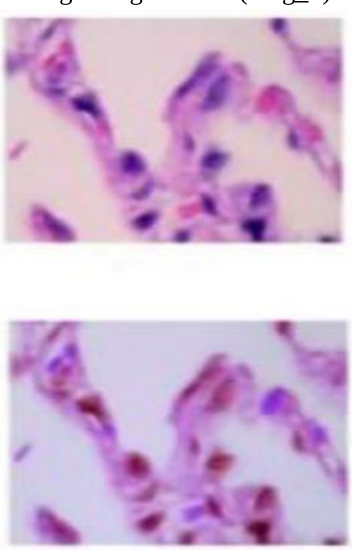
```
[[1153,    0,   65,    0,    0],
 [   0, 1239,   17,    1,    0],
 [  79,   13, 1169,    1,   13],
 [   0,    1,    0, 1215,   14],
 [   0,    2,    0,  100, 1168]]
```

iv. Final Outputs:

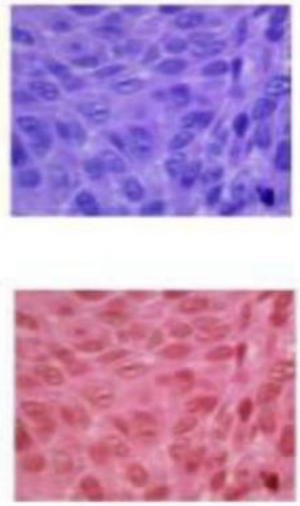
a) Lung Adenocarcinoma (lung_aca)



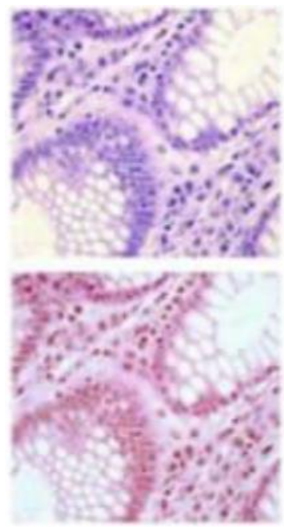
b) Lung Benign Tissue (lung_n)



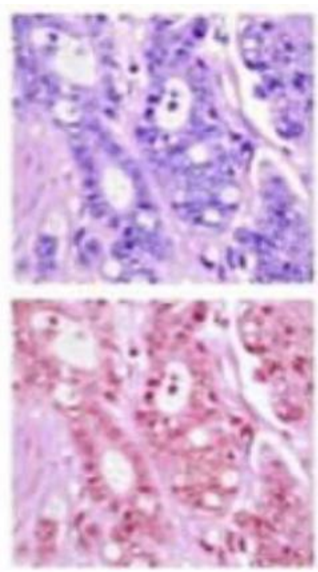
c) Lung Squamous Cell Carcinoma (lung_scc)



d) Colon Adenocarcinoma (colon_aca)



e) Colon Benign Tissue (colon_n)



IX. CONCLUSION

To sum up, CNN-based algorithms for the identification of lung and colon cancer have shown promising results in accurately identifying cancerous regions in medical images. These technologies have the potential to significantly improve cancer diagnosis and treatment results because they can detect cancer early, reduce the need for unnecessary biopsies, and raise the accuracy of diagnoses. However, there are a number of limitations that limit the effectiveness of these systems, including the likelihood of false positives and false negatives, interpretability concerns, and dependence on high-quality data. CNN-based lung and colon cancer detection systems should be used in conjunction with other diagnostic tools and by trained medical professionals to provide a precise and reliable cancer diagnosis.

Future advancements and integration of CNN-based lung and colon cancer detection systems into clinical practice are highly promising. As medical image collections grow, we should expect these systems to become more accurate. Additionally, advancements in CNN designs and training techniques should make it possible for these systems to be integrated with other diagnostic tools for personalized treatment.

REFERENCES

[1] “Lung cancer screening: the guidelines" by Moyer VA, U.S. Preventive Services Task Force (2014): Evidence-based recommendations for lung cancer screening are given in this article, along with advice on when to utilize low-dose computed tomography (LDCT) on people who are at high risk.

[2] "The American Cancer Society's Cancer Prevention Guidelines" by Kushi et al. (2012): The prevention of colon cancer is outlined in this article, along with suggestions for screening exams such colonoscopies, fecal occult blood tests, and stool DNA testing.

[3] "Screening for colorectal cancer: U.S. Preventive Services Task Force recommendation statement" by Bibbins-Domingo et al. (2016): The evidence-based recommendations for colon cancer screening in this article cover the age at which screening should begin, the number of screenings that should be performed, and recommended screening tests.

[4] "Lung cancer diagnosis and staging" by Vansteenkiste et al. (2013): An overview of lung cancer diagnosis and staging, including the use of imaging, biopsies, and molecular testing, is given in this article.

[5] "Colon cancer: diagnosis and treatment" by Chen et al. (2017): An overview of colon cancer diagnosis and treatment, including the use of imaging, colonoscopy, and surgery, is given in this article.

[6] "Lung cancer screening with low dose computed tomography" by Aberle et al. (2011): Evidence-based guidelines for patient selection, screening intervals, and result interpretation are included in this article about the use of LDCT in lung cancer screening.

[7] "The role of colonoscopy in colorectal cancer screening" by Lieberman et al. (2009): An summary of colonoscopy's function in colon cancer screening is given in this article, along with suggestions for patient preparation, selection, and follow-up.