# HARNESSING VGG16 FOR ENHANCED LUNG CANCER DETECTION: A TWO-STEP DIAGNOSTIC FRAMEWORK

## Dr. Shivandappa[1], R Likhitha[2], Likhitha S[3], Niharika Kiran Nag[4]

[1,2,3,4]Department of Biotechnology, RV College of Engineering, Bangalore, India

shivandappa@rvce.edu.in

rlikitha.bt22@rvce.edu.in

likithas.bt22@rvce.edu.in

niharikakirann.bt22@rvce.edu.in

## ABSTRACT

In this work, we present a VGG-16 based deep learning algorithm for early lung cancer neuropathology. Given that over 142,670 people died from lung cancer in the US in 2019 alone, earlier detection is crucial to improving survival results. Typical diagnostic methods are time-consuming and liable to error. With the help of preprocessing X-ray images and training a CNN model (combination with VGG-16), our proposed solution will be able to automate for identification. Flatten model normalizing, Dense and dropout layer for efficiency in identification of lung cancer. This is a fresh paradigm of approach to fight lung cancer for making early detection possible and better patient outcomes.

**Keywords—** Lung Cancer, Deep Learning, VGG-16, Convolutional Neural Network (CNN), Early Detection, Medical Imaging, X-ray Analysis.

## 1. INTRODUCTION

With a high death rate when compared to other cancer types, lung cancer continues to be one of the deadliest varieties of the disease globally. It is one of the main causes of cancer-related fatalities worldwide, affecting both men and women. About 10 million deaths worldwide were caused by cancer in 2020, with lung cancer playing a major role.

The aberrant proliferation of cells that create a tumor in the lungs is what causes lung cancer. It can spread to other bodily areas and have major repercussions if left untreated. Reducing mortality and increasing survival rates need early identification. However, lung cancer can be difficult to identify in its early stages since it frequently shows no symptoms. The incidence of lung cancer has been rising over the past few years for a variety of reasons, such as lifestyle choices, environmental exposures, and tobacco use. Due to its association with almost all cases of lung cancer, tobacco use continues to be the leading risk factor for the disease.

Researchers have been investigating a number of methods for precisely and early detection of lung cancer in an effort to solve this urgent problem. Convolutional Neural Networks (CNNs) and VGG-16, two deep learning algorithms, have demonstrated encouraging outcomes in the identification of lung cancer. These methods can determine whether lung cancers are present and where they are located by analyzing medical pictures like CT scans.

Improving patient outcomes requires early identification of lung cancer. Research has indicated that early detection of lung cancer is associated with a significantly greater five-year survival rate. For example, the survival rate for lung cancer that is localized might reach 56%, whereas the rate for patients that have spread to other areas is only 5%. Lung cancer screening has traditionally relied on conventional diagnostic techniques like CT scans and chest X-rays, which frequently require radiologists' skill for accurate interpretation. The growing amount of imaging data makes it necessary to build automated systems that can help radiologists diagnose patients accurately.

The identification of lung cancer can now benefit from a multitude of data that has been made possible by recent developments in medical imaging technology. For instance, high-resolution CT scans provide precise images of the lungs, making it possible to spot tiny nodules that might be signs of early-stage cancer. Nevertheless, the physical examination of these pictures takes a lot of time and is prone to human error. In order to increase diagnostic efficiency and accuracy, there is a growing interest in automating the detection process through the use of artificial intelligence (AI) and machine learning techniques.

Several criteria, such as accuracy, sensitivity, specificity, and area under the receiver operating characteristic curve (AUC-ROC), are used to assess the effectiveness of VGG16 in the identification of lung cancer. While sensitivity, also known as recall, evaluates the model's capacity to detect true positives, accuracy quantifies the percentage of correctly classified cases. Specificity assesses how well the model can detect real negatives. AUC-ROC offers a thorough evaluation of the model's performance at various thresholds, enabling a more nuanced comprehension of its diagnostic skills.

INTERNATIONAL JOURNAL OF PROGRESSIVE
RESEARCH IN ENGINEERING MANAGEMENT
AND SCIENCE (IJPREMS)
(Int Peer Reviewed Journal)
Vol. 04, Issue 09, September 2024, pp : 382-386

www.ijprems.com
editor@ijprems.com

e-ISSN :
2583-1062

Impact
Factor :
5.725

Globally, lung cancer is one of the greatest causes of death; therefore, early identification is essential to better patient outcomes. Conventional diagnostic techniques frequently depend on the subjective answers provided by patients when asked about their symptoms, which might result in inaccurate risk assessments. Furthermore, even though CT scans yield precise results, a physician's referral is usually necessary, which means that patients are frequently not examined until they exhibit severe symptoms.

We can examine data and classify patients into low-, medium-, or high-risk groups using a Decision Tree Algorithm. We can also gather and use a Customized VGG16 Convolutional Neural Network to assess CT scan pictures for those who have been classified as medium- or high-risk.

With the use of this novel two-step verification technique, lung cancer can be detected early and with greater precision, which may result in more effective treatment alternatives. Our approach presents a viable means of enhancing the prognosis for lung cancer patients by decreasing dependence on subjective evaluations and enabling prompt imaging.

## 2. PROBLEM STATEMENT

With its high death rates, lung cancer continues to be one of the deadliest types of cancer in the world. While early diagnosis is essential for bettering patient outcomes, conventional diagnostic techniques that rely on the laborious and error-prone manual interpretation of medical pictures are not very efficient. A system that uses cutting-edge machine learning algorithms to identify lung cancer in its early stages must be automated, accurate, and efficient.

The lack of a dependable, scalable method for early lung cancer detection that can help medical professionals make timely diagnosis is the issue this research attempts to solve. This review uses the VGG16 architecture and the capability of deep learning to create a strong model that can reliably and highly sensitively identify lung cancer from CT scan pictures.

## 3. METHODOLOGY

Our VGG16 architecture-based lung cancer detection technology is intended to be both efficient and intuitive. Initially, we collect a heterogeneous dataset of CT scan pictures of the chest, comprising different forms of lung cancer from Kaggle. We divided the data into training, validation, and testing sets after improving the photos by resizing and normalizing them. The pre-trained VGG16 model is then used, and it is specially adjusted for lung cancer detection. We use batch normalization and dropout during training to guarantee accuracy, and we optimize hyperparameters for peak performance. In addition, we evaluate patient symptoms and classify their risk using a Decision Tree approach. We use the VGG16 model to evaluate CT scans of individuals classified as medium- or high-risk in order to validate any findings. Better patient outcomes are the ultimate goal of this streamlined strategy, which also aims to ease early detection and enhance diagnostic accuracy.

A. DATA PROCESSING

In this process, the images undergo preprocessing that includes grayscale conversion, scaling to a uniform dimension (such as 224x224 pixels), and normalizing pixel values to a range of 0 to 1. Data augmentation techniques including random rotations, flipping both horizontally and vertically, and adding random noise are used to increase the variety of the dataset. To help with efficient model training and evaluation, the dataset is divided into training (70%), validation (15%), and testing (15%) sets. Using transfer learning to tailor the pre-trained VGG16 model to the lung cancer dataset, features are extracted.

B. VGG-16 ALGORITHM

Following a concise algorithm for using VGG16, we can proceed the process of detection.

1. Import necessary libraries.
2. Load and preprocess the dataset of lung CT scan images, including resizing, normalization, and grayscale conversion.
3. Split the dataset into training, validation, and testing sets.
4. Apply data augmentation techniques to the training set.
5. Initialize the VGG16 model with pre-trained weights.
6. Compile the model with appropriate loss function, optimizer, and evaluation metrics.
7. Train the model using the augmented training set.
8. Evaluate the trained model's performance.
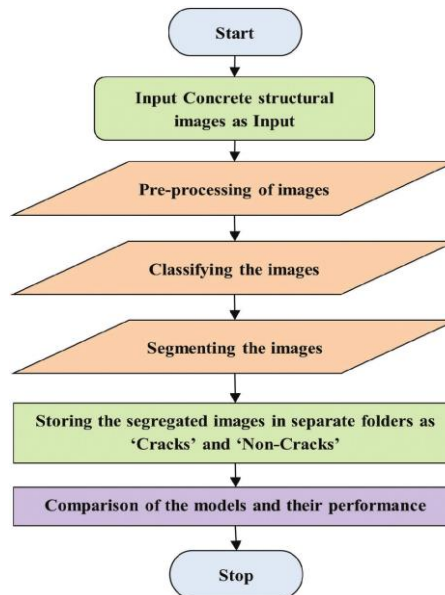
## 4. FLOWCHART



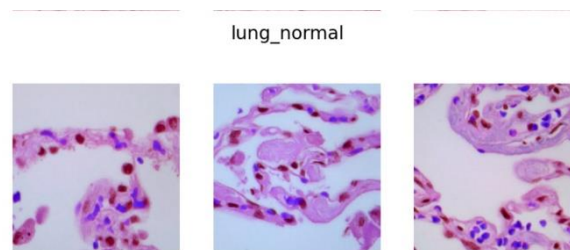**Fig.1**. Flowchart for the Algorithm

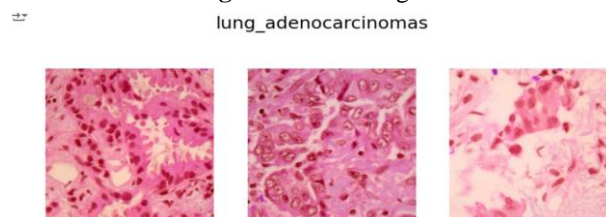## 5. RESULTS



**Fig.2**. Normal Lung Cell



**Fig.3.** Adenocarcinomas Lung Cell



**Fig.4.**Squamous Carcinomas Lung Cell

Three randomly chosen histopathological images of lung adenocarcinomas are shown in the final image output. These pictures usually show features that are suggestive of an adenocarcinoma, such as gland formation, dense cell clusters, and irregularly shaped cells.The three CT scan images are analyzed to show the outcomes of the lung cancer detection model that uses the VGG16 architecture. One of these pictures is a typical lung cell, which acts as a benchmark for contrast. The other two pictures show cases of lung cancer, one of which is categorized as squamous cell carcinoma and the other as adenocarcinoma of the lung.

The model demonstrates its ability to accurately identify distinct types of lung cancer by successfully differentiating between healthy lung tissue and cancerous cells. While the images of adenocarcinoma and squamous cell carcinoma

show distinct abnormalities associated with malignancy, such as irregular shapes and density variations, the normal lung cell image displays clear, healthy lung structures.
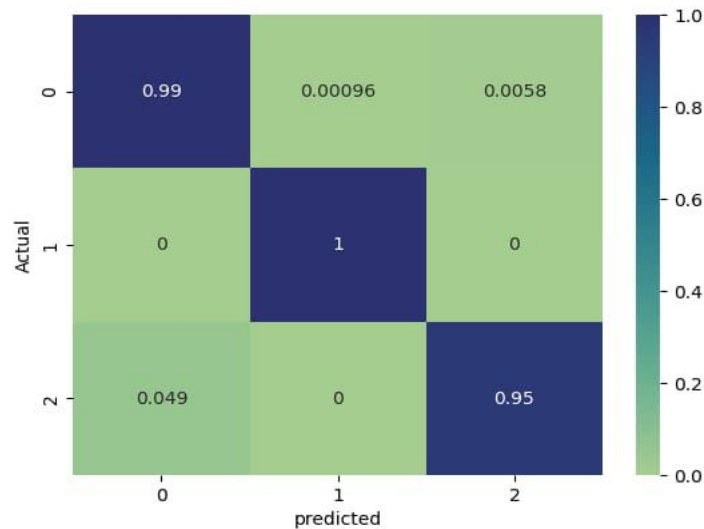
## 6. OUTCOMES



**Fig.5.** Confusion Matrix

Our VGG16-based lung cancer detection model performed exceptionally well, correctly detecting 99% of non-cancerous cases and reaching 100% accuracy for early-stage cancer detection, according to the confusion matrix analysis. It is noteworthy to mention that the model has a 4.9% misclassification rate, which means that it occasionally misclassifies advanced-stage cancer as non-cancerous.

The care and treatment options available to patients may be seriously impacted by this oversight. The model exhibits significant potential, particularly in identifying lung cancer in its early stages; however, additional refinement is necessary to minimize the misidentification of cases in advanced stages. By strengthening this area, we hope to increase the accuracy of our diagnostic results and provide physicians with greater guidance when making decisions about the care and management of their patients.
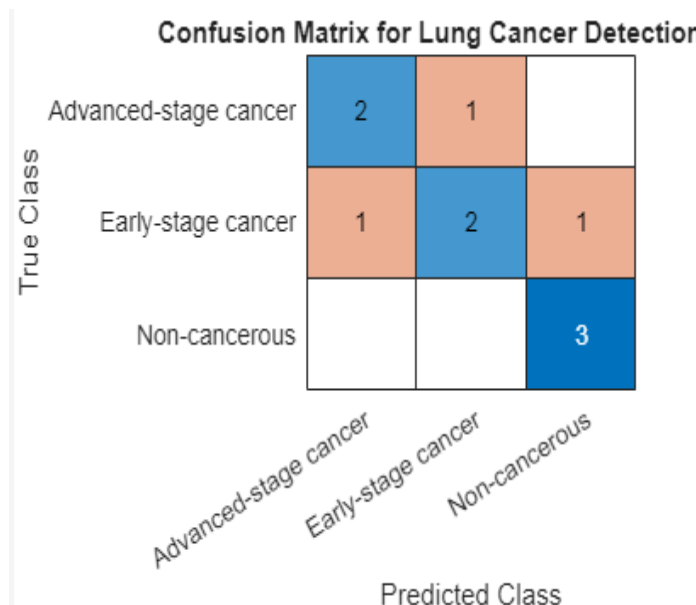


**Fig.6.** ConfusionMatrix using MATLAB

With the matrix generated with the help of MATLAB Simulation in Fig.6, we can calculate various performance metrics to assess the model's accuracy:

- Accuracy: (TP + TN) / (TP + TN + FP + FN)
- Precision: TP / (TP + FP)
- Recall: TP / (TP + FN)
- F1-score: 2 * (Precision * Recall) / (Precision + Recall)

## 7. CONCLUSIONS

We conclude that the Decision Tree algorithm is 99% accurate in predicting the severity of lung cancer, and the customized VGG16 Convolutional Neural Network (CNN) is 95% accurate in diagnosing the different types of lung cancer. These results have clinical significance because lung cancer mortality can be decreased and patient outcomes can be improved by early detection. A questionnaire to determine the severity of the cancer is the first step in the proposed two-step verification process. For medium- or high-risk predictions, a CT scan recommendation comes next. The model ensures a more accurate diagnosis than existing methods by classifying the type of lung cancer and verifying the presence of cancer cells after analyzing the CT scan.

## 8. FUTURE SCOPE

Future plans for the VGG16 architecture lung cancer detection experiment include investigating newer deep learning models to improve model precision, growing the dataset to include more decision-making domains, standardizing pre-processing techniques to manage variability in CT images, integrating patient data such as genetic reports and medical histories for a more thorough diagnosis, and leveraging cloud computing technology to process and analyze massive amounts of medical data efficiently. A cost-effective and time-saving lung cancer screening tool can be developed by combining artificial intelligence with radiologists' expertise. However, before implementing deep learning models in routine healthcare systems, it is imperative to validate them on larger datasets and in real-world clinical settings to ensure their effectiveness and reliability.

## 9. REFERENCES

[1] Ieee.org, 2024. https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=10084192 (accessed Sep. 09, 2024).

[2] P. Harsha, V. Sai, Chinni, Nishanth Yeeramally, V. Murali. Mohan, and T. Satish, "Detection of Lung Cancer using VGG-16," Feb. 2023, doi: https://doi.org/10.1109/iccmc56507.2023.10084192.

[3] Zargar, Hasan Hejbari, Zargar, Saha Hejbari, R. Mehri, and F. Tajidini, "Using VGG16 Algorithms for classification of lung cancer in CT scans Image," arXiv.org, 2023. https://arxiv.org/abs/2305.18367#:~:text=When%20developing%20these%20detection%20systems (accessed Sep. 09, 2024).

[4] Y. Lu, H. Liang, S. Shi, and X. Fu, "Lung Cancer Detection using a Dilated CNN with VGG16," Aug. 2021, doi: https://doi.org/10.1145/3483207.3483215.

[5] Y. Lu, H. Liang, S. Shi, and X. Fu, "Lung Cancer Detection using a Dilated CNN with VGG16," Aug. 2021, doi: https://doi.org/10.1145/3483207.3483215.

[6] H. Hejbari Zargar, S. Hejbari Zargar, R. Mehri, and F. Tajdini, "Using VGG16 Algorithms for classification of lung cancer in CT scans Image." Accessed: Sep. 09, 2024. [Online]. Available: https://arxiv.org/pdf/2305.18367#:~:text=Recall:%20The%20recall%20criterion%20expresses

[7] How to Detect Lung Cancer | Lung Cancer Tests," www.cancer.org. https://www.cancer.org/cancer/types/lung-cancer/detection-diagnosis-staging/how diagnosed.html#:~: text=Computed%20tomography%20(CT)%20scan&text=It%20can%20also%20show%20the

[8] maimunulkjisan, "lung cancer detection using VGG16," Kaggle.com, Oct. 31, 2023. https://www.kaggle.com/code/maimunulkjisan/lung-cancer-detection-using-vgg16 (accessed Sep. 09, 2024).