IJRAR.ORG E-ISSN: 2348-1269, P-ISSN: 2349-5138

INTERNATIONAL JOURNAL OF RESEARCH AND ANALYTICAL REVIEWS (IJRAR) | IJRAR.ORG

An International Open Access, Peer-reviewed, Refereed Journal

# Deep Learning-Based Computer Aided Diagnosis for Lung Cancer Detection

[1] Mr. J.Nagaraju, [2] Ramireddy Gopi Krishna, [3] Siddepalli Lasya Priya,[4]Sankurathri Gopi Raju

[1]Assistant Professor, [2]Student, [3]Student, [4]Student

[1]Department of Artificial Intelligence and Data Science,

[1]Lakireddy Bali Reddy College of Engineering, Mylavaram, India

***Abstract:*** The evolution Utilizing low-dose helical computed tomography (CT) has sparked renewed optimism for effective lung cancer screening. Justifying screening requires evidence of prolonged life expectancy at reasonable cost and manageable risk levels. An efficient screening test should accurately detect all existing cancers while minimizing unnecessary interventions. Therefore, ideal screening methods prioritize both high sensitivity and specificity. Presently, due to technological limitations, radiologists often prioritize heightened sensitivity, necessitating additional diagnostic procedures to address false positives. Published reports suggest that computer-aided diagnostic technology might potentially influence the cost-benefit analysis of CT's sensitivity and specificity in lung cancer screening protocols. This review aims to explore the ongoing discourse surrounding the effectiveness of lung cancer screening while evaluating the potential of state-of-the-art developments in Automated diagnostic design.

***IndexTerms*** - **Automated diagnostic systems, Deep Learning, Convolutional Neural Network , Artificial Nueral Networks**

## 1. Introduction

In the past decade, there has been a troubling increase in the mortality and morbidity rates associated with malignant tumors, presenting a significant challenge in both preventing and treating cancer. Among the top ten cancers, lung cancer stands out with the highest incidence and ranks first in cancer-related deaths. Notably, early symptoms of lung cancer often go unnoticed, with patients experiencing no apparent bodily abnormalities. Diagnosis typically occurs only when significant abnormalities become evident, signaling the presence of lung cancer. Consequently, diagnoses usually happen at intermediate to advanced stages. Unfortunately, current clinical surgical interventions often prove ineffective for advanced-stage patients, leading many to opt out of treatment, impacting their chances of survival.[3]

Studying the diagnostic system for lung cancer using text and images offers valuable insights for clinicians. This research has the potential to significantly lessen clinicians' workload by automating the manual screening process for early-stage lung cancer patients. Moreover, it serves as a safeguard against missed screenings caused by fatigue and other human factors.


**fig:1.1**- Advantages of CAD system

Over the past years, CAD systems have emerged as a viable solution to address the constraints of conventional lung cancer diagnosis methods. These systems employ computer algorithms and artificial intelligence (AI) techniques to aid radiologists in detecting and diagnosing lung cancer. By analyzing medical images, they offer automated and precise detection of suspicious lesions, significantly enhancing the efficiency and accuracy of lung cancer diagnosis.

## 2. LITERATURE SURVEY

The literature review on the diagnosis of Lung Cancer covers a range of deep learning techniques, each offering distinct insights. Lee et al. introduced an innovative template-matching technique utilizing different techniques for nodule detection. Their method underwent evaluation on 557 CT scans and achieving an accuracy of 0.72.[6]

Li et al. introduced an approach designed to aid radiologists in improving the identification of tumor nodules in CT scans, particularly emphasizing false-positive reduction. Their approach resulted in vast decrease of false positive rate by 44 percent and has increasd the true positive rate to 0.23.[7]

Katsuragawa et al. outlined an automated approach for dis-tinguishing different types of classes in lung cancer. They have used Artificial Neural Networks(ANN) for discriminating among fifty-five chest radiographs. By using ANN they have acheived a accuracy rate of 88.6 percent.[1]

Gurcan et al. has introduced a technique aimed at detecting lung nodules in CT scans. Initially, k-means clustering identified the lung regions in the first stage. Subsequently, they ap-plied rule-based classification and further refined the process using LDA to minimize false positives. Their study has analyzed and has acheived an accuracy rate of 88.4 percent and false positive rate of 5.48 per each scan. [4]

Arimura et al. has introduced a method for identifying lung tumor in different lung scan images. They have developed a model by using Artificial neural networks(ANN) to reduce the false positive rates. They have verified various images from different hospitals of various patients. They have acheived an accuracy of over 0.83.[2]

Suzuki et al. developed a technique by using Multiple Arti-ficial Neural Networks(ANN) to reduce the rate of false pos-itives. They have successfully eliminated the false positives of percent 66.8 and also has increased true positive rates.They have developed a model and that model aims to get an accuracy of 8l.4 percent.[8]

Ko et al. has introduced a method to identify lung tumors in patients.He has developed a model using CNN.They mainly focuses on to identify the classes accurately.They have got an accuracy rate of 0.92.[5]

## 3. METHODOLOGY

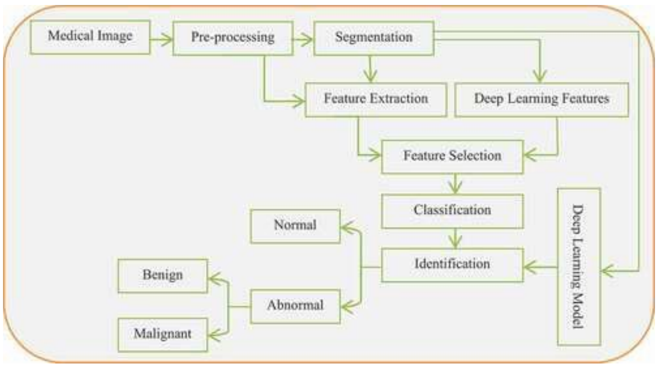


**fig:2.1**-Workflow of System

The Workflow of the system to detect lung cancer classes typically involves four steps: Data collection, noise removal, segmentation of images and identifying different cases. In the image acquisition step, medical images are obtained using imaging techniques such as CT scans or MRI. In the pre-processing step, the images are enhanced and filtered to improve the quality and reduce noise. Feature extraction involves extracting relevant features from the images, such as shape, texture, and intensity, which are then used to train the classification algorithm. The final step is classification, where the CAD system uses a trained algorithm to analyze the extracted features and classify the image as either normal or abnormal.

### 3.1 Image Acquisition:

In the realm of image processing, acquisition refers to the process of retrieving an image, typically from a hardware-based source, to prepare it for subsequent processing. This initial step stands as the starting point in the workflow sequence, as any further processing relies fundamentally on the availability of an image.

### 3.2 Pre-Processing:

In image analysis, pre-processing plays a crucial role in refining raw images to optimize their quality for later analysis or applications. Techniques such as noise reduction, contrast enhancement, normalization, and resizing are employed to address imperfections or distortions in the data, enhancing clarity and ensuring consistency for more accurate analysis.

### 3.3 Feature Extraction:

Feature extraction is integral to the dimensionality reduction process, where the initial raw data set is divided and condensed into more manageable groups, facilitating easier processing. Its role lies in identifying and consolidating variables into features, aiding in the selection of optimal attributes from large data sets. This process effectively reduces the data volume, making it more manageable and enhancing the efficiency of subsequent processing tasks.

### 3.4 Classification:

Image classification categorizes pixels or segments in an image using predefined characteristics. This technique uses algorithms that analyze features like color, texture, or shape to assign labels to different parts of the image, distinguishing between various objects or classes. Its goal is to accurately identify and categorize objects or patterns, enabling automated understanding of visual data.

### 3.5 System Architecture:

The System Architecture mainly consists of Three phases i.e Data collection and noise removal, building model and identifying different cases. The process execution occurs in a flow manner in these three steps.

The first step involves data collection from different sources and different hospitals and also data pre-processing like noise removal, removing unwanted data and lung segmentation. Second step involves Training of data ,here a model is developed using CNN and it consists of three layers i.e Convolutional, Pooling and FC layer. Third step involves Training phases in this phase the model classifies the images into different classes and also identifies the normal case or lung cancer cases.

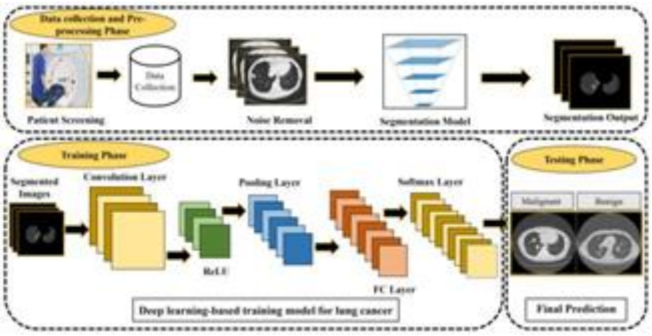


**fig:3.1**-Architecture of System

### 3.6 Dataset Description:

The dataset of lung cancer has gathered from Iraq-Oncology Teaching Hospital(IQ-OTH) from various specialized hospitals over a span of three months period in 2019.The dataset consists of various images of lung scan images from patients who have lung cancer and also from healthy persons who are not diagnosed with lung cancer. This dataset consists of 1190 images which are grouped into three different classes of two classes i.e benign, malignant are lung cancer classes and normal classes from healthy persons.

In each scan, there are multiple slices, typically ranging from 80 to 200 slices. These slices capture various perspectives of the human chest. Within the 110 cases, there’s diversity in gen-der, age, education, residential area, and living conditions. The individuals belong to different professions, such as employees in Iraqi ministries of Transport and Oil, as well as farmers. They primarily hail from regions in central Iraq, particularly from provinces like Baghdad, Wasit, Diyala, Salahuddin, and Babylon.

## 4 Results:

From the past years the CAD system has improved a lot to detect the early stage cancers accurately. These systems, often leveraging advanced imaging technologies like CT scans, have demonstrated promising outcomes, particularly in identifying early-stage lung cancer. The synergy of CAD with AI has notably improved feature extraction and classification accuracy from imaging data.

### 4.1 Model Accuracy:

The below image shows a graph of the model accuracy of a train. It is a line graph with two lines, one for train accuracy and one for validation accuracy. The x-axis is labelled ”Epoch” and the y-axis is labelled ”Accuracy.” The graph shows that the training accuracy starts at around 70 per cent and gradually increases over time, reaching a plateau of around 95 per cent after 14 epochs. The validation accuracy is slightly lower than the training accuracy, but it also follows a similar trend.

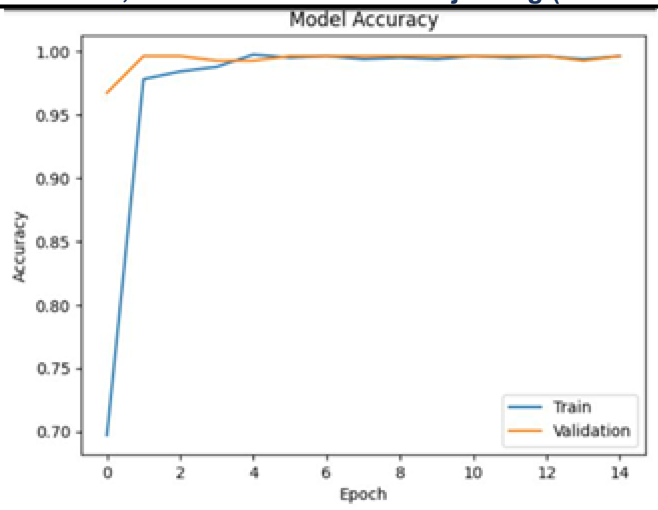


**fig:4.1.1**-Accuracy of the system

Overall, the graph shows that the model is learning well and can achieve high accuracy on data. This suggests that the model is likely to be able to generalize well to new data and could be used to develop a reliable computer-aided diagnosis system for lung cancer detection.

### 4.2 Model Loss:

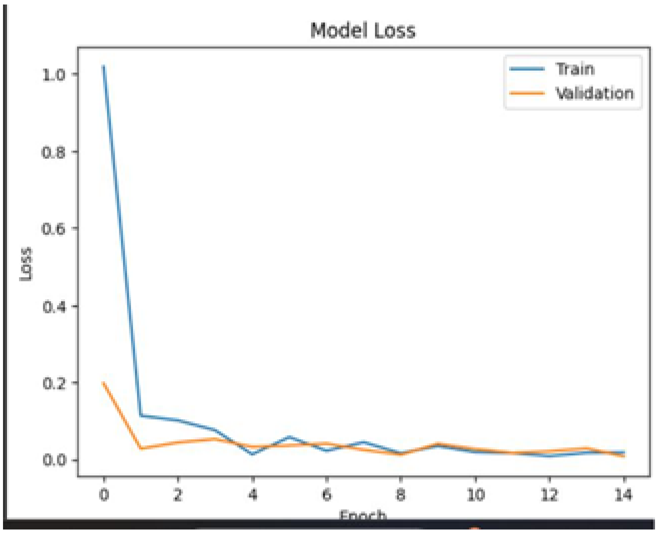


**fig:4.2.1**-Loss percent of the system

The graph shows that the training loss starts at around 1.0 and gradually decreases over time, reaching a plateau of around 0.2 after 14 epochs. The validation loss is slightly higher than the train loss, but it also follows a similar trend. Overall, the graph shows that the model is learning well and can achieve low loss. This suggests that the model is likely to be able to generalize well to new data.

## 5 CONCLUSIONS AND FUTURE ENHANCEMENT

In summary, leveraging Convolutional Neural Networks (CNNs) in for lung tumour identification represents a significant advancement in early detection methods. The integration of CNNs has shown promising outcomes, accurately categorizing and detecting lung tumors, thus contributing to the development of precise diagnostic tools. Incorporating this technology not only improves diagnostic precision but also holds potential for timely interventions and tailored treatment plans. This innovation stands to enhance patient outcomes and revolutionize healthcare strategies aimed at combatting lung cancer.

The performance metrics, including accuracy, sensitivity, and specificity, were assessed and compared with existing techniques. The outcomes demonstrated an impressive 0.97 accuracy in stage classification for the segmented images under test. These sophisticated deep learning models show promising capability in early tumor detection, facilitating effective virtual monitoring and diagnosis. Future research will focus on expanding and analyzing this system on larger databases, aiming to assess its ability to securely manage extensive hospital data. Furthermore, plans are in place to advance this prototype into a comprehensive cloud-based hardware module in upcoming phases of development.

### 5.1 FUTURE DIRECTIONS:

The future scope of this type of systems for lung cancer detection is vast. With the advancements in technology, CAD systems can be integrated with other diagnostic tools, such as biomarkers and genetic testing, to provide a more comprehensive diagnosis. These systems can also be trained on a larger dataset to improve their performance and accuracy.

## 6 Refernces

[1] Masahito Aoyama, Qiang Li, Shigehiko Katsuragawa, Heber MacMahon, and Kunio Doi. Automated computerized scheme for distinction between benign and malignant solitary pulmonary nodules on chest images. Medi-cal Physics, 29(5):701–708, 2002.

[2] Hidetaka Arimura, Shigehiko Katsuragawa, Kenji Suzuki, Feng Li, Junji Shiraishi, Shusuke Sone, and Kunio Doi. Computerized scheme for auto-mated detection of lung nodules in low-dose computed tomography images for lung cancer screening1. Academic Radiology, 11(6):617–629, 2004.

[3] Su Chen et al. Models of artificial intelligence-assisted diagnosis of lung cancer pathology based on deep learning algorithms. Journal of Healthcare Engineering, 2022, 2022.

[4] Metin N Gurcan, Berkman Sahiner, Nicholas Petrick, Heang-Ping Chan, Ella A Kazerooni, Philip N Cascade, and Lubomir Hadjiiski. Lung nodule detection on thoracic computed tomography images: Preliminary evalua-tion of a computer-aided diagnosis system. Medical Physics, 29(11):2552– 2558, 2002.

[5] Jane P Ko and Margrit Betke. Chest ct: automated nodule detection and assessment of change over time—preliminary experience. Radiology, 218(1):267–273, 2001.

[6] Yongbum Lee, Takeshi Hara, Hiroshi Fujita, Shigeki Itoh, and Takeo Ishi-gaki. Automated detection of pulmonary nodules in helical ct images based on an improved template-matching technique. IEEE Transactions on med-ical imaging, 20(7):595–604, 2001.

[7] Qiang Li, Shigehiko Katsuragawa, and Kunio Doi. Computer-aided diag-nostic scheme for lung nodule detection in digital chest radiographs by use of a multiple-template matching technique. Medical physics, 28(10):2070– 2076, 2001

[8] Kenji Suzuki, Junji Shiraishi, Hiroyuki Abe, Heber MacMahon, and Kunio Doi. False-positive reduction in computer-aided diagnostic scheme for de-tecting nodules in chest radiographs by means of massive training artificial neural network1. Academic Radiology, 12(2):191–201, 2005.