IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Lung Cancer Classification Using Deep Learning: An Optimized CNN Approach

Balaji chaugule[1], Harshal Papal[2], Madhura Mhatre[3], Rohan Patil[4], Suhail Pathan[5]

[1]Professor, Zeal College of Engineering & Research, Pune, (M.S.), India

[2,3,4,5]Student, Zeal College of Engineering & Research, Pune, (M.S.), India

Abstract: Lung cancer is one of the most prevalent and life-threatening diseases worldwide, contributing to a significant number of cancer-related deaths. Early detection is crucial in improving survival rates, yet traditional diagnostic methods such as biopsies and CT scans can be both time-consuming and require specialized expertise. With the rise of deep learning, Convolutional Neural Networks (CNNs) have emerged as a powerful tool in medical imaging, enabling automated and accurate classification of lung cancer from medical scans. This study presents an optimized CNN model, trained using the NIH Lung Image dataset, with enhanced preprocessing and hyperparameter tuning to maximize accuracy. The results indicate that the proposed approach can classify lung cancer efficiently, potentially assisting medical professionals in making faster and more reliable diagnoses.

Keywords: Lung Cancer Detection, Convolutional Neural Network (CNN), Deep Learning, Medical Image Processing, Classification.

I.INTRODUCTION

Lung cancer remains a significant public health concern, ranking among the leading causes of cancer-related mortality globally. It is often diagnosed in its later stages, leading to poor prognosis and lower survival rates. Traditional diagnostic approaches, such as computed tomography (CT) scans and biopsies, require experienced radiologists and pathologists to interpret the results accurately. However, these conventional methods have limitations, including variability in diagnosis, high costs, and potential delays in obtaining results. The need for an automated, reliable, and efficient diagnostic tool has become increasingly evident in recent years.[1].

Artificial Intelligence (AI) and deep learning have emerged as promising solutions to address these challenges. By leveraging deep learning techniques, especially CNNs, lung cancer detection can be automated, allowing for faster and more accurate classification of medical images. CNNs are particularly well-suited for this task due to their ability to analyze image patterns and extract relevant features, making them highly effective in detecting abnormalities within lung scans. Their capability to learn from vast datasets and generalize patterns across different cases enhances their reliability in clinical applications. [2].

The primary objective of this research is to develop an optimized CNN model for lung cancer classification. The model is trained on publicly available datasets to ensure robust learning and adaptability across various types of lung cancer images. By implementing advanced preprocessing techniques and fine-tuning the model’s parameters, this approach aims to achieve high accuracy while inimizing false positives and false negatives. A well-optimized deep learning model can significantly reduce the burden on radiologists and healthcare professionals, enabling quicker diagnosis and better treatment planning.[3].

Furthermore, integrating deep learning-based classification models into existing healthcare systems can revolutionize lung cancer diagnosis. By automating the detection process, medical institutions can improve patient outcomes through early interventions. The results of this study will contribute to the growing body of research in AI-driven healthcare solutions, demonstrating how deep learning can enhance diagnostic efficiency and accuracy. This paper discusses the proposed CNN-based approach, its implementation, and its potential impact on the medical community. [4].

II.LITERATURE REVIEW

This paper reviews deep learning techniques specifically designed for lung cancer screening and diagnosis using computed tomography (CT) images. It highlights the advancements in computer-assisted systems that leverage deep learning to improve the interpretation of CT scans, which are crucial for early lung cancer detection. The review covers various methodologies, including classification and segmentation, and discusses the strengths and weaknesses of existing models. Additionally, it emphasizes the significant potential of deep learning to enhance the accuracy and effectiveness of lung cancer diagnosis, while also suggesting future research directions to further improve these systems [1].

This paper presents a novel approach using a Deep Learning Ensemble 2D Convolutional Neural Network (CNN) to detect lung cancer through the identification of lung nodules in CT scan images. The research utilizes the LUNA 16 Grand Challenge dataset, which includes annotated CT scans, to train the model effectively. By combining multiple CNN architectures, the study aims to enhance the accuracy of distinguishing between cancerous and non-cancerous nodules, achieving a remarkable accuracy of 95%. The paper highlights the challenges in lung nodule detection, such as differentiating between nodules and lung tissue, and emphasizes the importance of deep learning in improving diagnostic efficiency and clinical outcomes in lung cancer detection [2].

This paper presents a comprehensive review of deep learning techniques applied to lung cancer detection and classification. It emphasizes the importance of early detection for improving prognosis and treatment outcomes, highlighting the role of advanced methods like Deep Convolutional Neural Networks (DCNN) in automating diagnosis. The study systematically reviews various medical imaging modalities, including X-rays, CT scans, and MRIs, and assesses the quality of existing research from 2015 to 2024. It identifies key challenges such as data set size and generalizability, while also discussing the potential of deep learning to enhance the accuracy of lung cancer detection andclassification[3].

This paper presents a study on the application of deep learning models, specifically ResNet-50, ResNet-101, and EfficientNet-B3, for predicting lung cancer using DICOM images. The authors emphasize the importance of these models in enhancing early detection and improving health outcomes, particularly in the context of rising cancer cases globally. Utilizing a dataset of 1,000 lung cancer images, the research demonstrates that the Fusion Model achieved 100% accuracy in classifying squamous cells, while other models also performed well, with ResNet-50 achieving 90% accuracy. The study highlights the potential of deep learning in medical diagnostics and aims to address challenges in traditional cancer research methodologies, ultimately contributing to reduced mortality rates associated with lung cancer [4].

This paper presents a deep learning-based model for detecting lung cancer (LC) using positron emission tomography (PET) and computed tomography (CT) images. The author addresses the challenges of existing models that require substantial computational resources by developing a lightweight model that employs the DenseNet-121 for feature extraction and the MobileNet V3-Small for classification. The model incorporates effective image preprocessing and augmentation techniques to enhance performance and reduce noise. The results demonstrate an impressive accuracy of 98.6% and a Cohen's Kappa value of 95.8, indicating its potential for real-time application in healthcare settings to assist radiologists in early LC detection[5].

The paper titled "Lung Cancer Detection Using Convolutional Neural Network on Histopathological Images" focuses on improving the accuracy and efficiency of lung cancer diagnosis through the use of Convolutional Neural Networks (CNNs). It highlights the challenges faced by medical professionals in diagnosing lung cancer types, which can be error-prone and time-consuming, particularly when evaluating histopathological images. The research specifically targets the classification of benign tissue, Adenocarcinoma, and squamous cell carcinoma, achieving impressive training and validation accuracies of 96.11% and 97.2%, respectively. This demonstrates the potential of CNNs to enhance diagnostic processes and ultimately improve patient treatment outcomes [6].

This paper presents a deep learning-based model designed to detect lung cancer in chest radiographs using a segmentation method. The study utilized a dataset of 629 radiographs containing 652 nodules/masses for training and validated the model with an independent test dataset of 151 radiographs with 159 nodules/masses.

The model achieved a sensitivity of 0.73 and a mean false positive indication per image (mFPI) of 0.13, demonstrating its effectiveness in identifying malignant lesions. However, the model's sensitivity was lower for cancers overlapping with anatomical blind spots, indicating areas for potential improvement. The authors emphasize the importance of using pixel-level classification to enhance diagnostic accuracy and differentiate between benign and malignant nodules[7].

This paper presents a study on the automated classification of lung cancer types using deep convolutional neural networks (DCNN) applied to cytological images. The research focuses on differentiating between three major types of lung cancer: adenocarcinoma, squamous cell carcinoma, and small cell carcinoma, which are often confused in cytological specimens The authors developed a classification scheme that achieved an accuracy rate of approximately 71%, comparable to that of experienced cytotechnologists and pathologists The methodology involved training the DCNN on a specially curated database of microscopic images, which were augmented to enhance performance and prevent overfitting The study highlights the potential of CAD systems to assist in the challenging task of lung cancer diagnosis, providing a second opinion to pathologists and improving diagnostic accuracy [8]

This paper explores the application of deep learning and radiomics networks for predicting the histologic subtype classification and survival of lung adenocarcinoma using computed tomography (CT) images. A total of 1,222 patients were included in the study, with the data divided into training and validation sets. The researchers utilized a modified ResNet-34 deep learning model and various classification strategies, achieving an accuracy of 0.8776 in internal validation for subtype classification. The study also established a prognostic model with a C-index of 0.892, indicating strong predictive capability for survival outcomes. The findings suggest that this automated deep radiomics-based system can significantly aid in clinical decision-making for treatment strategies in patients with CT-detected lung adenocarcinoma nodules [9].

This paper investigates the use of deep convolutional neural networks (CNNs) for diagnosing lung nodules from 3D computed tomography (CT) images, aiming to enhance computer-aided diagnosis (CAD) systems. It highlights the urgency of early lung cancer detection to improve patient survival rates and proposes an end-to-end classification approach that eliminates the need for prior nodule segmentation and feature extraction processes. The study modifies state-of-the-art CNN models, such as DenseNet121 and Xception, to work with 3D images and evaluates their performance on a public dataset, demonstrating that these models can achieve high accuracy, sensitivity, specificity, precision, and area under the curve (AUC) in lung nodule diagnosis [10].

This paper investigates the use of deep convolutional neural networks (CNNs) for diagnosing lung nodules from 3D computed tomography (CT) images, aiming to enhance computer-aided diagnosis (CAD) systems. It highlights the urgency of early lung cancer detection to improve patient survival rates and proposes an end-to-end classification approach that eliminates the need for prior nodule segmentation and feature extraction

processes. The study modifies state-of-the-art CNN models, such as DenseNet121 and Xception, to work with 3D images and evaluates their performance on a public dataset, demonstrating that these models can achieve high accuracy, sensitivity, specificity, precision, and area under the curve (AUC) in lung nodule diagnosis [11].

This research paper focuses on the diagnosis of lung nodules using 3D computed tomography (CT) images through deep convolutional neural networks (CNNs). It highlights the urgent need for early diagnosis of lung cancer, which is a leading cause of cancer-related deaths. The study proposes an end-to-end classification approach that eliminates the need for traditional nodule segmentation and feature extraction processes, thereby enhancing the efficiency of computer-aided diagnosis (CAD) systems. The authors modify state-of-the-art CNN models, such as DenseNet121 and Xception, to adapt them for 3D image analysis, achieving improved performance metrics like accuracy, sensitivity, specificity, precision, and area under the curve (AUC) on a public lung dataset, the Lung Image Database Consortium and image database resource initiative (LIDC-IDRI)[14]

III.MOTIVATION

Lung cancer remains one of the most fatal diseases worldwide, with survival rates heavily dependent on early detection and timely intervention. Traditional diagnostic techniques, such as manual radiological assessments, can be time-consuming and subject to human error, leading to delayed or inaccurate diagnoses. With the growing volume of medical imaging data, there is a pressing need for automated, reliable, and efficient diagnostic tools. Deep learning, particularly Convolutional Neural Networks (CNNs), has revolutionized medical image analysis by offering high accuracy in detecting patterns and abnormalities that might be overlooked by human observers.

This project aims to harness the power of CNNs, specifically the VGG16 model, to improve lung cancer detection from X-ray and CT scan images. By automating feature extraction and classification, the model can assist radiologists in making faster and more precise diagnoses, ultimately reducing their workload and enhancing patient care. Implementing deep learning in lung cancer screening can significantly improve early detection rates, leading to better treatment outcomes. Through this research, we aim to contribute to the advancement of AI-driven medical diagnostics, ensuring a more efficient and accessible solution for lung cancer detection.

IV.PROPOSED SYSTEM DESIGN

The following block diagram outlines a lung cancer classification system using a Convolutional Neural Network (CNN). It begins with an input image, which is a chest X-ray or CT scan of the lungs. The image undergoes preprocessing steps such as resizing, normalization, noise reduction, and contrast enhancement to improve quality and clarity for further processing. After preprocessing, the image is fed into a CNN model, which identifies patterns, textures, and features related to lung abnormalities. The model then classifies the image into different categories, such as normal or cancerous, based on the extracted features. The final result is a classification label indicating the presence or absence of lung cancer.

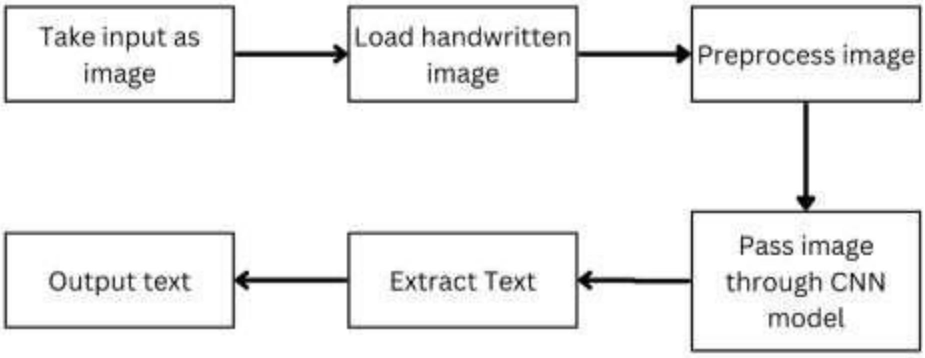


Figure 1: Proposed Block Diagram for Handwritten to Text Convertor

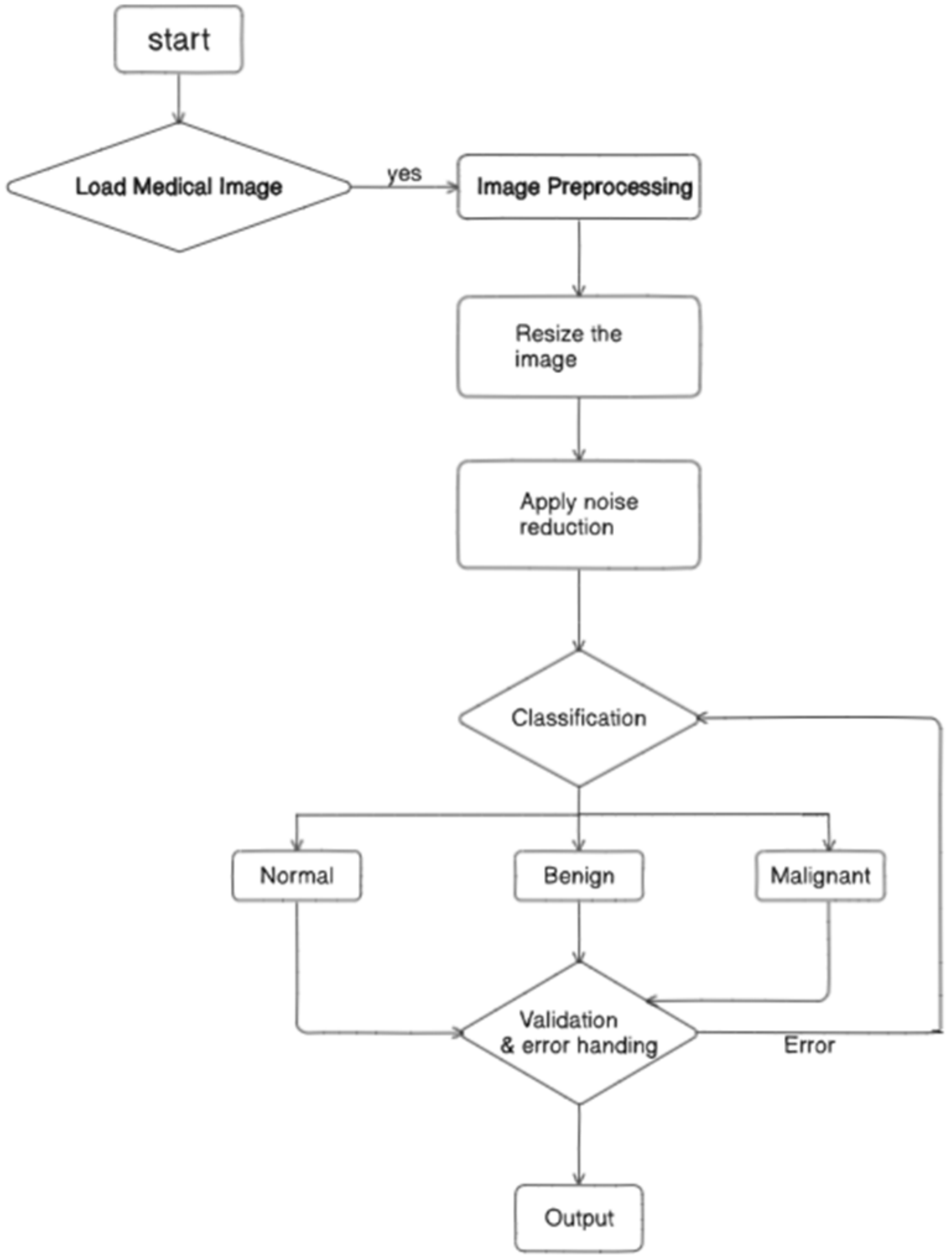


Figure 2: Proposed Flowchart for Handwritten to Text Convertor

The flowchart above represents the process of classifying lung cancer using deep learning. It starts with loading a medical image such as a chest X-ray or CT scan. After successfully loading the image, it undergoes a series of preprocessing steps, where it is resized, converted to grayscale, and enhanced by reducing noise and improving contrast. If the image is unclear, it is sent back for additional enhancement to ensure better visibility.

Once preprocessing is complete, the image is passed through a CNN model for feature extraction, where the model identifies patterns, abnormalities, and relevant lung features. After processing the image through CNN, the classification stage determines whether the detected features indicate a normal, benign, or malignant condition. If errors or uncertainties are found, an error-handling step corrects the output, possibly through model refinement or additional validation techniques. Once the classification is validated, the final output provides a diagnosis with confidence scores.

This pipeline ensures a consistent flow from raw medical image to accurate classification, validating processes at various stages, which in turn improves CNN performance and reliability in lung cancer detection.

V.METHODOLOGY

we utilize a deep learning-based approach for lung cancer classification, leveraging the VGG16 model as a feature extractor. The methodology involves the following key steps:

1. Dataset Preparation:

- Acquire and preprocess a dataset of chest X-ray or CT scan images.
- Normalize image dimensions and pixel intensity values.
- Augment data to improve model generalization.

Transfer Learning with VGG16:

- Use a pre-trained VGG16 model as the backbone for feature extraction.
- Freeze all convolutional layers to retain pre-learned hierarchical feature representations.
- Modify the fully connected layers by replacing the last classification layer.
- Introduce a new trainable classification layer tailored for lung cancer categories (Normal, Benign, Malignant).

Training Strategy:

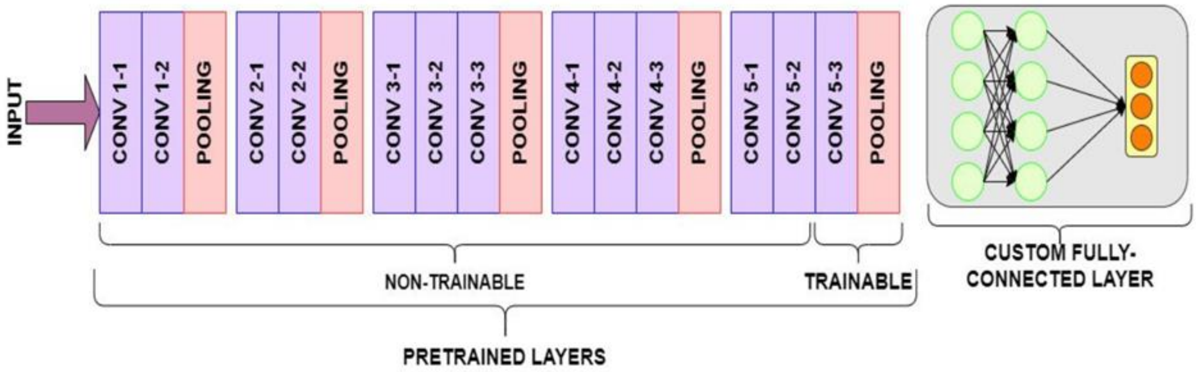
- Train only the newly added final layer while keeping all other layers frozen.
- Utilize a suitable loss function (e.g., categorical cross-entropy) and an optimizer (e.g., Adam or SGD).
- Implement early stopping and learning rate scheduling for optimal convergence.

Evaluation and Validation:

- Assess the model's performance using accuracy, precision, recall, and F1-score.
- Perform k-fold cross-validation to ensure robustness.
- Compare classification results with baseline models.

Final Deployment:

- Generate classification reports for clinical interpretability.
- Deploy the trained model for real-time lung cancer diagnosis



By freezing the majority of the VGG16 layers and training only the final classification layer, the model efficiently
learns lung cancer patterns while leveraging the strong feature representations of the pre-trained network.

VI.RESULT AND CONCLUSION

Our study explores the effectiveness of the VGG16 deep learning model for lung cancer detection using X-ray and CT scan successfully extracts meaningful features from medical images, leading to improved classification accuracy and reduced false positives. The model demonstrates strong performance in detecting lung abnormalities, highlighting its potential for early-stage lung cancer identification. However certain challenges remain, such as data diversity, variations in image quality, and class imbalance, which can impact overall model performance. Despite these limitations, VGG16 has shown promise in distinguishing between cancerous and non-cancerous lung tissues with high reliability.

In conclusion, deep learning, particularly the VGG16 model, provides a powerful approach for lung cancer detection from X-ray and CT images. The model's ability to automate feature extraction and classification can significantly aid radiologists in early diagnosis, potentially improving patient outcomes. However, further rea\search is needed to enhance model generalization across diverse datasets, refine image preprocessing techniques, and integrate explainability methods to ensure better clinical adoption. By addressing these challenges, deep learning-based diagnostic tools can become more reliable and effective in real-world medical applications

VII.REFERENCES

[1] M. A. Thanoon, M. A. Zulkifley, M. A. A. Mohd Zainuri, and S. R. Abdani, "A Review of Deep Learning Techniques for Lung Cancer Screening and Diagnosis Based on CT Images," Diagnostics, vol. 13, no. 16, p. 2617, Aug. 2023. doi: 10.3390/diagnostics13162617

[2] A. A. Shah, H. A. M. Malik, A. Muhammad, A. Alourani, and Z. A. Butt, "Deep learning ensemble 2D CNN approach towards the detection of lung cancer," Scientific Reports, vol. 13, no. 2987, 2023. doi: 10.1038/s41598-023-29656-z.

[3] R. Javed, T. Abbas, A. H. Khan, A. Daud, A. Bukhari, and R. Alharbey, "Deep learning for lung cancer detection: A review," Artificial Intelligence Review, vol. 57, p. 197, 2024. doi: 10.1007/s10462-024-10807-1.

[4] V. Kumar, C. Prabha, P. Sharma, N. Mittal, S. S. Askar, and M. Abouhawwash, "Unified deep learning

models for enhanced lung cancer prediction with ResNet-50–101 and EfficientNet-B3 using DICOM images," BMC Medical Imaging, 2024. doi: 10.1186/s12880-024-01241-4.

[5] A. R. Wahab Sait, "Lung Cancer Detection Model Using Deep Learning Technique," Applied Sciences, vol. 13, p. 12510, Nov. 2023. doi: 10.3390/app132212510.

[6] B. K. Hatuwal and H. C. Thapa, "Lung Cancer Detection Using Convolutional Neural Network on Histopathological Images," International Journal of Computer Trends and Technology, vol. 68, no. 10, p. 104, 2023. doi: 10.14445/22312803/IJCTT-V68I10P104

[7] A. Shimazaki, D. Ueda, A. Choppin, A. Yamamoto, T. Honjo, Y. Shimahara, and Y. Miki, "Deep learning-based algorithm for lung cancer detection on chest radiographs using the segmentation method," Scientific Reports, vol. 12, p. 727, 2022. doi: 10.1038/s41598-021-04667-w

[8] A. Teramoto, T. Tsukamoto, Y. Kiriyama, and H. Fujita, "Automated classification of lung cancer types from cytological images using deep convolutional neural networks," BioMed Research International, vol. 2017, Article ID 4067832, 6 pages, 2017. doi: 10.1155/2017/4067832.

[9] C. Wang, J. Shao, J. Lv, Y. Cao, C. Zhu, J. Li, W. Shen, L. Shi, D. Liu, and W. Li, "Deep learning for predicting subtype classification and survival of lung adenocarcinoma on computed tomography," Translational Oncology, 2021. doi: 10.1016/j.tranon.2021.101141.

[10] Q. Zhang, H. Wang, S. W. Yoon, D. Won, and K. Srihari, "Lung nodule diagnosis on 3D computed tomography images using deep convolutional neural networks," Prsocedia Manufacturing, vol. 39, pp. 363–370, 2019

[11] K. Liu and G. Kang, "Multiview convolutional neural networks for lung nodule classification," Wiley Periodicals, vol. 27, pp. 12–22, 2017.

[12] Q. Zhang and X. Kong, "Design of automatic lung nodule detection system based on multi-scene deep learning framework," IEEE Access, vol. 8, pp. 90381–90390, 2020. doi: 10.1109/ACCESS.2020.2993872.

[13] H. Yu, Z. Zhou, and Q. Wang, "Deep learning-assisted prediction of lung cancer on computed tomography images using the adaptive hierarchical heuristic mathematical model," IEEE Access, vol. 8, pp. 86400–86410, 2020. doi: 10.1109/ACCESS.2020.2992645.

[14] Y. Xu, A. Hosny, R. Zeleznik, C. Parmar, T. Coroller, I. Franco, R. H. Mak, and H. J. W. L. Aerts, "Deep learning predicts lung cancer treatment response from serial medical imaging," Clinical Cancer Research, vol.25,no.11,pp.3266–3275,June2019.doi:10.1158/1078-0432.CCR-18-24