|| Volume 9 || Issue 5 || May 2025 || ISSN (Online) 2456-0774
INTERNATIONAL JOURNAL OF ADVANCE SCIENTIFIC RESEARCH
AND ENGINEERING TRENDS

Lung Insight: Enhanced Detection of Lung Cancer Using Multi-Dataset Integration and Image Optimization

Prof. Rashmi Badave[1], Prof. Vanita Kshirsagar[2], Kunal Shinde[3], Aditya Waghmare[4], Mayuresh Kalal[5], Samyak Waghmare[6]

Artificial Intelligence and Data Science. D.Y. Patil Institute of Technology, Pimpri, Pune, India[1,2,3,4,5,6]

rashmi.badave@gmail.com[1],vanita.kshirsagar@gmail.com[2],shindekunal4285@gmail.com[3],adityawaghmare2303@gmail.com[4], mayureshkalal28@gmail.com[5], samyakwaghmare0503@gmail.com[6]

Abstract: Lung cancer is one of the most fatal types of cancer worldwide, and early detection plays a critical role in improving survival rates. This project, Lung Insight, aims to enhance lung cancer detection from thoracic CT scans by integrating multiple datasets and applying advanced image optimization techniques. The model will be based on deep learning, specifically leveraging Convolutional Neural Networks (CNN) to classify lung nodules as malignant or benign. Multi-dataset integration, including datasets like LIDC-IDRI, NSCLC, and LUNA16, will ensure model generalizability, while image enhancement techniques like Contrast Limited Adaptive Histogram Equalization (CLAHE) will improve the clarity of CT scans for better feature extraction. The ultimate goal is to build a robust and efficient system that aids radiologists in early lung cancer diagnosis, improving patient outcomes through timely intervention. The system will also employ model optimization techniques like pruning and knowledge distillation to ensure faster inference and deployment in real-time clinical environments. This study groups the existence of lung cancer in CT scan pictures and blood tests using a framework for detecting the disease using a support vector machine and image processing. The focus of the work described in this article is on the design and enhancement of a framework for determining lung cancer using CT scan pictures. In order to ensure improved survival rates, it is crucial to arrange different tumour types. Lung cancer's cycle of development is always being tested. The framework detects the many stages of lung cancer, enabling experts to accurately and swiftly diagnose lung cancer from a wealth of information.

Keywords: *NLP, machine learning, Kaggle, CNN,* CT scan pictures

---

I.INTRODUCTION:

Lung cancer remains one of the leading causes of cancer-related deaths worldwide, largely due to the lack of early detection and the difficulty in distinguishing malignant nodules from benign ones. Current diagnostic approaches rely heavily on radiologists' manual interpretation of imaging data, which is prone to human error, especially when dealing with small nodules or overlapping anatomical structures.

II.LITERATURE SURVEY

In [1] Cancer Cell Detection Using Hybrid Neural Networks (CCDC-HNN) is an innovative approach for early and precise detection. Deep neural networks extract characteristics from CT scan pictures. The precision of feature extraction is essential for the early identification of malignant cells; therefore, it safeguards the patient from this lethal illness. This research uses an advanced 3D-convolutional neural network (3D-CNN) to improve diagnostic accuracy. The proposed method also facilitates differentiation between benign and malignant tumors.

According to [2] They have used the LUNA 16 Grand Challenge dataset, which is available on their website. The dataset includes a CT scan along with annotations that improve understanding of the data and specifics of each CT scan. Artificial neural networks form the foundation of deep learning, which functions similarly to neurones in the human brain. We produce a comprehensive dataset of CT scans to train the deep learning mode. We train convolutional neural networks (CNNs) using a dataset that distinguishes between malignant and non-cancerous photos. We create a collection of training, validation, and testing datasets for our Deep Ensemble 2D CNN. Deep Ensemble 2D CNN comprises three distinct CNNs, each with varying layers, kernels, and pooling methodologies.\\

In [3] Lung-EffNet is a new transfer learning-based predictor designed to categorize lung cancer. We construct Lung-EffNet based on the EfficientNet architecture, and further modify the classification head by adding supplementary top layers. We assess Lung-EffNet using five variations of EfficientNet, specifically B0 through B4. The trials use the benchmark dataset "IQ-OTH/NCCD" for lung cancer patients, categorized as benign, malignant, or normal according to the presence or absence of lung cancer. Several data augmentation techniques addressed the class imbalance problem and mitigated biases. We compared the effectiveness of the proposed fine-tuned pre-trained EfficientNet to alternative pre-trained CNN architectures. The anticipated results indicate that Lung-EffNet, based on EfficientNetB1, surpasses other CNNs in accuracy and efficiency.

According to [4] Researchers have thoroughly investigated deep learning methodologies to aid in the interpretation of CT scans for lung cancer detection, in line with the progression of computer-assisted systems. The objective of this study is to provide a comprehensive analysis of the deep learning methodologies established for the screening and diagnosis of lung cancer. This study provides an overview of deep learning (DL) techniques, recommended DL methods for lung cancer applications, and the innovations of the examined approaches. This paper examines

two primary techniques of deep learning in the screening and diagnosis of lung cancer: classification and segmentation approaches. We will also examine the merits and drawbacks of contemporary deep learning models. The investigation indicates a substantial promise for deep learning techniques to provide accurate and efficient computer-assisted lung cancer screening and diagnosis using CT images.

According to [5] Using six different deep learning algorithms—Convolutional Neural Network (CNN), CNN Gradient Descent (CNN GD), VGG-16, VGG-19, Inception V3, and ResNet-50—the method shows better performance across all criteria. We evaluate the suggested technique using CT scan pictures and histopathology images. We developed six deep learning models for the efficient detection of lung cancer. The techniques used for lung cancer detection include CNN, CNN GD, Inception V3, ResNet-50, VGG-16, and VGG-19. We conducted the experimental studies using CT scan pictures and histopathological images. We evaluated the suggested method using recognition accuracy, F-Score, precision, sensitivity, specificity, and other metrics. The suggested technique has intrinsic advantages, making it an effective approach for lung cancer screening that will benefit those in need.

According to [6] employs image processing and k-Nearest Neighbour algorithms to classify lung cancer stages from CT scan pictures. The primary aim of this work is to develop an image processing approach for extracting lung cancer characteristics from CT scan pictures. Extracting characteristics from the segmented picture may facilitate the detection of lung cancer. The proposed approach involves the following image processing techniques: data collection, data pre-processing, feature selection, and lung cancer classification. The preprocessing used a median filter to eliminate noise present in the photos. We must extract three attributes: area, perimeter, and centroid. We ultimately used the dataset containing these attributes as the input for lung cancer categorization.

According to [7] We evaluate numerous segmentations, feature extraction, and classification algorithms such as artificial neural networks, convolutional neural networks, support vector machines, grey-level co-occurrence matrixes, and discrete wavelet transforms, among others. Early identification of lung cancer is crucial due to the rising death rate, since it might prolong the patient's life. The essential components of any CAD system are pre-processing, segmentation, feature extraction, and classification. We use various methods like SVM, DWT, ANN for contour identification, the watershed algorithm, and CNN to achieve high accuracy and sensitivity in CT scan pictures. This paper provides an overview of current technologies for lung cancer detection and highlights advancements in these technologies over recent years.

According to [8] An automated technique employs computed tomography (CT) scans to identify lung cancer in its initial phase. The primary aim of this research project is to attain standard performance accuracy. We came up with a new way to find lung cancer by using different parts of computed tomography images. These parts could be enhanced, filtered on the median, divided into groups, given features, and categorised using support vector machines.

According to [9] The system for diagnosing lung cancer consists of two main components: a segmentation module based on the UNETR network, and a classification module that uses a self-supervised network to classify the segmentation output as benign or malignant. The suggested method offers a robust instrument for the early diagnosis and treatment of lung cancer using 3D-input CT scan data. We have conducted comprehensive trials to enhance segmentation and classification outcomes.

According to [10] YOLOv5, a sophisticated object recognition system, is utilized in medical imaging to detect lung cancer. We acquired a dataset from Kaggle, which included chest X-rays and their corresponding annotations, to train and evaluate the algorithm. We developed an algorithm for the detection of malignant lung lesions using the YOLOv5 model. The training methodology included optimizing hyperparameters and using augmentation approaches to improve the model's performance. The trained YOLOv5 model demonstrated remarkable efficacy in detecting lung cancer lesions, exhibiting elevated accuracy and recall rates. An independent test set demonstrated its accuracy in identifying cancerous regions in chest radiographs, surpassing previous methodologies. The YOLOv5 model exhibited computational efficiency, facilitating real-time identification and rendering it appropriate for incorporation into clinical operations.

III.RESEARCH MTHODOLOGY

In general study the lung cancer diagnosis involves many different technologies. The CAD systems are developed for identification of lung cancer in initial stage. A general Computer Aided Diagnosis set-up consists of several steps in identification of the lung cancer. The following techniques are 1) Pre-processing and segmentation 2) Nodule Detection 3) Nodule Segmentation 4) Feature Extraction 5) Classification. Figure shows the structure of a basic CAD set-up in identification of lung cancer.

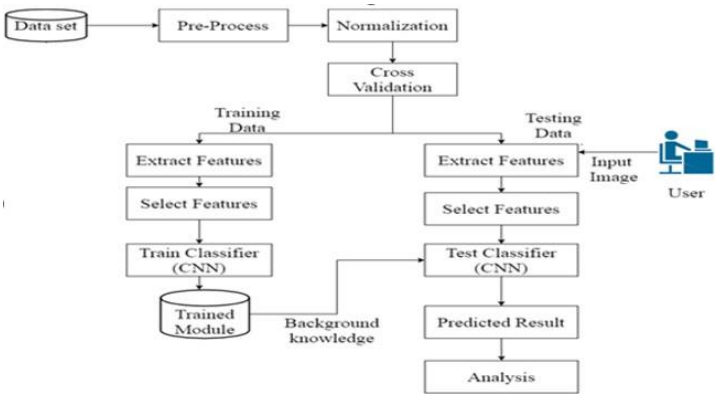


Figure 1: Research methodology of proposed system Implement Model

Data Collection

To ensure robustness, diverse datasets from different modalities

(CT scans, X-rays, MRI) will be gathered. Public datasets like LIDC-IDRI can be combined with private clinical data.The dataset used in this project is the 'IQ-OTH/NCCD Lung Cancer Dataset'. It contains CT scan images categorized into three classes:

- Benign
- Malignant
- Normal

These categories represent different stages and types of lung cancer, and the images are stored in respective folders.

The dataset is suitable for image classification tasks and is used here to train the MobileNetV2 model.

Data Preprocessing and Augmentation

Keras' ImageDataGenerator is used for real-time image preprocessing and augmentation.

This includes:

- Rescaling pixel values to [0, 1] range
- Random rotations (20 degrees)
- Zoom and shear transformations
- Horizontal flipping

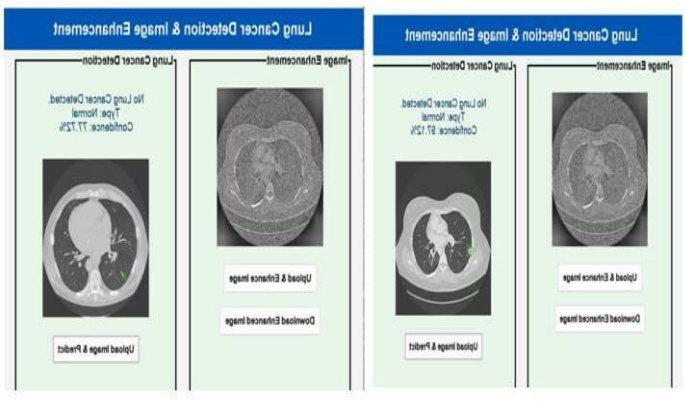
These augmentations help prevent overfitting and improve model generalization.

Model Architecture

The base model used is MobileNetV2, which is pre-trained on the ImageNet dataset.The top classification layers are removed (`include_top=False`), and a custom head is added for lung cancer classification. The custom layers are:

- GlobalAveragePooling2D: Reduces the spatial dimensions and retains features
- Dense layer with 128 units and ReLU activation
- Dropout layers (0.4 and 0.3) to prevent overfitting
- Output Dense layer with 3 units and softmax activation to classify into 3 categories

IV.RESULT AND ACCURACY



.

The model achieved high training and validation accuracy after fine-tuning. Predictions are displayed in the GUI with class labels and confidence. This project provides a complete workflow from preprocessing to real-time detection

V.CONCLSUION

This project integrates deep learning and user interface design to provide a powerful lung cancer detection tool. The MobileNetV2 model ensures fast and accurate classification, while image enhancement improves visual quality. With further enhancements and validations, this tool can assist radiologists in early lung cancer diagnosis.

VI.REFERENCES

[1] Wankhade, Shalini, and S. Vigneshwari. ”A novel hybrid deep learning method for early detection of lung cancer using neural networks.” Healthcare Analytics 3 (2023): 100195.

[2] Shah, Asghar Ali, et al. ”Deep learning ensemble 2D CNN approach towards the detection of lung cancer.” Scientific Reports 13.1 (2023): 2987.

[3] Raza, Rehan, et al. ”Lung-EffNet: Lung cancer classification using EfficientNet from CT-scan images.” Engineering Applications of Artificial Intelligence 126 (2023): 106902.

[4] Thanoon, Mohammad A., et al. ”A review of deep learning techniques for lung cancer screening and diagnosis based on CT images.” Diagnostics 13.16 (2023): 2617.

[5] Rajasekar, Vani, et al. ”Lung cancer disease prediction with CT scan and histopathological images feature analysis using deep learning techniques.” Results in Engineering 18 (2023): 101111.

[6] Abdullah, Mohd Firdaus, et al. ”Classification of lung cancer stages from CT scan images using image processing and k-Nearest neighbours.” 2020 11th IEEE Control and System Graduate Research Colloquium (ICSGRC). IEEE, 2020.

[7] Katiyar, Preeti, and Krishna Singh. ”A Comparative study of Lung Cancer Detection and Classification approaches in CT images.” 2020 7th International Conference on Signal Processing and Integrated Networks (SPIN). IEEE, 2020.

[8] Hoque, Ariful, et al. ”Automated detection of lung cancer using CT scan images.” 2020 IEEE Region 10 Symposium (TENSYMP). IEEE, 2020.

[9] Said, Yahia, et al. ”Medical images segmentation for lung cancer diagnosis based on deep learning architectures.” Diagnostics 13.3 (2023): 546.

[10] Gunasekaran, Karthick Prasad. ”Leveraging object detection for the identification of lung cancer.” arXiv preprint arXiv:2305.15813 (2023).