International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

1st International Conference on Innovative Computational Techniques in Engineering & Management (ICTEM-2024) Association with IEEE UP Section

AI in Healthcare: Deep Learning Solutions for Lung Cancer Detection

Himanshu Tyagi[1], Aditya Gupta[2], Saurabh Srivastava[3], Neeraj Kumari[4], Aryan Sharma[5], Nikhil Sharma[6]

[1]*Department of Computer Science & Engineering (AI & ML),Moradabad Institute of Technology, Moradabad, India*
[2]*Department of Computer Science & Engineering (AI & ML),Moradabad Institute of Technology, Moradabad, India*
[3]*Department of Computer Science & Engineering (DS),Moradabad Institute of Technology, Moradabad, India*
[4]*Department of Computer Science & Engineering (AI & ML),Moradabad Institute of Technology, Moradabad, India*
[5]*Department of Computer Science & Engineering (AI & ML),Moradabad Institute of Technology, Moradabad, India*
[6]*Department of Computer Science & Engineering (AI & ML),Moradabad Institute of Technology, Moradabad, India*
[a)]*12tyagihimanshu@gmail.com,* [b)]*adityaofficial1056@gmail.com,* [c)]*srbh.spn@gmail.com,* [c)]*neelu.rajpoot80@gmail.com*
[e)]*aryan070sharma@gmail.com,* [f)]*nsharma.mit21@gmail.com*
DOI: https://doi.org/10.55248/gengpi.6.sp525.1957

ABSTRACT:

Lung cancer remains one of the leading causes of cancer-related deaths worldwide, emphasizing the urgent need for early detection and accurate diagnosis. Traditional diagnostic methods, such as imaging and biopsy, often face limitations in sensitivity and specificity, leading to delayed diagnosis and suboptimal treatment outcomes. Artificial intelligence (AI), particularly deep learning, has emerged as a powerful tool in healthcare, offering innovative solutions for lung cancer detection. Deep learning algorithms, leveraging large datasets of medical imaging, such as computed tomography (CT) scans and chest X-rays, have demonstrated remarkable accuracy in identifying lung cancer nodules and distinguishing between malignant and benign lesions. Moreover, AI-driven approaches can assist in predicting patient outcomes and guiding personalized treatment plans, potentially improving survival rates. In this study, a deep learning-based model has been developed that uses the X-ray image of the chest to predict whether a person is suffering from lung cancer. The developed deep learning model offers real-time analysis, decreases false positives and negatives, and increases diagnostic accuracy. Furthermore, by providing a user-friendly interface for medical image processing, it aims to supplement the knowledge of healthcare professionals and ultimately improve clinical decision-making.

Keywords: Lung Cancer, Deep Learning, X-rays

1. INTRODUCTION

Lung cancer is one of the leading causes of death worldwide and is considered one of the most dangerous malignant tumors affecting human health. It has the highest mortality rate among all malignant tumors, making it the leading cause of cancer-related deaths for both men and women globally(Pang et al., 2020). Lung adenocarcinoma (LUAD) and lung squamous cell carcinoma (LUSC) are the two primary forms of lung cancer, and they have different molecular and clinical characteristics(Ye et al., 2020). The LUSC is more advanced and intrusive in older males. Age and other comorbidities like smoking contribute to its bad prognosis. LUSC is a prevalent malignancy with a diverse molecular profile. Patients with LUSC often have mutations in genes like TP53, CDKN2A, PTEN, PIK3CA, KEAP1, MLL2, HLA-A, NFE2L2, NOTCH1, and RB1(Lima et al., 2024).

Lung cancer is a complex illness that has a high death rate and incidence. Lung cancer, with an estimated incidence of 11.6% in men and women, was the most common cancer diagnosed in 2018, according to global cancer data, behind non-melanoma skin cancer. At 18.4% of all cancer deaths, lung cancer is the primary cause of cancer-related deaths. Lung cancer is a complex illness that has a high death rate and incidence. Lung cancer, with an estimated incidence of 11.6% in men and women, was the most common cancer diagnosed in 2018, according to global cancer data, behind non-melanoma skin cancer. At 18.4% of all cancer deaths, lung cancer is the primary cause of cancer-related deaths(Lima et al., 2024). According to the Surveillance, Epidemiology, and End Results (SEER) program data from 1992 to 2008, 1,450,837 non-pulmonary cancer survivors were identified, with 25,472 developing Second primary lung cancer (SPLC) over an average follow-up of 5.7 years (standard deviation of 3.6 years). More than half of these patients (57%) succumbed to the disease (Liu et al., 2021). There are about 1.8 million new cases of lung cancer per year (13% of all tumors), 1.6 million deaths (19.4% of all tumors) in the world(Pang et al., 2020). The United States of America is expected to have around 1,958,310 new instances of cancer and

609,820 cancer-related fatalities in 2023, with lung cancer accounting for 350 of those deaths every day(Naseer et al., 2023). Non-small cell lung cancer (NSCLC) represents approximately 85% of all lung cancers, with a 5-year overall survival rate of about 15% (Gupta et al., 2022).

Early detection of lung cancer improves the prognosis and is critical to effective therapy. Lung cancer can be detected by several techniques, including X-ray, blood test, biopsy, and CT. The diverse sizes, forms, locations, and concentrations of pulmonary nodules make nodule detection a difficult process. Timely lung cancer detection has been achieved by the application of computational intelligence approaches(Naseer et al., 2023). Clinically approved therapy for EGFR-positive lung cancer are predictive biomarkers. Through molecular testing on tissue removed during a biopsy, the oncogene mutation status is traditionally determined. But more sophisticated, automated methods have also been developed, like computer-aided diagnostics (CAD), which uses CT analysis(Silva et al., 2021). Predicting cancer survivability holds significant theoretical value and addresses important practical needs. It can offer valuable insights for improving clinical decision-making and advancing personalized medicine (Liu et al., 2021). CT and X-rays are the conventional methods used to detect the existence of lung cancer(Yu et al., 2020). The field of image classification has benefited from the application of deep learning and deep neural networks (Yu et al., 2020). Image classification is a computer vision task where a model assigns labels to images based on their content. Once trained, the model can classify new images into predefined categories, making it useful in applications like medical diagnosis, autonomous vehicles, and facial recognition. Popular frameworks for image classification include TensorFlow, PyTorch, and OpenCV(Srivastava & Ahmed, 2021, 2024).

Deep learning has emerged as a powerful approach for lung cancer detection, leveraging advanced neural networks to analyze medical imaging data such as computed tomography (CT) scans and chest X-rays. By automatically learning complex patterns and features within the images, deep learning algorithms can identify lung cancer nodules with high accuracy, distinguish between benign and malignant lesions, and even predict disease progression. These AI-driven techniques have the potential to surpass traditional diagnostic methods in terms of sensitivity and specificity, enabling earlier detection and more precise diagnoses. In this study, a deep learning-based model has been developed that will be very beneficial for detecting lung cancer.

2. METHODOLOGY

Artificial Intelligence (AI) has revolutionized healthcare, particularly in early disease detection. Deep learning, a subset of AI, plays a pivotal role in lung cancer detection by analyzing complex medical imaging data with exceptional accuracy. Algorithms trained on vast datasets of CT scans for identifying subtle patterns and abnormalities, often missed by traditional methods, enable earlier and more reliable diagnoses. The work flow diagram is show

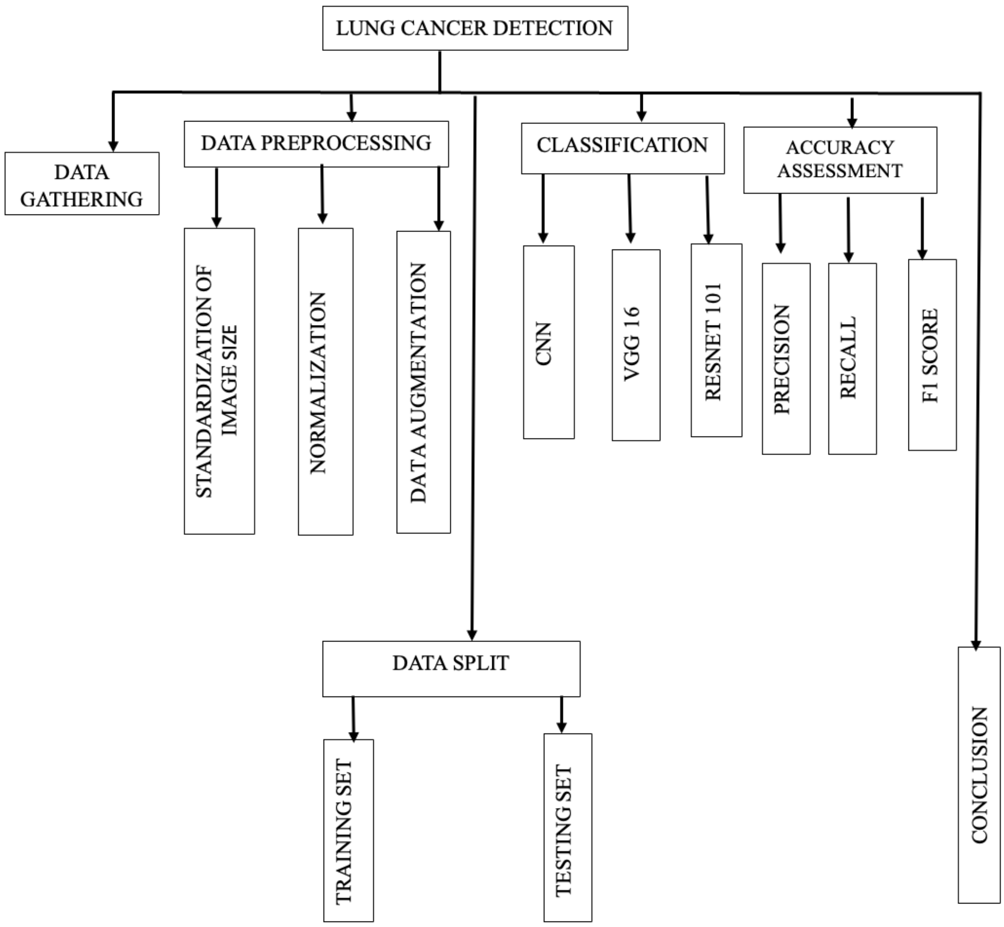

Fig 1 Work Flow Diagram

2.1 Data Gathering: The dataset used to build the model was collected from reliable medical imaging sources. It contains 1786 computed tomography (CT) scan images categorized into two classes:

2.2 Data Preprocessing

In this phase, the CT scan images were pre-processed to make the original data more suitable for the development of deep learning and transfer learning models. The required steps is shown in Figure 2.

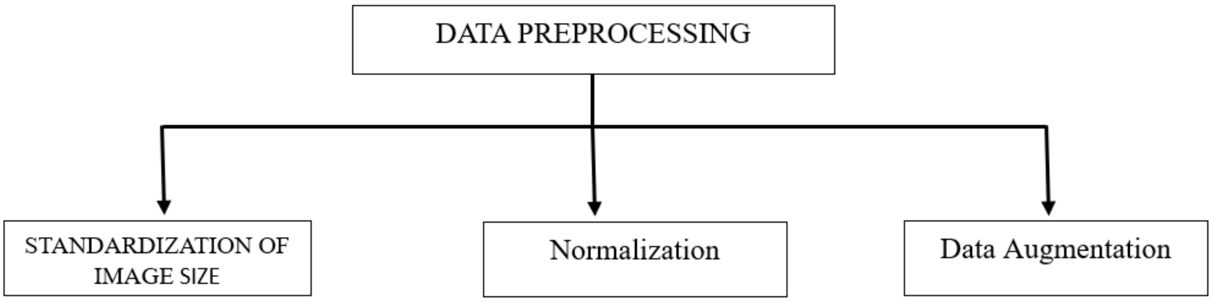

Fig 2 Data Preprocessing

a. **Standardization of Image Size:** All images were resized to a uniform size of 150 × 150 pixels to ensure consistency.
b. **Normalization:** The pixel values of all images were scaled to a range of 0 to 1 to improve model performance.
c. **Data Augmentation:** Techniques like rescaling, rotation, flipping, and random transformations were applied to increase the diversity of the training data.

2.3 Data Split

The dataset was divided into training and testing subsets as shown in Figure 3.

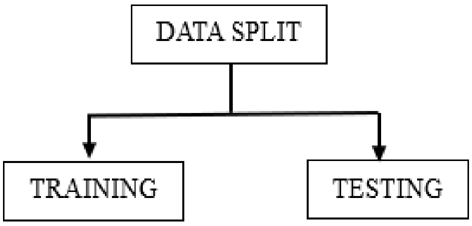

Fig 3 Data splitting

- **Training Set:** Contained 1373 high-quality CT scan images from all two categories which is used to train the model.
- **Testing Set:** Contained 413 high-quality CT scan images from all two categories for model evaluation.

2.4 Classification

Three different transfer learning-based classification techniques such as CNN, VGG16, and RESNET101as shown in Figure 4 were used to train the model for classifying CT scan images.

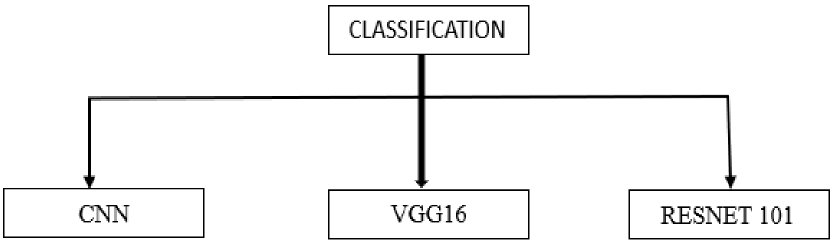

Fig 4 Model Used

a. **CNN (Convolutional Neural Network):** Used for feature extraction and learning spatial hierarchies in the data.
b. **VGG16:** A deep learning model with 16 weight layers, chosen for its ability to process high-resolution images effectively.
c. **ResNet101:** A residual network with 101 layers designed to handle deeper architectures by solving gradient issues using shortcut connections.

2.5 Accuracy Assessment

The models were evaluated based on training accuracy, and testing accuracy in terms of precision, recall, and F1-score as shown in Figure 5 for each class.

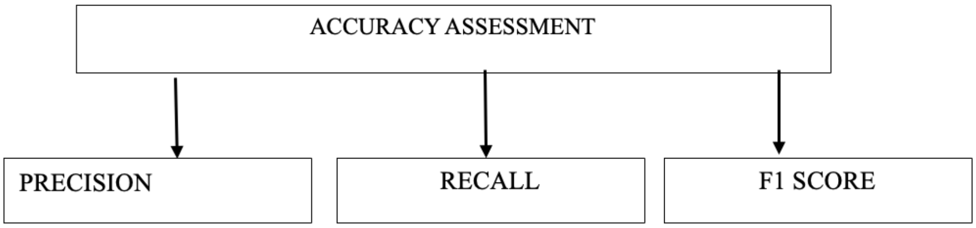

Fig 5 Accuracy Assesment

2.6 DISCUSSION

Lung cancer is one of the most common and fatal types of cancer, and early detection significantly improves patient outcomes. In this study, 1373 CT scan images were utilized to train and evaluate three models: CNN, VGG16, and ResNet101. **ResNet101** demonstrated superior performance with the highest testing accuracy and robustness in handling complex image data, the sample image is shown in Figure 6.

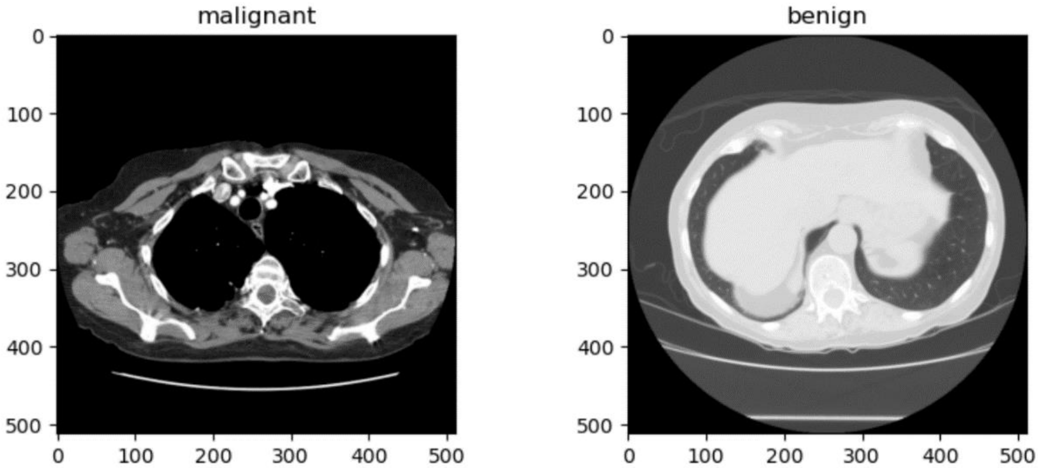

Fig 6 Sample Images

After the data pre-processing, the dataset was separated into 2 subsets, the first one used for training the model which contains 1323 images, and the second one used for testing which contains 413 images. Both training and testing subset contains all TWO types of images i.e. MALIGNANT, and BENIGN.

After the splitting of data, By using these high-quality magnetic resonance images, 3 models were trained efficiently to perform the accurate classification. The models trained are CNN, VGG 16, and RESNET101. All three are chosen for their good performance in handling complex image data. CNN is a foundation of deep learning model used for the model's efficiency for the feature extraction, which help in analyzing the complex pattern inside the MRI data. VGG 16, a deep network and transfer learning model containing 16 weight layers, picked for model's ability to gather spatial hierarchies and each detail of the images as it have sequential structure of convolutional layers. ResNet 101, a model of residual network.

The ResNet101 model, containing 101 layers, was used to address the gradient problem common in deep networks by building shortcut connections, which help it learn more complex patterns in a deeper structure. The results for the trained and tested models are as follows:

For **CNN**, the trained accuracy is 95.70%, and the tested accuracy is 90.79%, with an average precision, recall, and F1-score of 0.91 each. For **VGG16**, the trained accuracy is 94.78%, and the tested accuracy is 90.79%, with an average precision, recall, and F1-score of 0.91 each. For **ResNet101**, the trained accuracy is 94.67%, and the tested accuracy is 91.04%, with an average precision, recall, and F1-score of 0.91 each, shown in Table 1 .

Model	**Training Accuracy**	**Testing Accuracy**	**Average Precision**	**Average Recall**	**Average F1-Score**
CNN	95.70%	90.79%	0.91	0.91	0.91
VGG16	94.78%	90.79%	0.91	0.91	0.91
ResNet101	94.67%	91.04%	0.91	0.91	0.91

Table.1 Obtained Results

After training and testing the models, it was concluded that while all three models performed comparably in terms of precision, recall, and F1-score, **ResNet101** achieved a slightly higher testing accuracy, indicating its potential for handling complex images more effectively. However, **VGG16** and **CNN** also demonstrated robust performance, making them reliable alternatives depending on specific use cases.

3. CONCLUSION

Deep learning has shown great potential in transforming lung cancer detection and diagnosis, offering enhanced accuracy and efficiency compared to traditional methods. By leveraging large datasets from medical imaging, deep learning algorithms can identify cancerous nodules, differentiate between malignant and benign lesions, and predict disease progression with impressive precision. These advancements promise to improve early detection, enable personalized treatment, and ultimately increase survival rates for lung cancer patients. However, for deep learning solutions to be fully integrated into clinical practice, challenges such as model interpretability, data diversity, and regulatory compliance must be addressed. Continued research, robust validation, and collaboration between AI developers and healthcare professionals are essential to overcome these barriers and harness the full potential of deep learning in lung cancer detection. As AI-driven technologies advance, they can significantly enhance patient care, clinical decision-making, and overall healthcare outcomes.

REFERENCE:

Gupta, S., Vundavilli, H., Osorio, R. S. A., Itoh, M. N., Mohsen, A., Datta, A., Mizuguchi, K., & Tripathi, L. P. (2022). Integrative Network Modeling Highlights the Crucial Roles of Rho-GDI Signaling Pathway in the Progression of non-Small Cell Lung Cancer. *IEEE Journal of Biomedical and Health Informatics*, *26*(9), 4785–4793. https://doi.org/10.1109/JBHI.2022.3190038

Lima, D. V. C., Terrematte, P., Stransky, B., & Neto, A. D. D. (2024). An Integrated Data Analysis Using Bioinformatics and Random Forest to Predict Prognosis of Patients With Squamous Cell Lung Cancer. *IEEE Access*, *12*, 59335–59345. https://doi.org/10.1109/ACCESS.2024.3392277

Liu, P., Jin, K., Jiao, Y., He, M., & Fei, S. (2021). Prediction of Second Primary Lung Cancer Patient's Survivability Based on Improved Eigenvector Centrality-Based Feature Selection. *IEEE Access*, *9*, 55663–55672. https://doi.org/10.1109/ACCESS.2021.3063944

Naseer, I., Akram, S., Masood, T., Rashid, M., & Jaffar, A. (2023). Lung Cancer Classification Using Modified U-Net Based Lobe Segmentation and Nodule Detection. *IEEE Access*, *11*, 60279–60291. https://doi.org/10.1109/ACCESS.2023.3285821

Pang, S., Zhang, Y., Ding, M., Wang, X., & Xie, X. (2020). A Deep Model for Lung Cancer Type Identification by Densely Connected Convolutional Networks and Adaptive Boosting. *IEEE Access*, *8*, 4799–4805. https://doi.org/10.1109/ACCESS.2019.2962862

Silva, F., Pereira, T., Morgado, J., Frade, J., Mendes, J., Freitas, C., Negrao, E., De Lima, B. F., Silva, M. C. Da, Madureira, A. J., Ramos, I., Hespanhol, V., Costa, J. L., Cunha, A., & Oliveira, H. P. (2021). EGFR Assessment in Lung Cancer CT Images: Analysis of Local and Holistic Regions of Interest Using Deep Unsupervised Transfer Learning. *IEEE Access*, *9*, 58667–58676. https://doi.org/10.1109/ACCESS.2021.3070701

Srivastava, S., & Ahmed, T. (2021). Feature-Based Image Retrieval (FBIR) System for Satellite Image Quality Assessment Using Big Data Analytical Technique. In *PSYCHOLOGY AND EDUCATION* (Vol. 58, Issue 2). www.psychologyandeducation.net

Srivastava, S., & Ahmed, T. (2024). DLCD: Deep learning-based change detection approach to monitor deforestation. *Signal, Image and Video Processing*, *18*(Suppl 1), 167–181. https://doi.org/10.1007/s11760-024-03140-1

Ye, X., Zhang, W., & Sakurai, T. (2020). Adaptive Unsupervised Feature Learning for Gene Signature Identification in Non-Small-Cell Lung Cancer. *IEEE Access*, *8*, 154354–154362. https://doi.org/10.1109/ACCESS.2020.3018480

Yu, H., Zhou, Z., & Wang, Q. (2020). Deep Learning Assisted Predict of Lung Cancer on Computed Tomography Images Using the Adaptive Hierarchical Heuristic Mathematical Model. *IEEE Access*, *8*, 86400–86410. https://doi.org/10.1109/ACCESS.2020.2992645