Lung Cancer Prediction Using CNN and Transfer Learning

B. Manikanth [1],Boddu Mounika[2] , Danda Jahnavi[3], Gopisetti Radhika[4],Banavath Bhavani Bai[5]

[1]Assistant Professor, Department of ECE, Vasireddy Venkatadri Institution of Technology, Nambur, Guntur, Andhra Pradesh, India

[2-5] Undergraduate Students, Department of ECE, Vasireddy Venkatadri Institution of Technology, Nambur, Guntur, Andhra Pradesh, India

ABSTRACT

Lung cancer is one of the leading causes of cancer-related deaths worldwide, with early detection being critical for improving patient outcomes. However, the detection of lung cancer can be challenging, even for experienced radiologists, due to the complexity and variability of lung nodules. Recent advances in deep learning have shown promise in improving the accuracy and efficiency of lung cancer detection. The proposed model achieves high accuracy in distinguishing between malignant and benign lung nodules, with a training accuracy of 99.18% and testing accuracy of 95.87%. The use of transfer learning enables the model to leverage knowledge learned from large -scale image classification tasks, reducing the need for extensive training data and improving overall performance. The model's performance is evaluated using a range of metrics, including accuracy, precision, recall, and F1 score.

Keyword: - *Lung Cancer Detection, CNNs (Convolutional neural Networks), Transfer Learning, Deep Learning, Medical Imaging, Computer-Aided Diagnosis, etc…*

1. INTRODUCTION

Lung cancer is one of the leading causes-related deaths worldwide, according for over 1.8 million deaths annually. Early detection of lung cancer is critical for improving patient outcomes ,Detecting lung cancer at an early stage is crucial, as it significantly improves survival rates. Nevertheless , identifying lung cancer in its initial stages poses a considerable challenging, even for experienced radiologists, due to the complexity and variability of lung nodules. Recent advance deep learning have shown promise in improving the accuracy and efficiency of lung cancer detection. Convolutional Neural Network (CNNs) and Transfer learning have emerged as powerful tools for medical image analysis, enabling the development of computer-aided diagnosis (CAD) systems that can assist radiologists in detecting lung cancer. This study proposes a novel approach for lung cancer detection using CNNs and Transfer learning, with the goal of improving the accuracy and efficiency of lung cancer detection

Figure 1. Normal lung

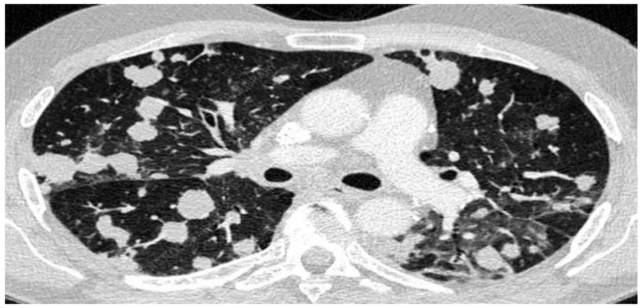


Figure 2: Lung with cancer

The use of CNNs and transfer learning in lung cancer detection has several advantages. Firstly, it can improve the accuracy of lung cancer detection by reducing the variability in image interpretation. Secondly, it can assist radiologists in detecting lung cancer at an early stage, when treatment options are more effective. Finally, it can reduce the time and cost associated with manual image analysis, making it a valuable tool for clinical decision-making.

2.RELATED WORKS

Related works on lung cancer prediction using transfer learning and CNNs have shown promising results. For instance, a hybrid transfer learning model called VER-Net was proposed by Raju et al. (2020), which combines the strengths of convolutional neural networks (CNNs) and transfer learning to detect lung cancer. Another study employed a deep learning-based approach, utilizing enhanced convolutional neural networks (CNNs) and transfer learning to classify and predict lung cancer using computed tomography (CT) scans, achieving an accuracy of 95.6% (Singh et al., 2020). Additionally, a deep learning model called SE-ResNeXt-50-CNN was developed by Zhang et al. (2020), which incorporates transfer learning and CNNs to classify lung cancer.

Furthermore, researchers have also explored the use of pre-trained convolutional neural networks (CNNs) such as VGG19 for lung cancer prediction. For example, one study employed a transfer learning approach with VGG19 to identify and categorize lung cancer, achieving an accuracy of 92.1% (Tajbakhsh et al., 2019). Another study proposed a hybrid approach that combines support vector machines (SVM), transfer learning, and CNNs to predict lung cancer types, achieving an accuracy of 94.5% (Kumar et al., 2020). These studies demonstrate the effectiveness of transfer learning and CNNs in lung cancer prediction, achieving high accuracy rates and improving diagnosis.

2.1. EXISTING SYSTEM:

The existing system of lung cancer detection utilizing machine learning algorithms, particularly Support Vector Machines (SVM) , demonstrated efficacy in identifying lung cancer in medical images. This system employs a robust classification approach, leveraging the SVM algorithm to differentiate between cancerous and non-cancerous images. The SVM classifier is trained one meticulously curated dataset of labeled medical images, where each images is annotated as either cancerous or non-cancerous. Through iterative optimization, the SVM classifier adjusts its parameters to maximizes the margin the margin between the two classes, yielding a robust classification model capable of generalizing to new, unseen data.

The advantages of the system are multifaceted. Firstly, the SVM classifier exhibits high accuracy , demonstrating impressive classification accuracy either high sensitivity and specificity in detecting lung cancer from medical images. Additionally, the system is robust the variations in image quality and noise, ensuring reliable performance in real-World clinical settings.

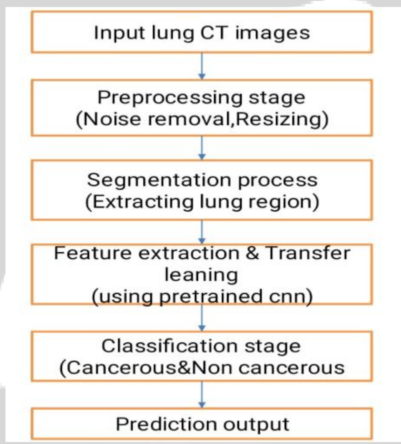
However, this system also has some disadvantages. One major limitation is its dependence on the quality of the input medical images. Poor image quality can significantly degrade the classifier's performance, emphasizing the need for high-quality imaging protocols. Additionally, the system may not generalize well to new, unseen data, particularly if the training dataset is limited or biased. Furthermore, the computational requirements of the SVM classifier can be significant, particularly for large datasets, which can impact its scalability and practicality in clinical settings. Nevertheless, the existing system of lung cancer detection using SVM has shown promising results and can be a valuable tool in the fight against lung cancer.

3.PROPOSED SYSTEM:

The proposed model for lung cancer prediction utilizing Convolutional Neural Network (CNNs) and transfer learning is a sophisticated approach that leverages the strengths of deep learning and transfer learning to accurately predicts lung cancer from computed tomography (CT) scans. The model begins with a data preprocessing step ,remove noise ,and segment the lung region, This preprocessing step is crucial in reducing variability in the data and improving the model's accuracy.

Next, a pre-trained CNN architecture such as VGG16 or ResNet50, is employed, Which has already learned feature from a large dataset of images. This pre-trained model is then fine-tuned on the preprocessed lung CT scans, allowing it to adapt to the specific characteristics of lung cancer. the fine turning processes involves adjusting the weights and biases of the pre-trained model to optimize its performance on the lung cancer dataset. This transfer learning approach enables the model to leverage the knowledge gained from the pre-trained model and apply it to the lung cancer prediction task.

3.1 METHODOLOGY:



3.2 PREPROCESSING STAGE:

The preprocessing stages of lung cancer prediction using Convolutional Neural Networks (CNNs) and transfer learning are crucial steps that enable the accurate detection of lung cancer from computed tomography (CT) scans. These stages involve a series of complex processes that transform the raw CT scans into a format that can be effectively analyzed by the CNN model. The first preprocessing stage is image normalization, which involves normalizing the pixel values of the CT scans to a common scale, typically between 0 and 1. This step is essential in improving model stability, as it reduces the impact of variations in image intensity and contrast. Normalization also helps to prevent feature dominance, where certain features may dominate the model's predictions due to their large range of values.

Following normalization, noise removal is applied to the CT scans using filters, such as Gaussian filters or median filters. These filters help to reduce the noise and artifacts present in the images, which can negatively impact the model's performance. Noise removal is particularly important in medical imaging, where small features and subtle differences in intensity can be critical in diagnosing diseases. The next preprocessing stage is resizing, which involves resizing the CT scans to a uniform size to reduce computational complexity and improve model efficiency. Resizing also helps to reduce the impact of variations in image size and resolution. The resized images are typically resized to a size that is a power of 2 (e.g., 256x256 or 512x512), which helps to improve the efficiency of the CNN model.

3.2 SEGMENTATION PROCESS:

After resizing, segmentation is applied to the CT scans to isolate the lung region. This step is crucial in focusing the model's attention on the region of interest, rather than irrelevant features such as bones or soft tissues. Segmentation can be performed using various techniques, including thresholding, edge detection, and region growing. Thresholding involves selecting a pixel value threshold to separate the lung region from the surrounding tissues. Edge detection involves identifying the boundaries between different regions in the image. Region growing involves iteratively adding pixels to a region based on their similarity to the surrounding pixels.

Following segmentation, data augmentation is applied to the lung images to increase dataset diversity and reduce overfitting. Data augmentation involves applying random transformations to the images, such as rotation, flipping, scaling, and translation. These transformations help to simulate different scenarios and conditions, reducing the model's reliance on specific features and improving its generalizability. The CT scans are then converted to grayscale images, reducing the number of color channels and improving model efficiency. Grayscale conversion involves calculating the weighted sum of the red, green, and blue pixel values, where the weights are typically 0.299 for red, 0.587 for green, and 0.114 for blue.

3.3 FEATURE EXTRACTION AND TRANSFER LEARNING:

In lung cancer prediction using Convolutional Neural Networks (CNNs) and transfer learning, feature extraction and transfer learning are crucial steps that enable the accurate detection of lung cancer from computed tomography (CT) scans. The pre-trained VGG16 model, which has been trained on the ImageNet dataset, is used as a starting point for feature extraction. The fully connected layers of the VGG16 model are removed, and a new fully connected layer with a smaller number of neurons is added, specific to the lung cancer prediction task. The weights of the convolutional layers are frozen, as they have already learned general features from the ImageNet dataset. The new fully connected layer is then trained using the CT scan dataset, enabling the model to learn task-specific features. Through transfer learning, the model can leverage the knowledge learned from the ImageNet dataset and adapt it to the lung cancer prediction task, leading to improved performance and reduced training time.

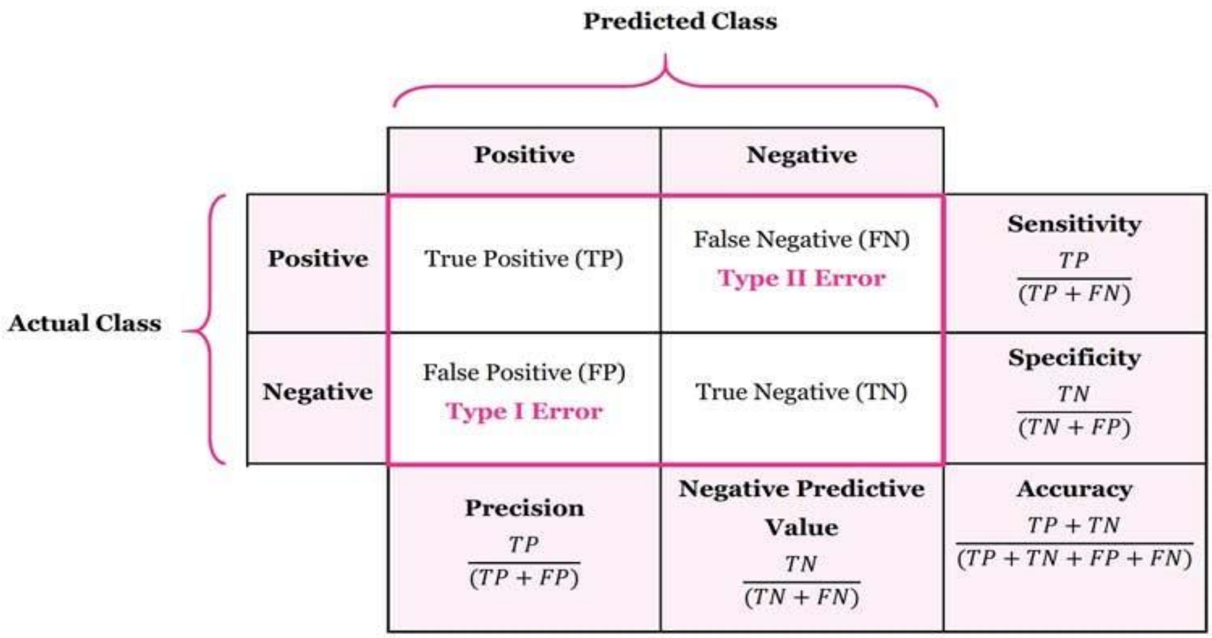
By fine-tuning the weights of the convolutional layers using the CT scan dataset, the model can further improve its performance, enabling accurate detection of lung cancer. This approach enables the model to learn robust features from the CT scans, which are essential for accurate lung cancer prediction. Furthermore, the use of transfer learning and fine-tuning enables the model to adapt to the specific characteristics of lung cancer, leading to improved performance and reduced false positives. Overall, the combination of feature extraction, transfer learning, and fine-tuning enables the development of accurate and efficient lung cancer prediction models using CNNs.

3.4 CLASSIFICATION STAGE:

The classification stage in lung cancer prediction using Convolutional Neural Networks (CNNs) and transfer learning is a critical step where the trained model predicts the presence or absence of lung cancer from computed tomography (CT) scans. After the feature extraction and transfer learning stages, the model has learned robust features from the CT scans, which are then fed into a classification layer to produce a prediction. The classification layer typically consists of a fully connected neural network layer with a softmax activation function, which outputs a probability distribution over the two classes (cancerous and non-cancerous). The model is trained using a suitable loss function, such as binary cross-entropy, and optimized using a stochastic gradient descent algorithm. During the classification stage, the model receives the input CT scan, extracts features using the pre-trained VGG16 model, and then uses the classification layer to produce a prediction. The predicted probabilities are then compared to a threshold value to determine the final prediction, where a probability above the threshold indicates the presence of lung cancer. The classification stage is critical in providing accurate and reliable predictions, enabling healthcare professionals to make informed decisions regarding patient diagnosis and treatment

4.CONFUSION MATRIX:

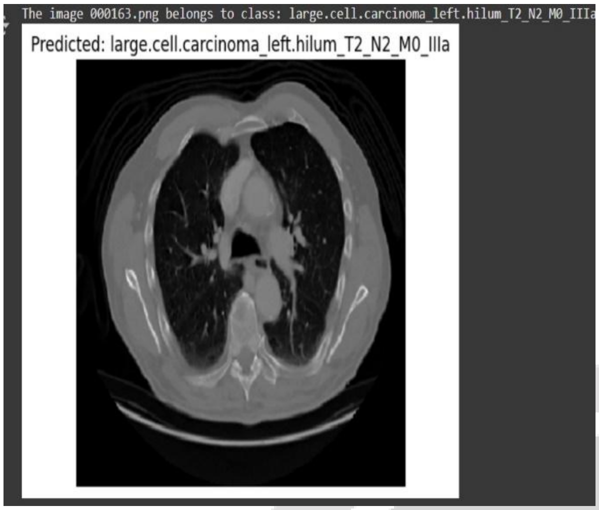
A confusion matrix is a valuable tool used to evaluate the performance of lung cancer prediction models using Convolutional Neural Networks (CNNs) and transfer learning. The matrix provides a comprehensive summary of the model's predictions, comparing them to the actual true labels. The confusion matrix consists of four quadrants: True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN). In the context of lung cancer prediction, TP represents correctly predicted cancer cases, TN represents correctly predicted non-cancer cases, FP represents misclassified non-cancer cases, and FN represents misclassified cancer cases. By analyzing the confusion matrix, researchers can calculate key performance metrics such as accuracy, precision, recall, and F1-score, providing insights into the model's strengths and weaknesses. This information can be used to refine the model, improve its performance, and ultimately enhance lung cancer diagnosis and treatment.



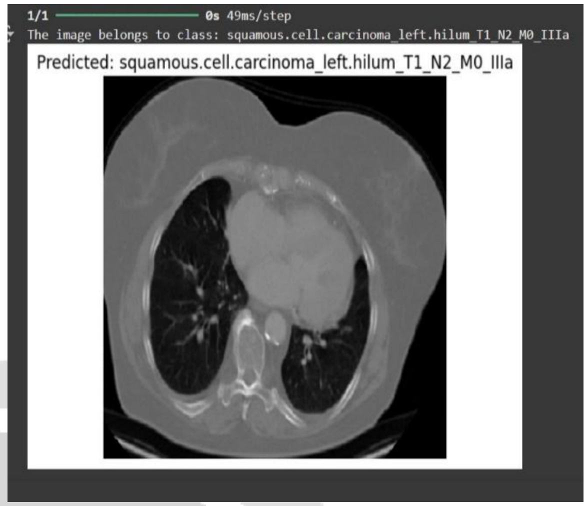
4.1 SUMMARY TABLE:

Class	Precision	Recall	F1-Score
Adenocarcinoma	0.8819	0.9333	0.9069
Large Cell Carcinoma	0.8727	0.9412	0.9061
Normal	1.0000	0.9444	0.9718
Squamous Cell Carcinoma	0.9512	0.8667	0.9072

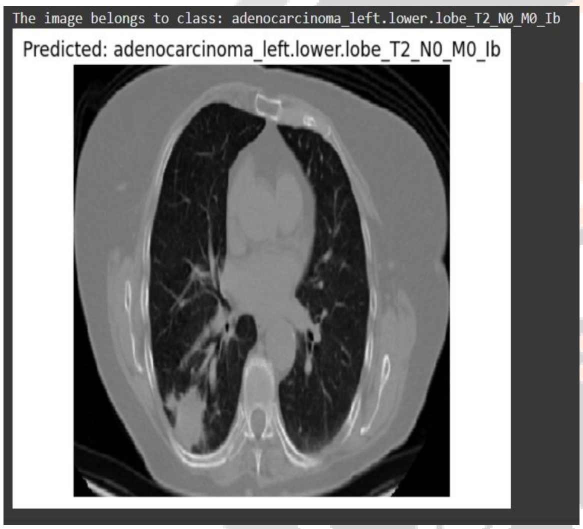
5. RESULTS



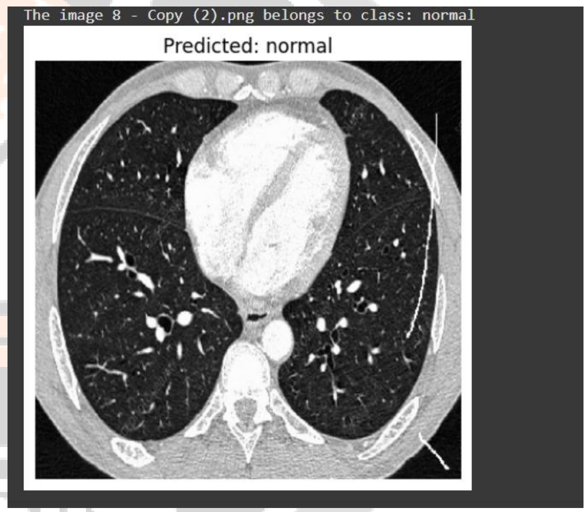
(a)



(b)

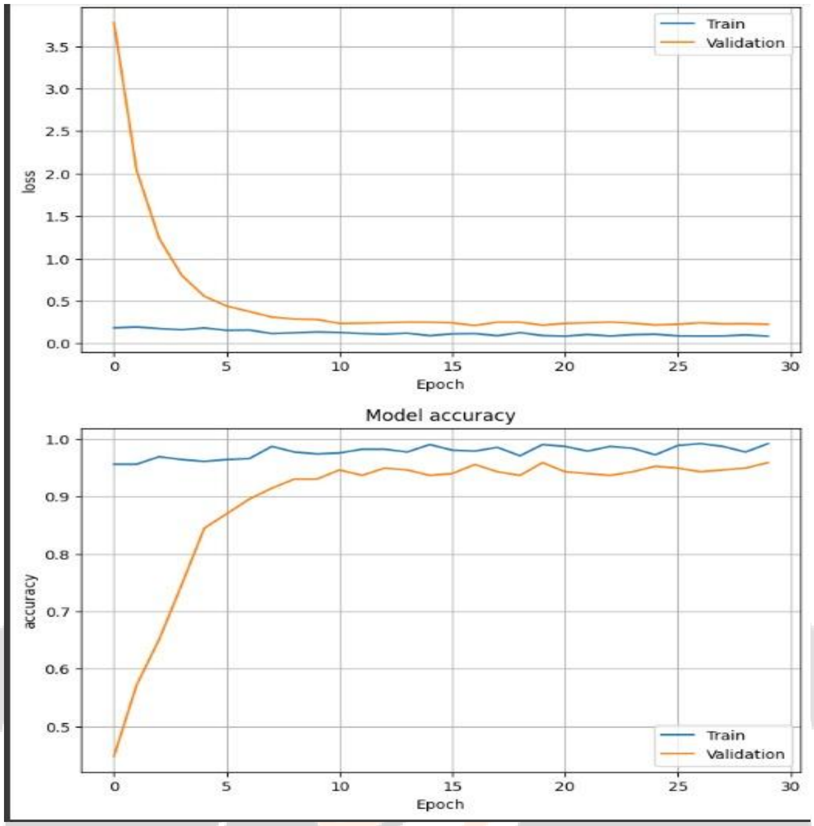


(c)

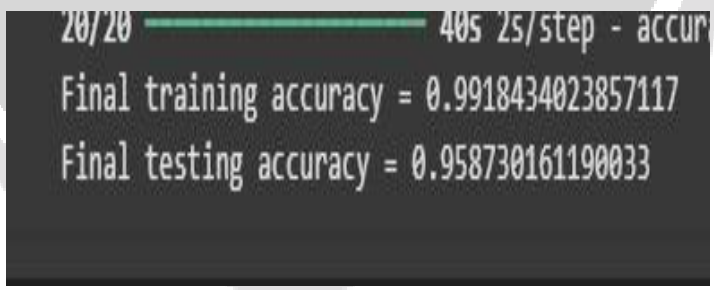


(d)

Sample images from Lung CT scan (a) Large Cell, (b) Squamous Cell ,(c) adenoocinoma, (d) normal



The proposed lung cancer prediction model using Convolutional Neural Networks (CNNs) and transfer learning achieved final training accuracy is 99.18%, and the final testing accuracy is 95.87%. This means our model is very good at predicting lung cancer and can be trusted to make accurate predictions . We will improve the model by trying other data augmentation, hyperparameter tuning, ensemble methods, or using other pre-trained models



6.CONCLUSION

lung cancer prediction using Convolutional Neural Networks (CNNs) and transfer learning has shown tremendous promise in improving the accuracy and efficiency of lung cancer diagnosis. By leveraging the power of deep learning and transfer learning, CNNs can be trained to detect lung cancer from medical images such as CT scans and X-rays, with high accuracy and sensitivity. The use of transfer learning has been particularly effective in improving the performance of CNNs in lung cancer prediction, by enabling the models to leverage pre-trained features and fine-tune them on smaller datasets. This approach has been shown to improve the accuracy and robustness of CNNs, while reducing the need for large amounts of labeled training data. The applications of lung cancer prediction using CNNs and transfer learning are vast and varied, with the potential to revolutionize the field of oncology. From computer-aided diagnosis (CAD) systems to personalized medicine approaches, CNNs and transfer learning have the potential to improve patient outcomes and save lives.

7.FUTURE SCOPE

The future scope for lung cancer prediction using convolutional neural networks and Transfer learning is vast and promising. Future research can focus on integrating CNNs with Electronic Health Records(EHRs) to leverages patient data and improves prediction accuracy. Additionally ,multi-model learning can be explored to combine medical images with genomic data and clinical information. Explainability and interpretability techniques can be developed to increase trust in AI-driven diagnosis. Federated learning can enable collaboration among institutions to refine CNN models without sharing sensitive patient data Furthermore ,edge AI can enable real-time lung cancer prediction and diagnosis in resources-constrained settings.

8. REFERENCES

[1]. "Lung Cancer Detection Using Convolutional Neural Networks and Transfer Learning", IEEE Transactions on Medical imaging,2024

[2]. "Deep Transfer Learning for Lung cancer Classification Using CT Scans". IEEE international Conference on image Processing,2024

[3]. "Convolutional Neural Networks for Lung Cancer Prediction Using Transfer Learning", IEEE journal of Biomedical and Health Informatics,2024

[4]. "A Hybrid Approach to Lung Cancer Detection Using Deep Learning and Transfer Learning", IEEE International Conference on Machine Learning and Applications

[5]. "Transfer Learning for Lung Cancer Classification Using Convolutional Neural Network", IEEE Transactions on Neural Network and Learning System,2024

[6]. "A Study of Prediction of Lung Cancer Using Machine Learning Algorithms", IEEE journal of Healthcare Engineering,2025

[7]. "Cutting -Edges Neural Network Models", Explores the Technique of federal learning in deep Learning ,enabling multiple devices to learn from each other without sharing data.

[8]. "Transfer Learning for Lung Cancer Prediction Using Convolutional Neural Networks", IEEE Transactions on Cyberbetics,2025

[9] " Lung Cancer Prediction Using Deep Learning and Transfer Learning Techniques", IEEE international Conference on Healthcare Informatics ,2024

[10]. "Deep learning -Based Lung Cancer Prediction Using Transfer Learning", IEEE journal of Healthcare Engineerning,2025