Journal of Neonatal Surgery
ISSN(Online): 2226-0439
Vol. 14, Issue 26s (2025)
https://www.jneonatalsurg.com

OPEN ACCESS

Hierarchical Multiclass Classification Of Non-Small Cell Lung Cancer Leveraging Efficientnet-B3 For Enhanced Diagostic Precision

Swathi Bonthala[1*], Suhasini.A[2], Suvarchala.K[3]

[1]Ph.D Research Scholar,Department of CSE, Annamalai University, Chidambaram, Annamalai University, Chidambaram, [3]Associate Professor,Department of CSE, Institute of Aeronautical Engineering and Technology, Hyderabad.

[2]Professor,Department of CSE, Annamalai University, Chidambaram, [3]Associate Professor,Department of CSE, Institute of Aeronautical Engineering and Technology, Hyderabad.

Email ID: swathi.redy03@gmail.com

***Corresponding Author:**

Email ID: suha_babu@yahoo.com Email ID: k.suvarchala@iare.ac.in

Cite this paper as: Swathi Bonthala, Suhasini.A, Suvarchala.K, (2025) Hierarchical Multiclass Classification Of Non-Small Cell Lung Cancer Leveraging Efficientnet-B3 For Enhanced Diagostic Precision. *Journal of Neonatal Surgery*, 14 (26s), 633-644.

ABSTRACT

Among all forms of cancer, lung cancer has the highest global prevalence and continues to be the leading cause of cancer-related mortality. Lung cancer is categorized into two distinct forms: non-small cell lung cancer (NSCLC), which constitutes approximately 85% of cases, and small cell lung cancer (SCLC), a less common but more aggressive subtype. The primary subtypes of NSCLC include large cell carcinoma, squamous carcinoma, and adenocarcinoma which differ in histopathological features, molecular markers, and treatment responses. CT scan imaging plays a crucial role in accurately diagnosing lung cancer, facilitating the generation of high-quality diagnostic images for further analysis. Deep learning-based convolutional neural networks (CNNs) have proven highly effective in detecting and classifying lung cancer, leveraging hierarchical feature extraction from CT scan images. While CNNs demonstrate strong performance in lung cancer detection, they suffer from high computational costs and prolonged training times. Transfer learning addresses these limitations by leveraging pre-trained models, making it particularly effective for medical imaging with scare-labelled data. In this paper, we investigate the use of EfficientNet-B3 to enhance the accuracy of lung cancer detection. The model's performance is evaluated using various metrics, including accuracy, precision, recall, F1-score, and ROC curve. Comparative analysis shows that ResNet achieves 91% accuracy, while VGG-16 attains 94%. Further improvements are observed with EfficientNet-B3, which achieves an accuracy of 98% making it the most accurate model among the assessed.

Keywords: *NSCLC classification, Detection, EfficientNet-B3, Transfer Learning, ImageNet.*

1. INTRODUCTION

Lung cancer remains the leading cause of cancer-related mortality worldwide, exceeding breast cancer in fatality rates due to late-stage diagnosis and limited therapeutic success in advanced cases. [1]. According to 2025, the estimated statistical report for lung cancer deaths is around 124,730. From various cancers, lung cancer is categorized into two primary histological subtypes: small cell lung cancer (SCLC) and non-small cell lung cancer (NSCLC). Non-small cell lung cancer (NSCLC) represents 85% of lung cancer cases, while small cell lung cancer (SCLC) accounts for the remaining 15%. NSCLC is further classified into adenocarcinoma, squamous cell carcinoma, and large cell carcinoma, with adenocarcinoma being the most prevalent, comprising approximately 40% of cases. The complexity of these subtypes necessitates robust computational models, such as deep learning, for accurate classification and early diagnosis. ADC primarily develops in the outer region of the lung and is more commonly diagnosed in women and non-smokers. SCC, the second most prevalent NSCLC subtypes, accounts for approximately 25-330% of cases. It is strongly associated with heavy smoking and rapid tumor progression, typically originating in the central lung region and bronchial tubes, where it arises from squamous epithelial cells constitutes 10-15% of NSCLC cases and can occur in any lung region but is predominantly found in the peripheral lung zones, with a higher prevalence among smokers [2].

Many research studies have revealed and researchers suggested that many of the patients are diagnosed with NSCLS as early diagnosis is required to reduce the death rate. Traditional diagnostic methods, including biopsy, sputum cytology, and bronchoscopy don't work accurately in terms of time-consuming to detect lung cancer. Later, Image-based methods like

Chest X-Ray, PET-CT, MRI, and CT scan have emerged as more effective alternative for detecting lung cancer at earlier stages. Among of all CT scan imaging provides accurate results for the detection of lung cancer [3]. CT imaging plays a fundamental role and is mostly used in medical imaging to detect lung disease. It generates various image techniques for lung detection such as conventional CT scans detecting small lung nodules and large tumors, High-resolution CT is used for detecting the ideal small nodules and thin slices ranging from 0.5-1mm and Low-dose CT scan is used for screening the lung, especially in smokers.4D CT scan captures the dynamic imaging in breathing cycles to examine lung abnormalities which are easy for doctors [4].

Medical imaging research has seen significant advancements, particularly in the use of CT and PET-CT scans for diagnosing lung cancer with deep learning and machine learning techniques. Traditional machine learning approaches, including supervised and unsupervised methods like SVM stands for support vector machine, DT, RF have been applied for diagnosis [5]. However, these models face challenges such as computational complexity, accuracy constraints, annotation difficulties, and lack of interpretability. Deep learning leveraging neural networks such as Convolutional neural networks (CNN), Feedforward neural network (FNN), and deep neural networks (DNN), has demonstrated superior performance in lung cancer detection by automatically extracting and learning hierarchical feature from medical images [6].

Convolutional Neural Networks (CNNs) are designed to refine the raw medical imaging data automatically, extracting the textures, shape, edges, and complex patterns for identifying the abnormalities in the lung such as tumors and lung nodules in lung CT images. Many researchers have worked and conducted experimental results on the multiclassification of NSCLC by applying CNN architecture/models. The existing method worked on the detection of lung abnormalities. It focused on discriminating benign and malignant cells in the lung with an accuracy rate of around 87% to 95%. Even though it faces a few challenges like computational cost and complexity, minimum accuracy, and generalization. To address the mentioned issues, an innovative Fine-tuned was developed in this paper for classifying NSCLC.

The primary contributions of this paper are described below:

i) The development of a hybrid classification method combining Fully connected (FC) layer with a SoftMax activation to enhance the multi-class classification of NSCLC subtypes.

ii) The proposed work in this study is EfficientNet-B3 to enhance the hierarchical feature representation for improve classification accuracy.

iii) Adam optimizer is used for its adaptive learning rate mechanism, requiring minimal hyperparameter adjustments while effective handling noisy and sparse data.

2. RELATED WORK

Medical imaging is a key application of artificial intelligence (AI) that has significantly transformed lung cancer analysis. By leveraging advanced AI techniques, including machine learning and deep learning, CT scan images can be effectively processed for accurate detection and classification of lung cancer. These technologies enhance diagnostic precision, automate feature extraction, and improve early detection, ultimately contributing to better patient outcomes.

Xueyun Tan et al. [2] developed an innovative multi-modality model for predicting tumour invasion status using deep learning algorithms. This multi-modality approach integrates radiomics, pathology testing, and clinical indicators, forming a multi-RPC (Radiomics-Pathology-Clinical) model to enhance diagnostic accuracy. The study involved 1,424 patients from two different hospitals, where data was segmented into multimodal training, pathology testing, and radiology testing groups. The combination of radiomics, pathology, and clinical indicators within the multi-RPC model achieved an Area Under the Curve (AUC) of 0.95, with a confidence interval (CI) of 0.89–0.939 for training and 0.939 (95% CI: 0.878–0.974) for validation. By integrating the multi-RPC model, the study demonstrated the potential to enhance predictive accuracy, thereby assisting clinicians in reducing unnecessary surgical procedures and optimizing patient-specific treatment decisions.

Yu et al, [3] examined the histological stages of tumors in the lung is highly recommended for doctors to diagnose and treat, especially for those suffering from NSCLC. To classify and predict the pathological stage of NSCLC by using random forest algorithm in supervised machine learning. while Huan Yang et al [4] developed six types of classifiers for pathological analysis of whole slice images (WSI) as adenocarcinoma, SCC, LCC, pneumonia, pulmonary tuberculosis, and normal lung by using deep learning like ResNet-50 and EfficientNet-B5. Collected patient data from four different hospitals to examine the accuracy among six classifiers and 1067 slides were tested and resulted in AUC of 0.970, 0.963, 0.918, and 0.978. At the same time Masud et al, [5] employed a CNN to train the CT images lung cancer categorization and nodule with 96.33% accuracy.

Wu, J., Chen, Z; et al,[6] compared the functioning of VGG-16 CNN with traditional ML techniques like SVM, Decision Tree, and RF, etc for the detection and classifying of NSCLC with PET/CT images.VGG-16 automatically extracts the convoluted feature from medical imaging without manual effort, unlike ML which works on human involvement. while Bebas et al,[7] performed an investigation on the classification of cellular subtypes of NSCLC using MRI/PET for analysis of texture images with ML classifiers particularly SVM with an accuracy of 75.48% which was better than the other

classifiers in ML. Ahmad et al,[8] show how machine learning models can predict the early symptoms of lung cancer. His research on prevention and early screening reduces risk factors to get better outcomes for patients.

Guo et al,[9] presented deep learning networks method to originating in minimal-dose imaging in lung spots and identify them as either it is benign or malignant through a precision of 99.02% with high accuracy, predicting the threat of metastatic pulmonary lesions in pleural effusion enables early-stage diagnosis of lung cancerous [10]. Pulmonary disease has a minimal lifespan of five years if the patient is affected by cancer. So, early analysis is more essential than treatment for the patients. To reduce this, the authors implemented a model called CPHM which works as a foundation for survival analysis and makes it easy for doctors to diagnose and regulate treatment for patients [11,12,13]. Huang et al[14], proposed XGBoost(Extreme Gradient Boosting) is a supervised algorithm used for classification and regression. As it predicts the one-year survival stage of cancer spreads in the bones. While Lu et al [15], implemented a U-Net model for segmentation of CT imaging, U-Net predicts an early diagnosis of NSCLC known survival rate of five years by integrating U-Net and CPHM (statistical model) to improve the prediction accuracy. Lai et al [16], combined Gene biomarker expression with clinical data to predict the five-year survival stage of NSCLC medical cases with an accuracy of AUC 0.8164 and diagnostic accuracy is 75.44%. In contrast, she et al[17], proposed DFNN with CPHM for post-diagnosis survival prediction of 3-year and 5-year in NSCLC.

PET is used to identify high-quality images such as tumor shape, size and location of a cancerous region [18]. The radiomic features show that qualitative features on PET/CT imaging have robust predictive capability of NSCLC [19]. Despite advancements, postoperative predicting NSCLC survives complex. Lee et al [20] designed a prediction method adopting a MLP with unsupervised and supervised learning to predict the three-year probability of tumor relapse after NSCLC resection that proved to be more efficient than the CPHM. Many medical assistant systems have incorporated ensemble learning to streamline pharmaceutical evaluations, Luo et al [21] applied ensemble methods to optimize a multi-agent drug therapy suggestion for NSCLC without relying on gene sequence data. The recommendation system resulted, in doctors having the ability to choose the suitable drug for individual patients. This ensemble learning makes the quick decision for the prostate gland that can help doctors to diagnose [22].

Chun-Hui Lin et al. [23] developed a deep learning (DL) framework for the early identification of lung cancer, aiming to enhance automated detection and reduce both the risk and prevalence of the disease. By leveraging advanced DL architectures, the proposed model improves diagnostic accuracy and facilitates timely intervention, thereby contributing to more effective lung cancer screening and management. This work created a generative adversarial network-based classifier for medical image analysis (CT images) and lung tumor identification. This computer-aided mechanism provides efficient training and mitigates the issue of spare data. The experimental outcomes demonstrated that it achieved 94.86% accuracy. However, this methodology is unstable and provides less convergence than others.

3. BACKGROUND WORK

3.1 Dataset

The LC15000 dataset, featuring images of lung cancer histopathology, has been compiled. This collection consists of 15,000 images depicting three types of lung tissues: SCC, LCC, and Adenocarcinoma. The images were sourced from the Kaggle repository. To expand dataset to 15,000 images, with 5000 images per type, the original photos were subjected to image augmentation techniques, including rotation and flipping. Before these techniques were applied, the images were resized to a square format of 768 by 768 pixels. Initially, the original images had dimensions of 1024 x 768 pixels.

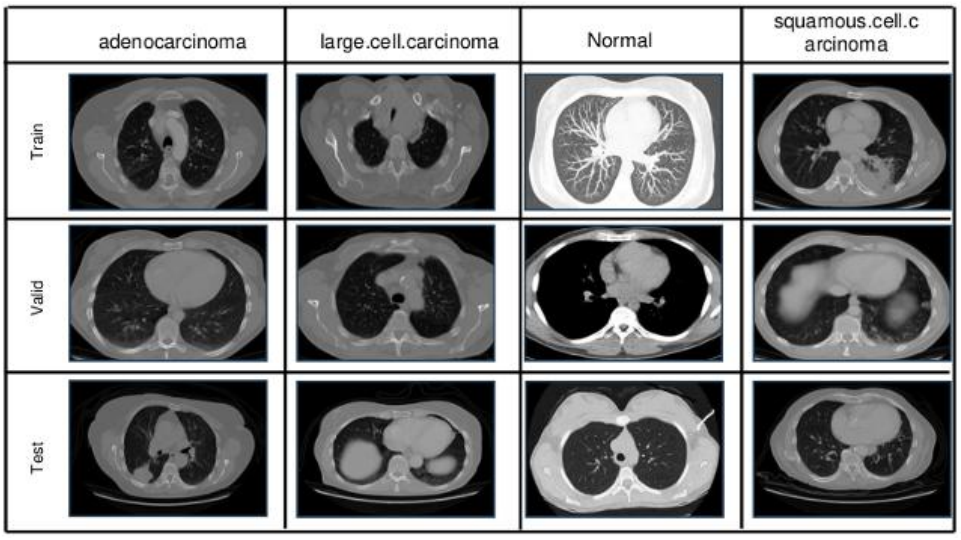


Fig-1: sample CT images of lung classification

3.2 Data Augmentation

To overcome the limitations of data availability, we applied augmentation techniques to enlarge our training dataset artificially. Data augmentation encompasses the generation of additional training samples through the application of randomized transformations to existing images, thereby improving the model's capacity for generalization to novel data. In this paper, we employed a range of data augmentation techniques: images were randomly zoomed to a size ranging from 85% to 115% of the original size (Zoom range), randomly shifted horizontally by as much as 20% in either direction (width shift range), moved vertically within a -20% to 20% range (height shift range), and randomly sheared between -15% to 15% (shear range). By implementing these techniques, our primary objective was to enhance the robustness and adaptability of the proposed deep learning model, thereby optimizing its performance for the accurate identification and classification of lung cancer. This approach aims to improve the model's generalization ability across diverse datasets, ensuring reliable and efficient diagnostic capabilities in clinical applications.

3.3 Transfer Learning

Transfer Learning referred to domain generalization acts as theoretical approach that utilizes prior insights across domains to enhance performance on transferable tasks, such as tagging or classification. CNN automatically learns and extracts features from the initial layer and deeper layers directly from its layers, distinguishing it from conventional machine learning methods that rely on manual feature engineering. The extracted features are converted into a 1-D feature vector, which is then passed through one or more fully connected layers to carry out classification tasks. While achieving notable success, a fundamental challenge is its dependence on a comprehensive dataset to enhance training efficiency and reduce overfitting. However, collecting volumized annotated data in medical imaging is a challenging task, and most of the existing data remains inaccessible to the public. Transfer Learning addresses this challenge by utilizing prior knowledge from pre-trained on fine-turned comprehensive benchmark datasets like ImageNet, to improve performance on tasks with limited data availability such as lung cancer classification from lung CT images.

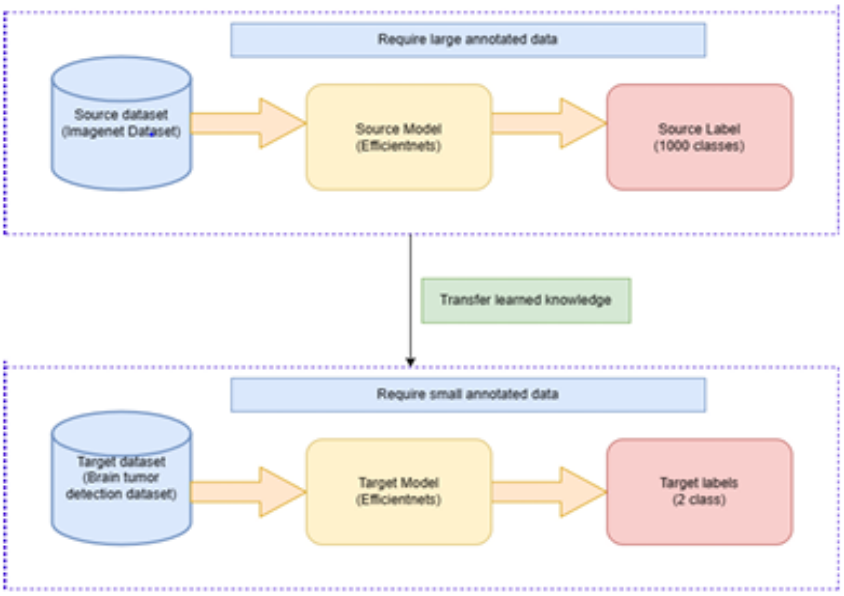


Fig-2: General structure of transfer learning

The figures demonstrate Transfer Learning, emphasizing the distinction between the target and source dataset, such as a collection of CT images. Consequently, no pre-trained CNN models are suitable for data interpretation and are assumed to effectively transfer learned knowledge to new test data. Instead, multiple layers of pre-trained models are fine-tuned based on observations to enhance compatibility with the target domain dataset.

3.4 Proposed EfficientNet-B3 Model

EfiicientNet-B3 is one of the types of Efficient Net family and has achieved notable proficiency in high-volume image recognition tasks particularly on the ImageNet dataset, ensuring exceptional accuracy and reliability. Efficient Net-B3 represents DCNN architecture denoting 8 times more efficient in inference and almost 6 times faster than well-known models such as ResNets, VGGNets, GoogleNets, InceptionNet, and Xception. Efficient Net implements compound scaling to optimize and enhance model versatility across the CNN family. The depth of a network indicates the layer count, while the width of a convolutional layer is defined by the number of filters used. The Equation represents:

Depth: $D = \emptyset^{\emptyset}$ (1)

Width: $W = \lambda^{\theta}$ (2)

Resolution: $R = \mu^{\theta}$ (3)

S.t.$\Theta.\lambda^{\Theta}.\mu^2 \approx 2$ (4)

$\emptyset \geq 1, \gamma \geq 1, \mu \geq 1$ (5)

Throughout the model development, employed the Kera's library to deploy an algorithm efficient net-B3 operating as the foundational architecture, initializing it with weights pre-trained on ImageNet. To maximize the utility of the predefined features, we locked the first half of the layers and then integrated additional custom layers on top of the base architecture. To reduce the spatial dimensions, we first employed a global average pooling layer with two fully connected layers afterwards comprising 64 and 128 neurons with ReLu activations. Additionally, dropout layers set at 0.5 were inserted at key points to enforce regularization and prevent overfitting.

For our four-class multiclass classification task, we configured the final dense layer of the model with a SoftMax activation function, enabling the network to generate probability distributions across the four classes and facilitating accurate classification. The model's unique architecture excels at capturing complex patterns, particularly in high-resolution images, while the compound scaling method secures a balanced trade-off between performance and computational cost.

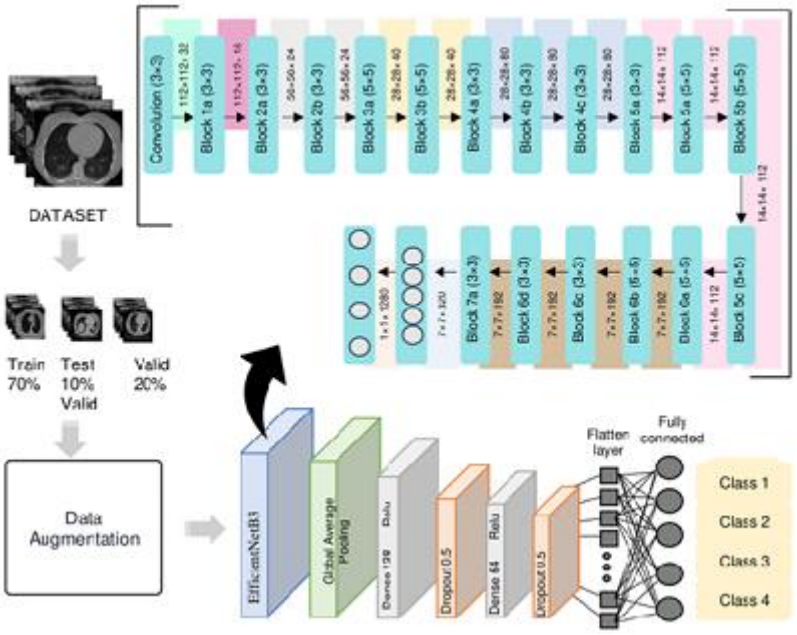


Fig-3 Proposed architecture of Multiclass classification of Lung CT dataset

3.5 Adam Optimizer Algorithm

The Adam (Adaptive Moment Estimation) optimizer efficiently updates network weights by combining momentum-based acceleration with adaptive learning rates. This optimization technique enhances convergence speed and stability, making it well-suited for training deep learning models.

The following are the steps of Adam

Algorithm:

Step-1: Initialize parameters and Hyperparameters

Given:

Θ (parameters to be optimized)

f(Θ) (loss function)

$g_t = \nabla f_t(\emptyset)$ (gradient of loss function at time step t)

Hyperparameters:

α (learning rate)

β_1(controls the decay coefficient for the first momentum estimate, which tracks moving average of past gradients, it is

typically 0.9)

β_2(Regulates the exponential decay rate for the second momentum estimate, capturing the moving average of squared gradients. Its default value is 0.999)

ε (A small constant, often 10^{-8} added to the denominator to avoid division stability during training)

Initialize:

$m_0 = 0$ (first momentum vector)

$v_0 = 0$ (second momentum vector)

t = 0 (time step)

Step-2: Iterative update of parameters:

i) Increment time step: t← $t + 1$

ii) Compute the partial derivative of the loss function with respect to the model parameters: $g_t = \nabla f_t(\emptyset)$

iii) Update Initial momentum approximation: $m_t = \beta_1 . m_{t-1} + (1 - \beta_1) . g_t$

iv) Update pre-corrected second-order momentum estimate: $v_t = \beta_2 . v_{t-1} + (1 - \beta_2) . g_t^2$

v) Calculate the unbiased first momentum estimate: $\hat{m}_t = m_t / 1 - \beta_1^t$

vi) Compute adjusted second moment estimate: $v\hat{t} = v_t / 1 - \beta_2^t$

vii) Refine model coefficients: $\emptyset = \emptyset - \propto . m\hat{t} / \sqrt{\hat{v}_t} + \in$

Step-3: Convergence condition:

Repeat step-2 until convergence (e.g., loss stops decreasing significantly).

4. EXPERIMENTAL DISCUSSION

within this segment will discusses the outcomes of the proposed model. The proposed model EfficientNet-B3 to generate the best accurate results.

4.1 Experimental Configuration

The designed algorithm of EfficientNet-B3 is executed using the python programming using operating system is windows 10 an intel CPU and 16 GB Ram.

4.2 Performance Metrics

The performance metrics employed to measure the performance of the developed classifier for lung cancer categorization are formally defined below, along with their corresponding mathematical expressions.

i) Accuracy:

Accuracy defines the classifier's effectiveness in categorizing both healthy and lung cancer classes and it quantifies the proportion of truly classified cases to the total instances. It is represented in eqn (6).

$$Accuracy = \frac{t_p + t_n}{t_p + t_n + f_p + f_n} \quad (6)$$

Where t_p, t_n, f_p and f_n represents TP, TN, FP, and FN, respectively.

ii) Precision:

Precision measures the classifier's efficiency in identifying true positive cases to the total true instances, and it is formulated in eqn (7).

$$\Pr ecision = \frac{t_p}{t_p + t_n} \quad (7)$$

iii) Recall\Sensitivity:

Recall determines the system's proficiency to detect all relevant instances and it defines the fraction of real positive cases to

the total positive cases, as defined in eqn (8).

$$\text{Re}call = \frac{t_p}{t_p + f_p} \quad (8)$$

iv) F1-Score:

The F1-score represents the harmonic mean of precision and sensitivity (recall), providing a balanced measure of a classifier's performance in distinguishing lung cancer cases from healthy subjects. It is mathematically defined as follows: eqn (9).

$$F1 - score = 2 \times \left(\frac{sensitivity * precision}{sensitivity + precision} \right) \quad (9)$$

v) *Specificity:*

Specificity measures the model's effectiveness in detecting and classifying healthy cases (true negative) instances. It is represented mathematically in eqn (10).

$$Specificity = \frac{t_n}{t_n + f_p} \quad (10)$$

vi) *False positive rate:*

FPR represents the ratio among true negative instance that the system incorrectly categorizes as positive, as formulated in eqn (11).

$$FPR = \frac{f_p}{t_n + f_p} \quad (11)$$

vii) *False negative rate:*

FNR denotes the ratio of real positive cases incorrectly categorized as negative by the system. It is expressed in eqn (12).

$$FNR = \frac{f_n}{t_p + f_n} \quad (12)$$

viii) *Execution time:*

The execution time determines the overall time consumed by the designed model for performing lung cancer identification and categorization.

The investigation of these parameters allows us to estimate how the classifier performs lung cancer classification.

4.3 Train and Test

4.3.1 Hyperparameters and its optimized values:

Hyperparameter	**Range**	**Optimized Value**
Base Model	-	EfficientNet-B3
Dense Layer 1 units	[32,64, 128]	128
Dense Layer 2	[64, 128]	64
Dropout 2 rate	[0.1, 0.3]	0.3
Activation	-	ReLU
Batch Size	[16, 32]	32
Epoch	[30, 50,100]	100

Learning rate	[0.001, 0.01, 0.1]	[0.01]
Dense layer – Output	-	4 x (total of classes)
Dense layer- Output	-	Softmax
Optimizer	[‘Adam’, ‘sgd’, ‘rmsprop’, ‘adagrad’]	Adam

Table-1: Hyperparameters of proposed model

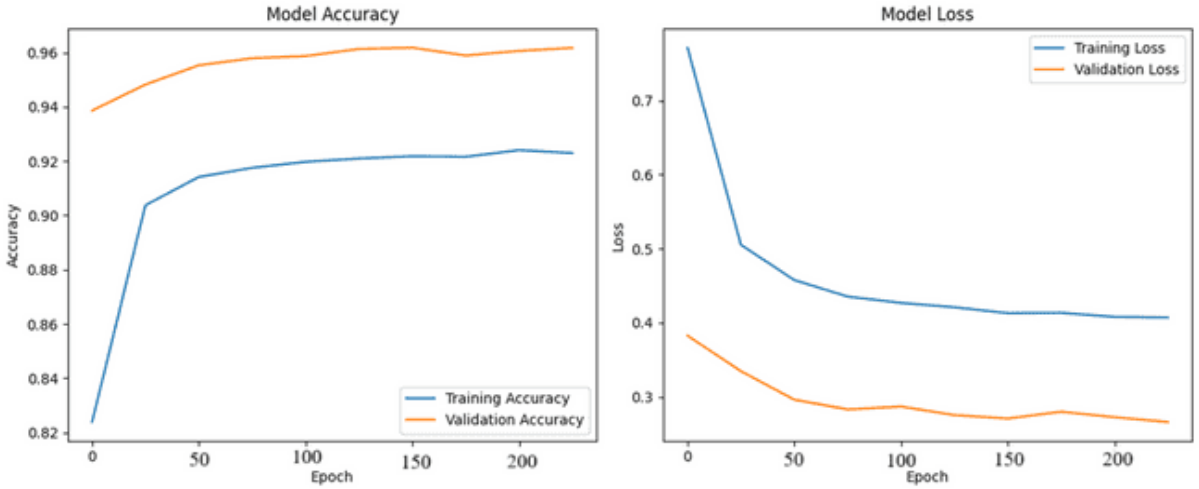


Fig 4: Model Accuracy and loss

4.3.2 Confusion Matrix and ROC curves for the EfficientNet-B3:

The confusion matrix is designed for the proposed model EfficientNet-B3, it classifies into multiclass classification of Lung Cancer into adenocarcinoma, SCC, Normal, and LCC as designed below**:**

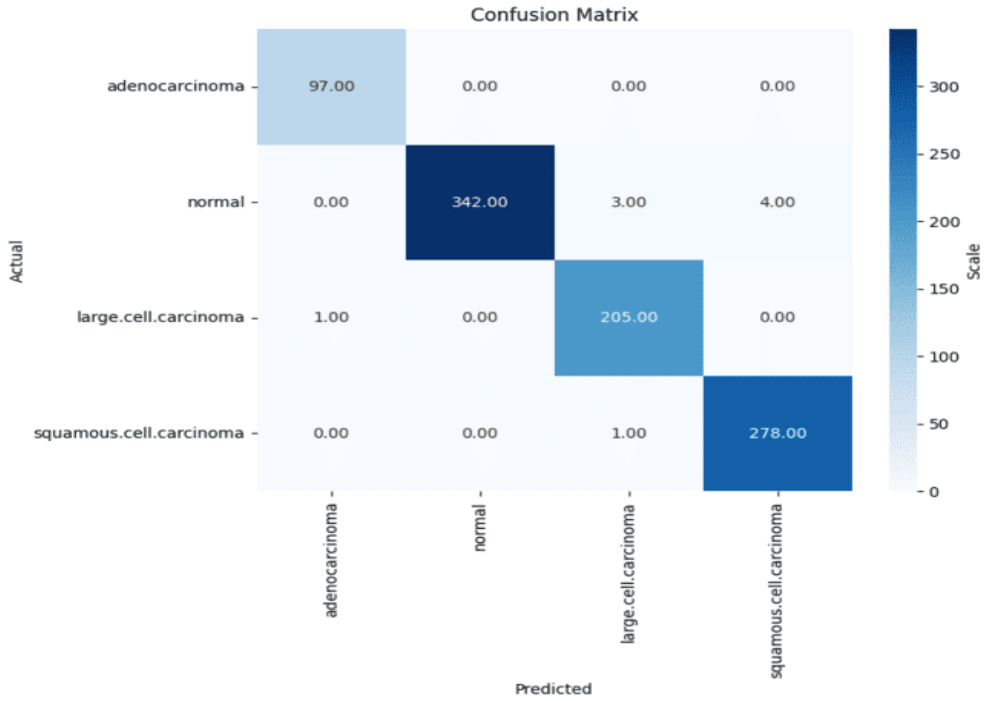


Fig 5: Confusion matrix of EfficientNet-B3

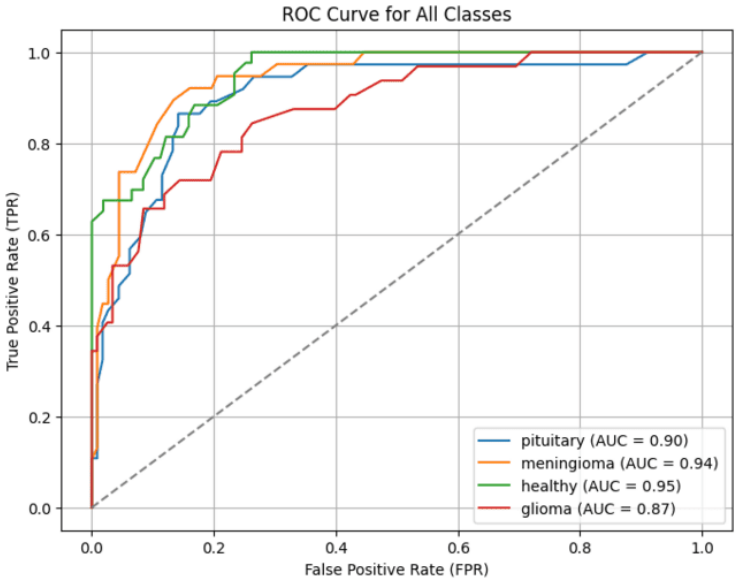


Fig 6: ROC curve

The Receiver Operating Characteristic (ROC) curve visually represents the classification effectiveness of the proposed EfficientNet-B3 model in detecting lung cancer. It demonstrates the model's ability to distinguish between lung cancer and healthy cases by analysing performance across multiple threshold values. The ROC curve is constructed by plotting the True Positive Rate (TPR) against the False Positive Rate (FPR), offering a detailed evaluation of the model's predictive strength.The Area Under the Curve (AUC) serves as a quantitative measure of classification performance, where values closer to 1.0 indicate superior accuracy, while lower values suggest diminished predictive capability. This metric is essential for assessing the robustness and reliability of the model in medical image classification tasks.

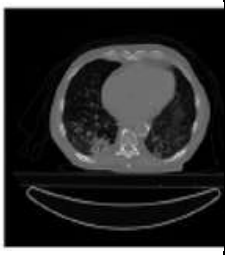
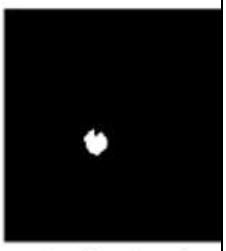
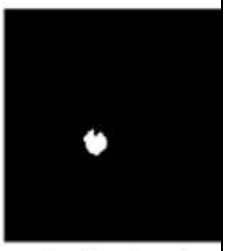
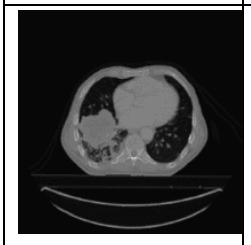
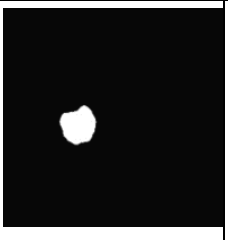
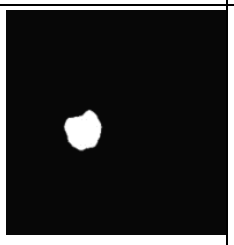
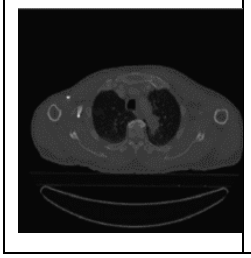
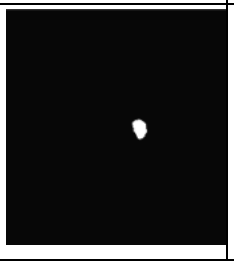
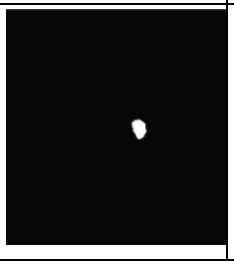
Ground truth			**Predicted**		**Accuracy**
Input	Mask image	Label	Mask	Label	
		Adenocarcinoma		Adenocarcinoma	**0.9835**
		LCC		LCC	**0.9843**
		SCC		SCC	**0.9859**

Table-2: Input and classification results with accuracy

The random input images and their classification are tabulated in Table 2. The proposed algorithm intends to categorize four classes: normal (healthy, adenocarcinoma, LCC, and SCC). The designed methodology achieved an accuracy of 0.9835, 0.9843, and 0.9859, respectively for healthy, adenocarcinoma, LCC, and SCC classes. This increased accuracy highlights that the presented framework offers the finest categorization of lung cancer from CT images.

4.4 Comparative Assessment

Model	Accuracy	Recall	Precision	F1-Score
ResNet-50	91%	0.91	0.91	0.91
ResNet-101	97%	0.97	0.97	0.97
VGG-16	94%	0.94	0.94	0.94
VGG-19	95%	0.95	0.95	0.95
EfficientNet-B0	96%	0.96	0.96	0.96
EfficientNet-B1	96.4%	0.96	0.96	0.96
EfficientNet-B2	97.3%	0.97	0.97	0.97
EfficientNet-B3	98.0%	0.98	0.98	0.98

Table-3: Comparisons of Performance with fine-tuned models

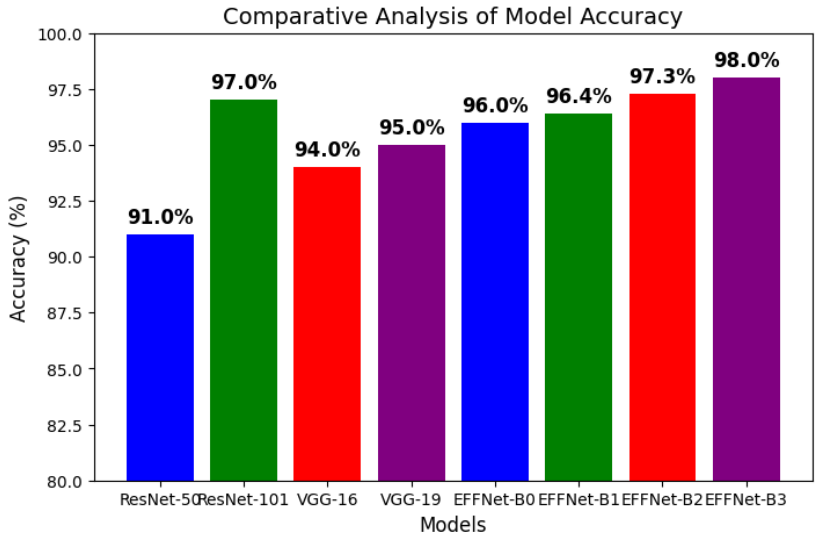


Fig 7: Comparative analysis of model accuracy

5. CONCLUSION AND FUTURE WORK

This paper suggests a deep learning-based approach utilizing EfficientNet-B3 architecture for detecting lung cancer, demonstrating its ability to enhance imaging technology and patient well-being. The designed algorithm effectively differentiates between abnormal and normal tissue in CT images, reaching an accuracy of 98%, outperforming conventional classifiers. Additionally, the model exhibits strong generalizability, highlighting its applicability across diverse clinical cases. The high classification accuracy underscores its potential to improve early diagnosis, treatment planning, and overall patient outcomes, thereby reducing the lung cancer burden pn healthcare.

Future exploration should emphasize validation across a wider, multi-institutional datasets, with integration of more clinical and genomic information to enhance diagnostic precision. Furthermore, advancements in deep learning architectures, transfer learning, and hybrid AI models may significantly refine both the efficacy and efficiency of automated lung cancer detection.

REFERENCES

[1] X Zhan, H Long, F Gou, X Duan, G Kong, A convolutional neural network-based intelligent medical system with sensors for assistive diagnosis and decision-making in non-small cell lung cancer. Sensors 2021, 21(23), 7996; https://doi.org/10.3390/s21237996.

[2] Xueyun Tan, Feng Pan, Na Zhan, Sufei Wang, Zegang Dong, Yan Li, Guanghai Yang, Bo Huang,Yanran Multimodal integration to identify the invasion status of lung adenocarcinoma intraoperatively. https://doi.org/10.1016/j.isci.2024.111421

[3] Yu, L.; Tao, G.; Zhu, L.; Wang, G.; Li, Z.; Ye, J.; Chen, Q. Prediction of pathologic stage in non-small cell lung cancer using machine learning algorithm based on CT image feature analysis. BMC Cancer 2019, 19, 464; https;//doi.org/10.1186/s12885-019-5646-9

[4] Huan Yang1, Lili Chen2, Zhiqiang Cheng3, Minglei Yang4, Jianbo Wang5, Deep learning-based six-type classifier for lung cancer and mimics from histopathological whole slide images: a retrospective study. http:// doi.org/10.1186/s12916-021-01953-2.

[5] Masud, M.; Sikder, N.; AlNahid, A.; Bairagi, A.K.; Alzain, M.A. A machine learning approach to diagnosing lung and colon cancer using a deep learning-based classification framework. Sensors **2021**, 21, 748.

[6] Wu, J.; Chen, Z.; Zhao, M. An efficient data packet iteration and transmission algorithm in opportunistic social networks. J. Ambient. Intell. Humaniz. Comput. **2020**, 11, 3141–3153.

[7] Bębas, E.; Borowska, M.; Derlatka, M.; Oczeretko, E.; Hładuński, M.; Szumowski, P.; Mojsak, M. Machine-learning-based classification of the histological subtype of non-small-cell lung cancer using MRI texture analysis. Biomed. Signal Process. Control **2021**, 66, 102446.

[8] Guo, J.; Wang, C.; Xu, X.; Shao, J.; Yang, L.; Gan, Y.; Yi, Z.; Li, W. DeepLN: An artificial intelligence-based automated system for lung cancer screening. Ann. Transl. Med. **2020**, 8, 1126

[9] Ahmad, A.S.; Mayya, A.M. A new tool to predict lung cancer based on risk factors. Heliyon **2020**, 6, e03402

[10] Cui, L.; Li, H.; Hui, W.; Chen, S.; Yang, L.; Kang, Y.; Bo, Q.; Feng, J. A deep learning-based framework for lung cancer survival analysis with biomarker interpretation. BMC Bioinform. **2020**, 21, 112.

[11] Zhang, Y.H.; Lu, Y.; Lu, H.; Zhou, Y.M. Development of a Survival Prognostic Model for Non-small Cell Lung Cancer. Front. Oncol. 2020, 10, 362.

[12] Wang, J.; Chen, N.; Guo, J.; Xu, X.; Liu, L.; Yi, Z. SurvNet: A Novel Deep Neural Network for Lung Cancer Survival Analysis with Missing Values. Front. Oncol. 2021, 10, 588990.

[13] Huang, Z.; Hu, C.; Chi, C.; Jiang, Z.; Tong, Y.; Zhao, C. An Artificial Intelligence Model for Predicting 1- Year Survival of Bone Metastases in Non-Small-Cell Lung Cancer Patients Based on XGBoost Algorithm. BioMed Res. Int. **2020**, 2020, 3462363.

[14] Lu, C.; Bera, K.; Wang, X.; Prasanna, P.; Xu, J.; Janowczyk, A.; Beig, N.; Yang, M.; Fu, P.; Lewis, J.; et al. A prognostic model for overall survival of patients with early-stage non-small cell lung cancer: A multicentre, retrospective study. Lancet Digit. Health **2020**, 2, e594–e606.

[15] Lai, Y.H.; Chen, W.N.; Hsu, T.C.; Lin, C.; Tsao, Y.; Wu, S. Overall survival prediction of non-small cell lung cancer by integrating microarray and clinical data with deep learning. Sci. Rep. 2020, 10, 4679.

[16] She, Y.; Jin, Z.; Wu, J.; Deng, J.; Zhang, L.; Su, H.; Jiang, G.; Liu, H.; Xie, D.; Cao, N.; et al. Development and Validation of a Deep Learning Model for Non-Small Cell Lung Cancer Survival. JAMA Netw. Open **2020**, 3, e205842.

[17] Cui, R.; Chen, Z.; Wu, J.; Tan, Y.L.; Yu, G.H. A Multiprocessing Scheme for PET Image Pre-Screening, Noise Reduction, Segmentation and Lesion Partitioning. IEEE J. Biomed. Heal. Inform. **2021**, 25, 1699–1711.

[18] Baek, S.; He, Y.; Allen, B.G.; Buatti, J.M.; Smith, B.J.; Tong, L.; Sun, Z.; Wu, J.; Diehn, M.; Loo, B.W.; et al. Deep segmentation networks predict survival of non-small cell lung cancer. Sci. Rep. **2019**, 9, 17286.

[19] Lee, B.; Chun, S.H.; Hong, J.H.; Woo, I.S.; Kim, S.; Jeong, J.W.; Kim, J.J.; Lee, H.W.; Na, S.J.; Beck, K.S.; et al. DeepBTS: Prediction of Recurrence-free Survival of Non-small Cell Lung Cancer Using a Time-binned Deep Neural Network. Sci. Rep. 2020, 10, 1952

[20] Luo, S.; Xu, J.; Jiang, Z.; Liu, L.; Wu, Q.; Leung, E.L.H.; Leung, A.P. Artificial intelligence-based collaborative filtering method with ensemble learning for personalized lung cancer medicine without genetic sequencing. Pharmacol. Res. 2020, 160, 105037.

[21] Wu, J.; Zhuang, Q.; Tan, Y. Auxiliary Medical Decision System for Prostate Cancer Based on Ensemble Method. Comput. Math. Methods Med. 2020, 2020, 6509596.

[22] Cui, R.; Chen, Z.; Wu, J.; Tan, Y.L.; Yu, G.H. A Multiprocessing Scheme for PET Image Pre-Screening, Noise Reduction, Segmentation and Lesion Partitioning. *IEEE J. Biomed. Heal. Inform.* 2021, *25*, 1699–1711.

[23] Lee, B.; Chun, S.H.; Hong, J.H.; Woo, I.S.; Kim, S.; Jeong, J.W.; Kim, J.J.; Lee, H.W.; Na, S.J.; Beck, K.S.; et al. DeepBTS: Prediction of Recurrence-free Survival of Non-small Cell Lung Cancer Using a Time-binned Deep Neural Network. *Sci. Rep.* **2020**, *10*, 1952.

[24] Luo, S.; Xu, J.; Jiang, Z.; Liu, L.; Wu, Q.; Leung, E.L.H.; Leung, A.P. Artificial intelligence-based collaborative filtering method with ensemble learning for personalized lung cancer medicine without genetic sequencing. *Pharmacol. Res.* **2020**, *160*, 105037.

[25] Wu, J.; Zhuang, Q.; Tan, Y. Auxiliary Medical Decision System for Prostate Cancer Based on Ensemble Method. *Comput. Math. Methods Med.* **2020**, *2020*, 6509596.

[26] Lin, Chun-Hui, Cheng-Jian Lin, Yu-Chi Li, and Shyh-Hau Wang. "Using generative adversarial networks and parameter optimization of convolutional neural networks for lung tumor classification." Applied Sciences 11, no. 2 (2021): 480.

[27] Meldo, A.; Utkin, L.; Kovalev, M.; Kasimov, E. The natural language explanation algorithms for the lung cancer computer-aided diagnosis system. *Artif. Intell. Med.* **2020**, *108*, 101952

[28] Song, Z.; Liu, T.; Shi, L.; Yu, Z.; Shen, Q.; Xu, M.; Huang, Z.; Cai, Z.; Wang, W.; Xu, C.; et al. The deep learning model combining CT image and clinicopathological information for predicting ALK fusion status and response to ALK-TKI therapy in non-small cell lung cancer patients. *Eur. J. Nucl. Med. Mol. Imaging* **2021**, *48*, 361–371

[29] Sung, H.; Ferlay, J.; Siegel, R.L.; Laversanne, M.; Soerjomataram, I.; Jemal, A.; Bray, F. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J. Clin.* **2021**, *71*, 209–249.

[30] Lin, Chun-Hui, Cheng-Jian Lin, Yu-Chi Li, and Shyh-Hau Wang. "Using generative adversarial networks and parameter optimization of convolutional neural networks for lung tumor classification." Applied Sciences 11, no. 2 (2021): 480.

[31] Dafni Rose, J., K. Jaspin, and K. Vijayakumar. "Lung cancer diagnosis based on image fusion and prediction using CT and PET image." Signal and image processing techniques for the development of intelligent healthcare systems (2021): 67-86.