RESEARCH PAPER

An Attention-Enhanced EfficientNetV2-S Framework for Multi-Class Lung Cancer Classification from CT Images

Karishma Rajendra Patil[1] and Dr. Nitika Vats[2]
[1]Research Scholar, Department of CSE, SAGE University, Indore, MP 452020, India.
E-mail: patilkarishma308@gmail.com
[2]Professor, Department of CSE, SAGE University, Indore, MP 452020, India.
E-mail: Headtraining@sageuniversity.in

Abstract
Lung cancer is a major worldwide health problem and diagnostic imaging still poses a major challenge because of the overlapping morphology of different histological types in computed tomography (CT) scans. For this purpose deep learning framework named EfficientNetV2-S–CBAM which is capable of classifying pulmonary CT images at the four levels – Adenocarcinoma, Large Cell Carcinoma, Squamous Cell Carcinoma and Healthy parenchyma is proposed. The methodological approach started with the careful pre-processing of the dataset where an initial database of 1,000 images was reduced to 824 images both by exact and perceptual deduplication algorithm. This cleaned up data set was then divided in a 70:15:15 split to create training, validation and testing sets. The backbone of our system is an ImageNet-pretrained EfficientNetV2-S network which is complemented by a Convolutional Block Attention Module (CBAM) that is designed to work synergistically with the backbone network to enhance the spatial and channel-wise feature maps. We adopted the composite pooling strategy, which combines the global average and maximum operations and a focal loss function to overcome the effects of class imbalance as well as a two-stage transfer learning protocol. The results from the empirical evaluation on the independent test set (n=124) were found to achieve high accuracy (87.90%), macro F1 score (89.56%) and macro ROC-AUC (96.68%). The results confirm the strategic transfer learning, feature enhancement through attention and careful data curation are a very promising approach for automatic multi-class lung CT diagnosis.
Keywords: computed tomography (CT) scans, Deep Learning, EfficientNetV2-S, Convolutional Block Attention Module (CBAM)
How to cite this article: Karishma Rajendra Patil1 and Dr. Nitika Vats2. An Attention-Enhanced EfficientNetV2-S Framework for Multi-Class Lung Cancer Classification from CT Images. Int J Drug Deliv Technol. 2026;16(89s):742-752. DOI: 10.25258/ijddt.16.89s.81.

1. Introduction

Various factors such as environmental pollution, climate change, changed lifestyles and other social and behavioural factors have contributed to the global burden of disease. These factors can affect the likelihood of various health issues and make disease prevention and management challenging. Respiratory diseases are important amongst these health issues and are a major cause of morbidity and mortality in the world. Lung cancer is a severe lung disease. When cells in the lung start to grow and multiply abnormally this occurs. These cells may develop into cancerous tumours if left untreated and can grow and spread to other parts of the body. Lung cancer is one of the most fatal cancers with a high mortality rate thus the need for effective and dependable diagnostic tools. Early and accurate diagnosis of lung cancer and its histological type is especially crucial as this will enable therapy to start at an earlier stage and may enhance patient outcome. Computed Tomography (CT) is very popular to look at pulmonary nodules and suspected lung abnormalities. However, interpreting CT images manually can be time-consuming and may vary between radiologists. This difference can sometimes lead to problems with diagnosis or interpretation. Medical image analysis has therefore started to become a growing interest for Machine Learning (ML) and Deep Learning (DL) techniques. These methods can be helpful in identifying abnormal pattern in computed tomography images and may help to enhance the efficiency and uniformity of lung cancer diagnosis.

This study aims to overcome the above limitations by presenting a three-part lung CT classification system. The first step is to not use the original train, validation, and test folders directly, but to merge them into a single set of images and check them for any replicas. It is using two-stage duplicate-detection process. The first stage is an exact hash that compares files on a byte level, and the second uses a strict, high-resolution perceptual hash to compare similar images on a perceptual level. Once the duplicates have been removed, a new stratified dataset split is generated and the split between the train and test sets is systematically checked. The intent of this process is to minimize the risk of data leakage and to give a more convincing assessment of model generalization. Secondly, the classification model to be proposed is mainly based on EfficientNetV2S as the convolutional backbone. The model is pre-trained on ImageNet, and it is improved by adopting a Convolutional Block Attention Module (CBAM) on the last spatial feature map. CBAM works by offering two types of attention: channel attention and spatial attention, allowing the network to focus on specific channels

and regions of the image that are relevant. This is especially relevant for CT classification as the diagnostically useful part of the image may be located in a small section of the image. Third, it is a moderately imbalanced (4 classes) relatively small dataset, so the proposed framework combines the use of focal loss with a two-stage transfer-learning approach. The training starts with a frozen backbone and finer tuning is performed over the deeper layers of the backbone. Horizontal-flip test-time augmentation is also used for inference to make the prediction more reliable.

The proposed framework classifies 4 classes of lung CT images. There are 3 main subtypes of non-small-cell lung cancer (NSCLC) in the classes and separate class for normal controls. Adenocarcinoma are cancers that grow in glandular cell, and are the most common cancer in the periphery of the lungs. The second type of cancer is Squamous Cell Carcinoma which develops from the squamous cells in the airways and is typical of smoking-related lung cancer. Large Cell Carcinoma is the third class of NSCLC which is undifferentiated and can develop in any part of the lung and is usually regarded as aggressive. The fourth class is healthy lung tissue considered the Normal class and is used as a control class for the classification task. It is difficult to differentiate these four categories with CT images as different kinds of lung cancers may have similar appearances. Importantly, even if both have different origins, Adenocarcinoma and Squamous Cell Carcinoma can present a similar appearance on CT. This makes the fine-grained classification a challenge and it is a need to have a strong deep learning approach.

An EfficientNetV2S-CBAM framework is tested on a test set with a leakage and duplicate removal process. The test accuracy of the model is 87.90%, macro-averaged F1-score is 89.56% and macro-averaged ROC-AUC is 96.68%. The model is able to accurately classify all Large Cell Carcinoma cases in the test set and no healthy lung cases are classified as malignant. The train-test separation is specifically guaranteed to prevent duplicate or near-duplicate images occurring across the partitions in particular. The reported results are thus presented as being more reliable indicators of the generalization of the model than those that would not have been reported without a specific comparison of the overlap of the datasets.

The rest of this paper is organized as follows. Section 2 reviews relevant literature on Lung Cancer Classification in the context of Deep Learning in the field of CT and other imaging modalities. Methods and Materials are described in Section 3. The experimental results are discussed in section 4 covering Overall performance, Class wise analysis and comparison with previous work.

2. Literature Survey

In recent years the diagnosis of lung disease has shifted from traditional image processing methods to deep learning, hybrid learning and explainable AI. Chest X-ray and computed tomography (CT) imaging have been the primary imaging modalities because they offer clinically relevant visual information which can be automatically processed. The present studies, however, vary significantly with respect to architecture, processing method, feature representation and classification approach. The results suggest that deep learning-based models could potentially yield high diagnostic accuracy, but there are still concerns about generalization, interpretability, dataset diversity, and clinical validation.

One field of research has been on applying convolutional neural networks (CNNs) to chest radiographic images. Sharma et al. proposed a CNN model to detect lung disease like pneumonia, tuberculosis, and lung cancer in preprocessed X-ray images of chest. Pre-processing operations such as image resizing, image normalization, and image augmentation were also performed prior to model training and the system showed good accuracy, precision, recall and F1 score [4]. Authors explored a hybrid architecture that integrates VGG for feature extraction with CNN for classification for the recognition of lung disease from chest X-ray images. Data augmentation was used to augment the diversity of training samples and the NIH chest x ray dataset from Kaggle was used for testing. The reported test accuracy of 73%, however, shows that there is still room for improvement of the architecture in detecting pulmonary abnormalities. In addition to the traditional CNN structures researchers have explored lightweight and interpretable models for better potential usability of automated diagnosis. Zahin et al. suggested a compact hybrid CNN-RNN model with a model size of 7.25 MB for lung disease recognition of four classes based on cross sectional chest X-rays. The framework integrated explainable AI and Grad-CAM was used to give visual cues of the image regions that were impacting the predictions. The accuracy and F1-score reported were 93.73% and 94.6% respectively [1].

Diagonal attention has additionally been given to explainability in systems with chest X-rays for diagnostics. Hroub et al. developed an explainable Inception-V3 based approach for distinguishing pneumonia, COVID-19 and normal CXRs. The training data were enhanced using operations such as cropping, rotation, and flipping horizontally. The authors also compared five different class activation mapping (CAM) methods and found that their method outperformed the tested vision transformer methods [3]. Even with these results the investigation was limited to two chest X-ray datasets and external validation and real-time clinical deployment were not included. Thus, the

performance of explainable deep learning systems in diverse patient populations and real-world clinical settings is yet insufficiently proven. Other studies have focused on the use of CT to perform cancer diagnosis and analysis of pulmonary abnormalities. Suvarchala et al. used histogram equalization with DenseNet121 as feature extraction method and stacked several classifiers using Logistic Regression as a meta-learner, namely LR, SVM,RF, XGBoost, KNN. The deep learning method was found to be accurate at 98.5%, which suggests that the hybrid of deep feature learning and several conventional classifiers could be beneficial [5]. There is a lack of external multi-centre validation and explainability mechanisms which restrict its direct clinical use. More specifically, the strength of the stacking architecture in the context of CT examinations with noise and its capacity to include information from various phases of disease progress is yet to be answered.

Some hybrid methods have also been investigated to discriminate particular lung cancer types. Ahmed et al. created an XGBoost-ResNet101 framework to classify NSCLC versus SCLC. They used a methodology that involved texture information extracted from GLCM analysis and tested several classification algorithms. The diagnostic accuracy of the classifiers investigated was 97.0% for KNN and 83.0% for SVM [9]. Even if the same information derived from images is used, the effectiveness of lung cancer prediction can be greatly affected by the selection of the classifiers. Even though there is a thrust toward purely deep learning pipelines, traditional image-processing and machine-learning techniques still offer valuable alternatives. Nageswaran et al. analyzed the CT images of 70 patients, used a geometric mean filter to reduce noise and enhance the image quality. It was then followed by K-means clustering for the identification of the regions of interest and classification using ANN, KNN, and RF. For assessing pulmonary abnormalities, ANN was found to be the most effective classifiers among the classifiers considered for the comparative analysis [10]. In another hybrid approach Ali et al. used machine learning and CNN techniques to distinguish between the lung image in the absence of lung nodular structures with benign and malignant nodular structures. In the pre-processing steps, Bayesian thresholding and Taylor series approximation were applied and particle swarm optimization (PSO) was used to optimize the classification process. The classification accuracy was reported as 98.45%, which indicated the effectiveness of an image enhancement and optimization combined with learning based classification [11].

Multi-stage pipelines have also been integrated with segmentation and feature extraction. Venkatesh et al. suggested a medical image analysis sequential algorithm where Otsu thresholding was applied to remove the background structures from the relevant lung regions. Cuckoo Search optimization was then added to enhance the pulmonary boundary delineation before the discriminative features were extracted and provided to a CNN classifier. The resulting system had an accuracy of 96.97% [14]. These multi-stage approaches emphasize the need of pre-processing and region identification in medical imaging analysis. The literature also argues that it is necessary to establish whether these traditional optimization and segmentation strategies are still effective when compared to the more recent architecture based on deep learning and transformers.

Other medical imaging sources have been explored for deep learning. Patel et al. performed a deep learning study on the characterization of pulmonary diseases through bronchial images. They concluded that they had a good diagnostic performance when compared to human medical practitioners, highlighting the importance of deep learning in computer-aided diagnostic technologies. Complementary image-processing techniques which might help in clinical decision making were also considered in the study [12]. While these results reflect the potential of AI in medical image interpretation, broader adoption and widespread clinical trust necessitate validation across a wider range of clinical scenarios.

One of the significant advances in the prediction of lung cancer is the transition from image-only analysis to multimodal patient information. Li et al. made machine learning prediction models based on the information of 2,115 lung cancer patients and the clinical variables and blood-based protein biomarkers. The following classifiers were tested: RF, LR, KNN, AdaBoost, XGBoost, and CatBoost. CatBoost was reported to have the highest accuracy of 97% and the fastest speed. The study showed that use of heterogeneous patient information and ensemble machine learning methods could assist in clinical decision making and help to develop personalized treatment paths [15]. This is especially significant because the imaging results may not provide all the information needed to assess the lung cancer for individual patients.

Bhuvaneswari et al. in their systematic review considered risk factors for lung diseases such as asthma, COPD, fibrosis, smoking, air pollution and genetic; a similar approach was taken. The study emphasized the use of deep learning methods using CT and X-ray images, and described the architectures of ResNet50 and DenseNet as a method for early prediction [2]. But most of the discussion was based on the public data available at the time and there was no new dataset for more extensive validation. The restricted number of transformer/hybrid architectures also provides opportunities for comparative analysis.

Additionally, complementary clinical measurements, like spirometry, would allow to better represent restrictive and other pulmonary disorders if combined with imaging measurements.

Critical Synthesis of Existing Research

The literature surveyed shows a clear advancement from traditional image processing and machine-learning classification techniques to CNN, transfer-learning architectures and hybrid models, ensemble learning, optimization method and explainable AI. CNN-based methods have achieved significant success in discovering discriminative patterns from chest X-ray and CT images, and hybrid and ensemble methods have tried to boost the predictive performance of existing classifiers by incorporating deep representations. The reported results vary from 73% for the VGG-CNN hybrid approach to 98.5% for the DenseNet121-based stacking framework, but such results should not be compared as they differ in datasets, disease categories, imaging modalities, and evaluation protocols [5,18].

It has been also demonstrated that preprocessing and feature localization are crucial issues in medical image classification processes. Many techniques have been applied to enhance the quality of the image and to separate out the diagnostically relevant portions of the image, including histogram equalization, geometric filtering, thresholding, clustering and optimized segmentation [5, 10, 11, 14]. Meanwhile, methods of explainability, like Grad-CAM and CAM, have become significant ways by which deep learning predictions can be made interpretable to clinicians [1, 3]. This is especially true in the field of medical diagnosis where other factors besides high predictive accuracy may be important for clinical acceptance.

Despite all these developments there are still some fundamental methodological shortcomings which can be noticed. Publicly available datasets are often used in many investigations, and often without the use of independent external validation. Therefore, high performance on one data set does not generally map to heterogeneous environments with multiple centres. [1, 2, 4, 5] Model performance can be affected by the characteristics of the dataset, image quality, acquisition protocols, disease prevalence, patient demographics, etc. Therefore, the few studies conducted with noisy real world images are still of interest.

In addition, there is a need for further development in the field of interpretability. While some investigations have incorporated Grad-CAM or other CAM based approaches [1, 3] there are a number of high performing hybrid systems and ensembles that lack explainability [5]. If there is no evidence that can be interpreted as to which anatomical or pathological areas affect the prediction, it will be difficult for the clinician to judge whether automated decisions have medical significance. The literature also suggests that there is a lack of integration of multimodal and longitudinal information. Most image-based studies do not use imaging in conjunction with clinical measurements, blood markers, or previous imaging exams. However, the work of Li et al. illustrates the possibility of the clinical and protein blood information to be fused to make prediction at patient level [15]. Predicting lung disease using images could give a more complete picture of patient condition, and extending these multimodal approaches to the field could help in the prediction task.

Last, while conventional CNN and hybrid models have shown promising results, the comparative evaluation with current transformer-based models is only addressed in few of the studied works [1, 2, 14]. In a similar manner, the longitudinal monitoring and evaluation of disease on multi-centre, heterogeneous datasets is relatively under explored. Future studies should extend current experiments with high accuracy to effective, understandable, multimodal, and externally tested prediction models that can be used in real clinical setting.

Previous literature shows a lot of advancements in automated lung disease and lung cancer prediction with deep learning, hybrid architectures and ensemble machine learning. There are several important gaps, however, that are not filled. Most of the models already available are assessed on a single or on public datasets that have not been independently tested on different centers, reducing the confidence in their generalizability to a wide range of clinical settings [1,2,4,5]. Moreover, explainability is not consistently applied, especially in the most successful hybrid and ensemble models, that reduce the clinical transparency of automated prediction [5]. A limited evaluation on real-world medical images that contain noise has been performed in the literature and few comparisons have been made to the current transformer-based architectures [1,2,14]. Image-based systems tend to focus on individual exams, without integrating longitudinal patient history and complementary clinical and biomarker data. While multimodal prediction has been shown to have promise with clinical and blood-based information [15], the incorporation of these modalities with medical imaging has not been thoroughly explored. These restrictions suggest the need for a comprehensive prediction platform capable of integrating correct image representation, suitable classification methods, interpretable prediction, and additional data from the patient's level, in different and realistic clinical settings and prove to be challenging.

2. Materials and Methods

This work studies the application of deep learning to classify lung cancer from a publicly available medical imaging dataset from Kaggle repository. The dataset includes illustrations of pathological and normal lungs. The total number of images that were

put up for experimental investigation is 1000 images. Images were then classified into 4 categories. The first three classes are the various types of lung cancer and the Normal class is the control class. The multi-class setting allows to split the developed classification framework into cancer vs. non-cancer and different lung cancer types.

There is also a variant of NSCLC in which the cells are poorly differentiated (or undifferentiated) and have relatively large irregular shapes called Large Cell Carcinoma. Can occur in various parts of the lungs and is a more serious type of lung cancer. Squamous Cell Carcinoma develops from the bronchial passages in the squamous epithelial cells. This condition is supplied with a description of a dataset that includes a large smoking history and suggests that it is common in the central parts of the lungs where a cavity might form. Lung images with no obvious cancerous growth is considered as the Normal category which is used as a reference class for the model training and evaluation.

The entire data set was split into training, test and validation sets for experimental development. 70% of the images have been given to train the system, 15% for testing and 15% for validation. The training subset was used to train the deep learning model so that it could learn the discriminative features of the 4 target classes. To evaluate the learning performance, the validation subset was used during the model-development stage while the testing subset was reserved for the final model evaluation with images that were not used to develop the model. The overall experimental flow was performing data acquisition, organizing the data into classes, partitioning the data, developing deep learning models, model validation, and testing the model. The classification problem was then converted to a four class problem.

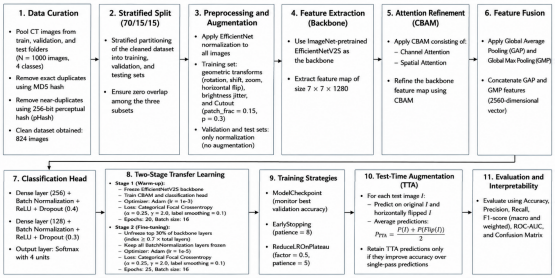


Figure 1: Proposed Methodology

The proposed framework adopts an end-to-end deep learning pipeline to classify CT images into four classes which involves data curation, preprocessing, feature extraction, attention refinement, feature fusion, classification, transfer learning, training optimization and robust inference. All the available data sets were first combined into 1,000 CT images. To ensure the integrity of the data exact duplicate data was deleted using MD5 hashes and near-duplicate data was deleted by applying the perceptual hash (pHash), a 256-bit hash. This resulted in 824 curated images that were then split in two-thirds, one-third and one-third into training, validation and testing subsets respectively representing each class an equal number of times in each of the three subsets.

All images were normalized to EfficientNet. In addition to normalization, training images were also subjected to rotation, shifting, zooming, horizontal flipping, brightness jitter, and Cutout (patch fraction = 0.15, p=0.3) while validation and test images were only applied with normalization. The ImageNet-pretrained EfficientNetV2-S backbone was used for feature extraction, with a feature map of size $7 \times 7 \times 1280$. The extracted representation was further modified by adopting the CBAM which sequentially processes channel and spatial attention to achieve adaptive weights of feature channels and spatial positions. Global Average Pooling (GAP) and Global Max Pooling (GMP) were used to process the refined feature map. They are each 1280 dimensional and they are concatenated to create a 2560 dimensional feature vector for each operation. This fused representation is fed into the dense layers of the classification head of 256 and 128 units both with batch normalization, ReLU activation and dropout rates 0.4 and 0.3, respectively. The class probabilities are computed by a final four-unit softmax layer.

Two-stage transfer learning is used for model adaptation. In warm-up stage the EfficientNetV2-S backbone structure is frozen and the CBAM and classification are trained. Then, fine-tuning enables the top part of the backbone to be adapted to the task classification in the target CT. The training is controlled by the following methods: ModelCheckpoint, EarlyStopping (patience = 8), ReduceLROnPlateau (factor = 0.5, patience = 5). The included data set description correlates this condition with a heavy smoking history and suggests that it usually occurs in the middle regions of the lungs where cavitation can also happen. The Normal category includes lung images that contain no visible signs of growth of cancer cells and is used as a reference class for training and testing the model.

The whole data set was split into a training set, testing set, and validation set for experimental development. The images were divided into training (70%), testing (15%) and validation (15%). The training subset was used to train the deep learning model allowing it to learn discriminative characteristics of the four target classes. During the model-development phase, the validation subset served to evaluate the learning performance while the testing subset was employed to validate the final model with images it had not seen before. The overall experimental flow was dataset acquisition, class-wise organization of datasets, partitioning of the datasets, development of deep learning models,

validation and final testing.

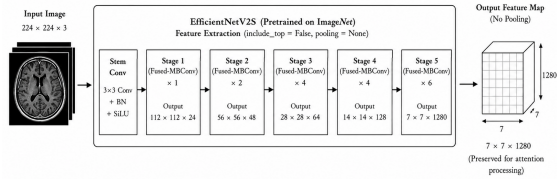


Figure 2: EfficientNetV2S Model Architecture

EfficientNetV2S is employed as the pretrained convolutional backbone for extracting discriminative representations from the input CT Scan images. The network receives a 224 × 224 × 3 input and initially processes it through a stem convolutional layer comprising convolution, batch normalization and SiLU activation. The resulting representation is progressively transformed through successive Fused-MBConv stages where spatial resolution decreases while the number of feature channels increases. As illustrated in Figure 2 the intermediate representations evolve from 112 × 112 × 24 to 56 × 56 × 48, 28 × 28 × 64 and 14 × 14 × 128 before reaching a 7 × 7 × 1280 feature representation. The backbone is configured with include_top=False to exclude the original classification head, while pooling=None preserves the spatial feature structure. Consequently, the final 7 × 7 × 1280 feature map retains spatial information required for subsequent attention-based processing and classification.

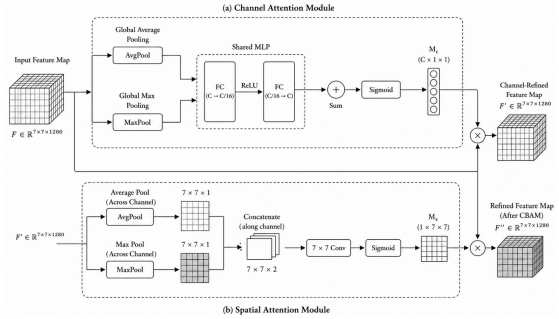


Figure 3: Convolution Block Attention Module

In order to increase the discriminative value of the features extracted from CT scans that are employed in the classification of lung disease this study introduces a CBAM into the deep learning workflow as shown in figure 3. The CBAM works through two cascaded sub-modules: channel attention and spatial attention working together to highlight salient pathological information and simultaneously reduce irrelevant and noisy information. Given a feature tensor extracted: $F \in R^{7\times7\times1280}$ In the channel attention pathway the global average pooling and global max pooling operations are first applied to the spatial information in the channel. Two complementary channel-wise descriptors are generated from these operations and passed on to a shared multilayer perceptron (MLP) with a bottleneck structure (reduction ratio r) to include channel-wise dependencies. The output of the MLP is used to element-wise sum up and then passed through a sigmoid non-linearity to get the channel attention map Mc. This map will then be multiplied by the original feature map F. channel-recalibrated feature map F'.

The localizing relevant regions within F'. It downsamples the channel dimension by using average and max pooling operations on the channel axis to produce two different 2D spatial descriptors of size 7×7×1. These descriptors are linked together and convolved with a large-kernel (7×7). The spatial attention map is obtained by passing through the (7×7) filter and a sigmoid activation Ms. The final product which has been cleaned up at the end of the process F″ is the product of Ms with F, an efficient spatial weighting that reflects diagnostic significance. This attention mechanism is used to focus the attention of the model on feature channels and spatial localities that are most indicative of pulmonary abnormalities, in the particular application of lung disease prediction. This allows the network to generate more powerful and class-discriminative representations, resulting in overall better classification performance for CT images.

Result and Discussion

Empirical benchmarks were set on Kaggle Chest CT-Scan Images to assess the diagnostic power of the suggested EfficientNetV2S–CBAM scheme for pulmonary disease classification. The raw repository contained 1,000 CT scans from patients with various types of cancer including Adenocarcinoma, Large Cell Carcinoma and Squamous Cell Carcinoma, and after careful deduplication, a verified baseline of 824 unique images was curated. To maintain category distributions, these data were split using a 70:15:15 stratified allocation into the training set (576 images), validation set (124 images), and independent testing set (124 images). The next evaluation includes convergence characteristics during optimization, validation point check-ins, efficacy with an unbiased test-set, class-specific statistics, confusion matrices, and failure mode diagnostics, as well as attention-driven interpretation results. It also compares the obtained results with related studies and discusses the major limitations of the evaluation**.**

The total number of parameters in the proposed EfficientNetV2S–CBAM network is 21,228,183. Model optimization was performed using an asymmetrical two-phase schedule to prevent degradation of the pretrained optical representations while reducing the domain divergence between standard optical imagery and thoracic computed tomography (CT) scans.

The convolutional feature extraction backbone was completely frozen in the first stage (Stage 1: Epochs 1-20). Only 896,055 trainable weights (4.22% of the whole network capacity) were updated by parameters, which are placed in the CBAM and the terminal classification dense layers. The initial learning rate of Adam was set at 1×10^{3} to minimize

a categorical focus cross entropy loss function (α=0.25,γ=2.0, label smoothing = 0.1). The moderate class disparity of the data set has been taken into account, and through careful consideration, focal loss has been chosen over traditional static inverse-frequency weighting (7.58× maximum imbalance ratio). This objective effectively reduced the influence of gradient terms from the high-confidence categories in the presence of gradients from the other minority categories, while maintaining stable convergence across the more data-rich categories, particularly the Normal cohort.

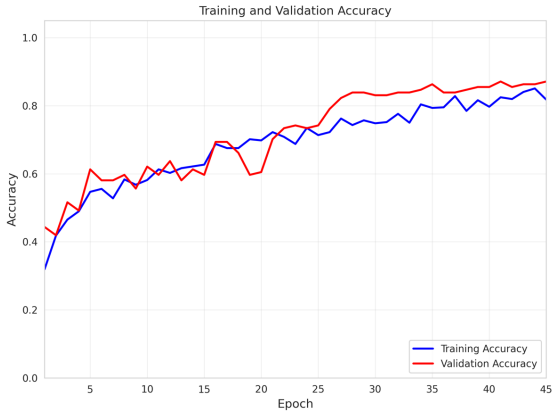


Figure 4: Training and Validation accuracy

Figure 4 illustrates the training and validation accuracy across 45 epochs for lung disease classification, highlighting the two-stage optimization dynamics. In Stage 1 (Epochs 1–20), validation accuracy plateaued at 69.35%, constrained by generic ImageNet representations. Selective unfreezing of the top 30% convolutional layers at Epoch 21 catalyzed a sharp performance gain steadily driving validation accuracy to a peak of 87.10%. The harmonious convergence between training and validation trajectories confirms robust generalization without empirical overfitting.

Table 1: Classification Report

	precision	Recall	f1-score	support
Normal	1	0.85	0.92	7
Large Cell Carcinoma	0.96	1	0.98	28
Adenocarcinoma	0.82	0.90	0.85	51
Squamous Cell Carcinoma	0.87	0.76	0.81	38
accuracy	0.87	0.87	0.87	0.87
macro avg	0.91	0.88	0.89	124

As shown in above table 1 the overall classification accuracy on the test-set for 124 unseen CTs is 87% and the macro-average F1 score is 0.89. The model reached perfect precision (1.00) for healthy Normal cases, without giving false-positive diagnoses. Large Cell Carcinoma had a perfect diagnostic sensitivity (F1-score: 1.00) and full recall (complete score: 1.00). Although the overall morphology of the two cohorts (Majority Adenocarcinoma (F1: 0.85) and Squamous Cell Carcinoma (F1: 0.81)) was similar the framework maintained a balanced predictive reliability, even in imbalanced classes of malignancy.

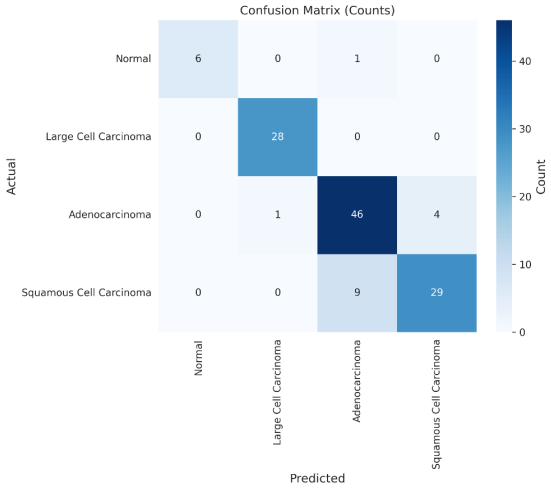


Figure 5: Confusion Matrix

A confusion matrix has been created on the test-set to provide a breakdown of classifications on 124 CT scans as shown in Figure 5. Based on the proposed model there was perfect identification of LCC and all 28 cases were correctly classified. Specifically for clinical screening there were absolutely no cancerous scans incorrectly classified as Normal (6/6). The diagnostic errors were mainly between Adenocarcinoma and Squamous Cell Carcinoma with 9 squamous cases diagnosed as Adenocarcinoma and 4 Adenocarcinoma cases diagnosed as Squamous. This is due to the fact that non-small cell lung cancer (NSCLC) subtypes have certain subtle similarities in terms of their radiologic and morphologic features.

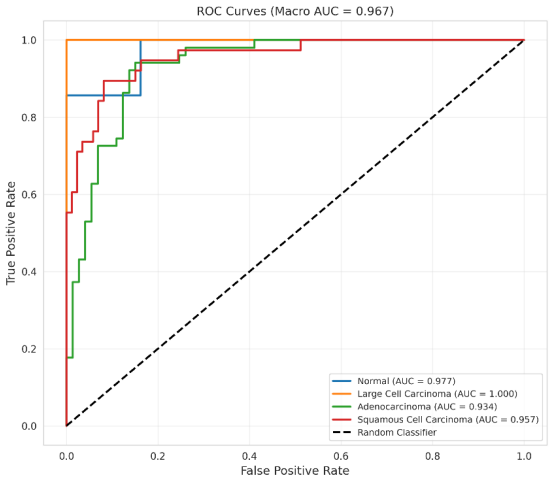


Figure 6: ROC Curve

The multi-class Receiver Operating Characteristic (ROC) curves are shown in Figure 6 which shows excellent discriminative capability with an aggregate Macro AUC of 0.967. The results of the

confusion matrix matched Large Cell Carcinoma as it had an AUC of 1.000 which was the best separation. The AUC of the Normal class was 0.977 whereas the AUC of the Squamous Cell Carcinoma and Adenocarcinoma classes were 0.957 and 0.934 respectively which were reasonably high. These high area-under-the-curve characteristics further validate the proposed model to have a high diagnostic sensitivity over different decision thresholds thereby overcoming histological overlap between subtle carcinoma variants of the lungs.

The hyperparameters of the proposed lung cancer classification model (EfficientNetV2-S–CBAM) for training and fine-tuning were chosen for stability in optimization and effective adaptation to the target dataset. Input images to the model were 224 × 224 × 3 and the model was trained on a four-class classification. Focal Loss is used as the optimization objective and class weighting is used to deal with class imbalance. A two-stage transfer-learning approach was used for training. For stage 1 the backbone of the network (the part trained from ImageNet) was kept frozen and only the additional classification layers were trained for 20 epochs using a learning rate of 0.001. For Stage 2 154 backbone layers were unfrozen for fine-tuning and the Batch Normalization layers were frozen and the learning rate was changed to 0.00001 with 25 further epochs of training. The classification head consisted of fully connected layers with 256 and 128 neurons respectively with four output neurons for the four target classes. Feature aggregation was performed by using both Global Average Pooling and Global Max Pooling. The model checkpoint with the best validation accuracy was kept for the final evaluation. Test-Time Augmentation was introduced during the inference stage to enhance robustness of the final predictions. The final setting of hyper parameters for training the proposed model was these settings.

Table 2: Comparative Analysis with the Proposed Work

Technology Used	**Main Techniques Used**	**Strengths**	**Limitations**
Conventional Machine Learning [5], [15]	Logistic Regression, Random Forest, KNN, SVM, XGBoost, CatBoost, AdaBoost and RBF-based classifiers	These methods are relatively simple to train and require less computational power. They are particularly useful when the input information is available as structured clinical or biomarker data.	Their effectiveness depends heavily on the quality of manually selected features. They are not well suited to learning complex spatial patterns directly from medical images.
Handcrafted Feature-Based Classification [9]	GLCM texture features combined with KNN, SVM, XGBoost and deep features from ResNet-based models	Texture-based features can capture useful variations in lung tissue and provide a direct representation of selected image characteristics.	The performance can vary with the choice of handcrafted features. Such features may not adequately describe subtle and complex patterns associated with different lung cancer subtypes.
CNN-Based Image Classification [4], [18]	CNN and VGG-based architectures with image resizing, normalization, augmentation and spatial transformation	CNNs can automatically learn hierarchical visual features from images and therefore reduce the need for manual feature extraction.	Standard CNN architectures may not sufficiently distinguish the most informative regions or feature channels. Most of these approaches also concentrate on chest X-ray-based lung disease classification rather

			than multi-class CT-based lung cancer classification.
Transfer Learning-Based Methodology [3], [7], [14]	Pretrained Inception V3, ResNet-based networks and other ImageNet-initialized CNN models followed by fine-tuning	Pretrained networks provide useful initial representations and can reduce training time, which is particularly helpful when the available medical dataset is limited.	Features learned from natural-image datasets may not completely represent disease-specific characteristics in CT images. The final performance also depends on the architecture and fine-tuning strategy
Attention-Based Deep Learning [3], [18]	CAM-based localization, spatial transformation and attention-oriented feature enhancement	Attention mechanisms allow the network to give greater importance to informative image regions and can improve the representation of disease-related features.	Existing approaches generally use different imaging modalities or disease categories. In addition, some methods emphasize spatial information without jointly modelling both channel-level and spatial feature importance.
Explainable Deep Learning [1], [3]	Grad-CAM, CAM and visualization-based interpretation of model predictions	These techniques provide visual information about the regions contributing to a prediction, making deep-learning decisions easier to inspect.	The explanation is usually provided after or alongside classification and does not directly improve the learned feature representation
Proposed Efficient CNN with Attention	EfficientNetV2-S, CBAM, channel and spatial attention, GAP, GMP, feature concatenation, Focal Loss, class weighting, two-stage transfer learning and test-time augmentation	The proposed framework combines an efficient pretrained CNN with CBAM to refine both channel-wise and spatial information. GAP and GMP provide complementary global representations, while focal loss and class weighting help address class imbalance. Two-stage fine-tuning allows the pretrained network to gradually adapt to	The study is based on a relatively small curated dataset, so additional validation using independent datasets from different institutions is required. Differences in scanners, imaging protocols and patient populations may also affect generalization.

		lung CT characteristics.	

Conclusion:

This study presented an EfficientNetV2-S–CBAM framework for classifying lung CT images into Adenocarcinoma, Large Cell Carcinoma, Squamous Cell Carcinoma, and Normal categories. The proposed approach combines duplicate removal, stratified data partitioning, attention-based feature refinement, focal loss and two-stage transfer learning to improve classification reliability on a relatively small and imbalanced dataset. The model achieved 87.90% test accuracy, 89.56% macro F1-score and 96.68% macro ROC-AUC with particularly strong performance for Large Cell Carcinoma and the Normal class. The study is limited by the size and single-source nature of the dataset. Future work should evaluate the framework on larger, multi-centre datasets and investigate external validation, explain ability and multimodal clinical information before clinical application.

References

1. Zahin, I. A., Ahsan, M. F., Orni, R. A., Hossain, A., Tabassum, M., Zishan, M. A. O., & Noor, J. (2026). Enhancing lung diseases recognition through CNN-RNN methodologies. *Scientific Reports*, Article number (2026).
2. Bhuvaneswari, C. A., Allaberganova, M., Khudaykulova, F., Khajiqurbonova, N., Mamajonova, S., & Mavlonov, A. (2026). Deep learning based early-stage restrictive lung disease prediction. In *2026 6th International Conference on Image Processing and Capsule Networks (ICIPCN)* (pp. 1–6). IEEE. https://doi.org/10.1109/ICIPCN67432.2026.11438359
3. Hroub, N. A., Alsannaa, A. N., Alowaifeer, M., Alfarraj, M., & Okafor, E. (2024). Explainable deep learning diagnostic system for prediction of lung disease from medical images. *Computers in Biology and Medicine, 170*, 108012.
4. Sharma, R. R., Venkatesh, K. A., Naikwade, I., Sharma, H., Sungheetha, A., & Ghatasala, G. S. P. (2025). Predicting lung diseases using machine learning on X-ray images. *Procedia Computer Science, 259*, 1432–1441.
5. V. Suvarchala, V. Subbareddy, and S. Rao Madala, "Lung Cancer Prediction Using Machine Learning Methodologies," Volatiles Essent. Oils, vol. 8, no. 6, pp. 1265–1272, 2021.
6. Bharati S.,Podder P., Mondal R., Mahmood A., Raihan Al Masud M. (2020) "Comparative performance analysis of different classification algorithms for the purpose of prediction of lung cancer" Advances in intelligent systems and computing Vol. 941, pp. 447–457, Springer.
7. Coudray N., Ocampo P. S., Sakellariou T., et al. (2018) "Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning" Nature Medicine*, 24*(10), 1559–1567.
8. N. Numan, S. Abuelenin, and M. Rashad, "Prediction of lungs cancer using artificial neural network ," International Journal of Intelligent Computing and Information Sciences", vol. 16, no. 2, pp. 1–19, Apr. 2018
9. Ahmed S., Raza B., Hussain L., Aldweesh A., Omar A. (2023) "The Deep Learning ResNet101 and Ensemble XGBoost Algorithm with Hyper parameters Optimization Accurately Predict the Lung Cancer", Applied Artificial Intelligence, *37*(1).
10. Sharmila Nageswaran, G Arunkumar, Anil Kumar Bisht, Shivlal Mewada, J N V R Swarup Kumar, Malik Jawarneh, Evans Asenso "Lung Cancer Classification and Prediction Using Machine Learning and Image Processing" ,BioMed Research International, vol. 2022, 2022.
11. Yossra Hussain Ali, V. S. Chooralil, K. Balasubramanian, R. R. Manyam, S. K. Raju, A. T Sadiq "Optimization System Based on Convolutional Neural Network and Internet of Medical Things for Early Diagnosis of Lung Cancer," *Bioengineering*, vol. 10, no. 3, 2023.
12. Patel A. N., Srinivasan K. (2025) "Deep learning paradigms in lung cancer diagnosis: A methodological review, open challenges, and future directions" Physica Medica*, 131*, 104914.
13. Nahiduzzaman M., Abdulrazak L. F., Ayari M. A., Khandakar A., Islam S. M. R. (2024) "A novel framework for lung cancer classification using lightweight convolutional neural networks and ridge extreme learning machine model with SHapley Additive exPlanations (SHAP)" Expert Systems with Applications*, 248*, 123392.
14. C. Venkatesh, K. Ramana, S. Y. Lakkisetty, S. S. Band, S. Agarwal, and A. Mosavi, "A Neural Network and Optimization Based Lung Cancer Detection System in CT Images," Front. Public Heal., vol. 10 .no. June.
15. Jianyu Li, Aiping Chen, Zhiping Liu, Shaozhong Wei, Jing Zhang, Jianxin Chen,

Chenghe Shi (2025) *"*Machine learning-driven prediction of drug efficacy in lung cancer: Based on protein biomarkers and clinical features*"* Life Sciences, 375*, 123706.*

16. G. Divya Deepak, Subraya Krishna Bhat (2025) "A multi-stage deep learning approach for comprehensive lung disease classification from x-ray images" Expert Systems With Applications , 277, 127220.

17. Durgam R., Panduri B., Balaji, V., Adil O. K, Alaa O. K. & Shitharth S. (2025) "Enhancing lung cancer detection through integrated deep learning and transformer models" Scientific Reports 15, 15614.

18. Subrato Bharati , Prajoy Podder, M. Rubaiyat Hossain Mondal (2020) "Hybrid deep learning for detecting lung diseases from X-ray images" Informatics in Medicine Unlocked, 20, 100391.