**ISTP**

RESEARCH ARTICLE

# Enhancing Lung Cancer Detection: Optimizing Deep Learning With Convolutional Block Attention Module

**Sriramganesh Gokavarapu,**[1,2] **K Venkata Rao,**[1] **Gorla Srinivas,**[3] **and Dasari Siva Krishna**[4]

[1]Computer Science and Systems Engineering, Andhra University, Visakhapatnam, India

[2]Department of Computer Science and Engineering, Sri Vasavi Engineering College (A), West Godavari, India

[3]Department of Computer Science and Engineering, Anil Neerukonda Institute of Technology & Sciences, Visakhapatnam, India

[4]Department of Computer Science and Engineering, Gandhi Institute of Technology and Management, Visakhapatnam, India

*Abstract*: Lung cancer manifests as the uninhibited propagation of cells within tissues in lung. Early exposure of malignant cells within the lungs, which play crucial roles in oxygenation and carbon dioxide exchange vital for human physiology, is imperative. Deep learning in detecting lung cancer has gained considerable attention because of its impending to enhance patient identification of diagnosis and treatment. Nonetheless, existing algorithms demonstrate less-than-ideal performance in crucial areas such as recognition accuracy, precision, sensitivity, F-score, and specificity. To address this issue, a novel method Unet with CBAM is leveraged by the different deep learning pre trained models such as VGG-16, VGG-19, ResNet-50, Resnet-101 and EfficientNet. In this proposed methodology we introduce Unet with Convolutional Block Attention Module (CBAM) for the prediction of lung cancer. The proposed model has a **base learner** consists of two modules: Unet and CBAM attention. In the attention mechanism there are also two modules: channel attention module and spatial attention modules. The features of each module are integrated and supplied to the **meta-learner** as a fully connected model for the precision. This model improves results comparison with other models.

## I. INTRODUCTION

Cancer is a life-threatening disease with potentially severe consequences. Among the most prevalent types of cancer worldwide, lung cancer remains a leading cause of cancer-related deaths in both developed and developing countries. Non–Small Cell Lung Cancer (NSCLC) accounts for the majority of these cases, with a poor five-year survival rate of only 18%. Despite significant advancements in medical technology that have improved overall cancer survival rates, progress in lung cancer treatment remains slow. This is primarily due to most cases being diagnosed at a late stage. The spread of cancer cells from the lungs to the lymph nodes and eventually into the bloodstream, known as metastasis, plays a critical role in lung cancer progression [1]. Cancer cells typically migrate toward the mediastinum due to natural lymphatic flow. Early detection is crucial in preventing metastasis from spreading to other organs. Regular radiographic follow-ups are essential for monitoring therapeutic responses and tracking changes in tumor characteristics over time.

The role of the primary tumor in lung cancer is similar to that in breast cancer, influencing staging, treatment options, and patient outcomes. Accurate lymph node examination is essential for proper staging and prognosis determination. Traditionally, pathologists have relied on manually reviewing glass slides under microscopes, a process that is both time-consuming and labor-intensive. However, advancements in digital pathology have transformed this practice through slide scanning technology, enabling the digitization of classical pathology. This shift allows for computer-assisted image analysis, reducing manual effort and improving diagnostic efficiency.

Deep learning (DL) approaches, particularly Convolutional Neural Networks (CNNs) [2], such as VGG-19 [8,9], have shown great potential in medical disease diagnosis, including pathology. CNNs have been successfully employed to detect tumor cells across various types of cancer. Accurate assessment of lymph nodes is crucial for classifying cancer and determining the most appropriate therapeutic interventions. Recent studies have demonstrated that histopathological images [3] serve as reliable predictors of therapeutic biomarkers. However, manual screening of multiple slides remains cumbersome and prone to human error. AI-driven solutions have exhibited superior efficiency in detecting micrometastases compared to pathologists working under high workloads, offering promising advancements in cancer diagnostics.

The workflow of the proposed methodology as follows

- Introducing the Unet architectures with Convolutional Block Attention Module (CBAM) attention mechanism. The CBAM which consists of two modules channel attention(CA) and spatial attention (SA).

- In the CA module, global average pooling is followed by two fully connected layer for feature extraction.

- The SA module generates spatial attention maps by processing feature maps through convolutional layers.

Corresponding author: Sriramganesh Gokavarapu (e-mail: sriramganesh8107@gmail.com).

- Combine attention maps with original feature maps to generate refined representations.

- Evaluate and compare the proposed model's performance against other models.

The article is structured as follows: Section II reviews related works, while Section III describes the dataset used for training the proposed model and also discusses various methodologies, including VGG, ResNet, and EfficientNet. Section IV introduces the proposed Unet architecture with CBAM attention. Section V presents experimental results and comparative analysis. Finally, the article concludes with Section VI.

## II. RELATED WORK

Lung cancer, characterized by the uncontrolled proliferation of malignant cells within lung tissues, poses a significant global health challenge, contributing to rising mortality rates. While complete eradication remains unattainable, efforts to mitigate its impact are crucial [1,7]. The incidence of lung cancer is strongly correlated with persistent smoking habits. Various classification methodologies, including machine learning (ML) models, have been explored to improve lung cancer diagnosis and treatment. Younis *et al.* [5] proposed preprocessing strategies using Unet and ResNet algorithms to emphasize lung regions. Extracted features were classified using methods like AdaBoost and Random Forest, and ensemble predictions were employed to estimate the likelihood of cancer presence in CT scans. In another study, Shin *et al.* [6] utilized deep learning methodologies to analyze the characteristics of cell exosomes, achieving an impressive 95% classification accuracy. Their deep learning algorithm revealed a high similarity between plasma exosomes from patients and extracellular vesicles from lung cancer cells, suggesting potential diagnostic applications. Another study explored the use of clinical factors and radiomic markers to predict lung ADC subtypes using LIFEx software. Additionally, a deep learning network incorporating tumor cell and metastatic staging systems was developed to estimate survival rates and provide personalized therapy recommendations through a user-friendly interface. Furthermore, abnormal lung detection techniques, leveraging segmentation-based ML approaches, have significantly improved performance, particularly when employing Support Vector Machines (SVMs). Ensemble ML models have demonstrated strong predictive power, with performance metrics such as an RMSE of 0.64, recall of 0.97, and an F1-score of 0.96. Recent literature also highlights the effectiveness of ensemble models, including stacking approaches [10], which utilize different embedding techniques for feature extraction.

Medical image segmentation has seen significant advancements with deep learning models and ensemble techniques, particularly in disease detection and classification. Various Unet-based architectures, CNNs, and hybrid networks have been developed to tackle challenges in medical image analysis, including precise tumor detection, disease segmentation, and feature extraction. Meng *et al.* [12] introduced AFC-Unet, a CNN-transformer model incorporating full-scale attention mechanisms to enhance segmentation accuracy. This model effectively integrates convolutional and transformer components to capture both local and global image features. Similarly, Wan *et al.* [13] proposed UMF-Net, a multi-branch Unet architecture designed for colon polyp segmentation, employing feature fusion techniques to improve performance by combining features from multiple scales. These approaches emphasize the importance of attention

mechanisms and feature fusion in addressing complex medical segmentation tasks, ultimately improving accuracy in detecting and segmenting abnormalities.

Yaohong Suo *et al.* [11] developed a novel approach for diabetic retinopathy (DR) classification using two modules: the Full Convolution Spatial Attention Module (FCSAM) and CS-ResNet. Additionally, they introduced an image enhancement algorithm to improve image quality, effectively reducing overfitting. Their system achieved an impressive accuracy of 98.1%. Wan *et al.* [13] also proposed a segmentation-based Unet architecture for polyp segmentation, incorporating multi-fusion techniques to combine attention-based features across different scales. Similarly, Hussain *et al.* [14] introduced the MAGRes-Unet model, which integrates multiple attention mechanisms with residual connections. This approach employs diverse activation functions and optimizers to enhance feature extraction. Ortega-Ruíz *et al.* [15] developed DRD-Net for breast cancer segmentation, incorporating dilation, residual, and dense connections to refine imaging features. Meanwhile, Iriawan *et al.* [16] introduced YOLO-Unet for brain tumor detection, combining the YOLO detection framework with Unet to improve segmentation accuracy in localized tumor regions. This architecture effectively addresses challenges in tumor localization within MRI scans, demonstrating the advantages of integrating object detection with segmentation in medical imaging.

Li *et al.* [17] contributed to deep-space rock image analysis using the Diamond-Unet, which combines global and local feature extraction with CNN-based cross-fusion paths. Wang *et al.* [18] implemented attention-based techniques in the Attention-Dual Unet, designed for fusing infrared and visible images, with potential applications in both medical and industrial domains. AlArfaj *et al.* [19] developed a segmentation-based model for detecting pepper bell leaf diseases, employing a transfer learning mechanism with a Unet-based architecture. Their approach was compared with other transfer learning models, such as InceptionV3, demonstrating enhanced performance. Abuhayi, Bezabh, and Ayalew *et al.* [20] introduced INVGG, an involution-based VGG network for lumbar disease classification, while Rehman and Gruhn *et al.* [21] designed an adaptive VGG16-based CNN for knee osteoarthritis stage detection. This model integrated fine-tuned pre-trained VGG19 and CNN features to improve classification performance. Similarly, Sahoo *et al.* [22] developed an enhanced VGG-19 model to optimize feature pooling, improving the detection of dynamic objects in complex video scenes. This approach is beneficial for medical and non-medical applications requiring high-precision detection in intricate environments.

In lung cancer research, Ardila *et al.* [24] demonstrated the effectiveness of a 3D CNN-based approach for lung cancer screening using low-dose CT scans, setting a benchmark for automated screening in oncology. Kareem *et al.* [25] explored Support Vector Machines (SVMs) for lung cancer detection, highlighting the relevance of non-deep learning approaches in image-based diagnostics. Rajasekar *et al.* [23,26] applied deep learning techniques to predict lung cancer using CT scans and histopathological images, emphasizing the benefits of multi-modal data integration. Comparative studies by Radhika *et al.* [27] and Pradhan and Chawla *et al.* [28] have analyzed various machine learning algorithms for lung cancer detection, providing insights into the evolution from traditional ML techniques to advanced deep learning architectures in medical diagnostics. These studies underscore the shift towards sophisticated neural network-based solutions, significantly improving diagnostic accuracy and efficiency in lung cancer detection.

# III. DATASET AND METHODOLOGY

This section presents the dataset used in this study and the methodology employed. It includes the use of pre-trained models such as VGG-16, VGG-19, ResNet-50, ResNet-101, and EfficientNet, all trained on the dataset. The study leverages these models to analyze their effectiveness in achieving the desired outcomes. The dataset serves as the foundation for training and evaluating these architectures, ensuring a comprehensive comparison of their performance. By utilizing well-established deep learning models, this approach enhances the reliability of the results. The methodology focuses on optimizing model performance while maintaining consistency across different architectures for a thorough evaluation.

## A. DATASET DESCRIPTION

The Iraq-Oncology Teaching Hospital/National Center for Cancer Diseases (IQ-OTH/NCCD) lung cancer dataset was collected in the above-mentioned specialist hospitals over a period of three months in fall 2019. It includes CT scans of patients diagnosed with lung cancer in different stages, as well as healthy subjects. IQ-OTH/NCCD slides were marked by oncologists and radiologists in these two centers. This dataset features CT scans from patients, as well as scans from healthy individuals. In total, the dataset comprises 1,097 CT slices from 110 cases, divided into three categories: normal, benign, and malignant. Specifically, 561 scans are classified as malignant, 120 as benign, and 416 as normal. The dataset can be accessed through the dataset [4], and sample images for each category are provided in Fig. 1.

## B. METHODOLOGY (PRE-TRAINED MODELS)

In this section, we conduct experiments using various deep learning architectures, including VGG, ResNet, and EfficientNet. These models are selected based on existing literature to facilitate a comprehensive comparison with the proposed method. VGG is a well-established deep CNN architecture known for its effectiveness in various computer vision tasks, particularly image classification [17,18]. The VGG-16 model consists of multiple convolutional layers followed by fully connected layers, forming a structured design. It comprises five convolutional blocks, each containing multiple convolutional layers, followed by max-pooling layers, which optimize feature extraction and representation. ResNet-50, originally designed for large-scale image classification, is adapted for lung cancer classification in this study [19]. The key

innovation in ResNet is the introduction of residual blocks, which help mitigate the vanishing gradient problem and facilitate efficient training of deep networks. Mathematically, a residual block can be defined as follows: given an input X[l], the transformation within the block can be expressed as $X[l+1] = X[l] + F(X[l], W[l])$, where F represents the residual function, and W[l] denotes the learned weights. The fully connected layers at the end of the ResNet-50 model further refine the extracted features for accurate classification.

EfficientNet is designed to achieve high accuracy while maintaining computational efficiency [29–31]. It addresses the challenge of network scaling by balancing three key dimensions: depth, width, and resolution. Deeper networks capture more complex features but require more computation, while wider networks process more information in parallel but demand additional memory. Higher input resolution provides finer details but increases computational cost. EfficientNet incorporates MBConv (Mobile Inverted Bottleneck Convolution) layers, which enhance feature representation and improve model performance. This architectural design is particularly beneficial for medical applications, such as lung cancer detection, where achieving high accuracy with optimized computational resources is crucial.

# IV. PROPOSED METHODOLOGY

The proposed model as shown in Fig. 2, takes resized images of input images of size $3 \times 224 \times 224$, which means it's designed for RGB images with a $224 \times 224$ pixel resolution. The network is structured with multiple stages of convolutional layers (CL), each augmented with CBAM to enhance feature extraction by directing attention to important areas. Each layer is followed by normalization and sigmoid activation function, which helps improve the network's learning capabilities and overall stability.

The input consists of an RGB image of size $224 \times 224$ pixels. A CL with 32 filters is applied, followed by ReLU activation and batch normalization. Output Feature Map reduces the image dimensions to $32 \times 112 \times 112$, indicating a downsampling operation, likely from stride or pooling. The network progresses through multiple blocks that involve sequential convolutions, with CBAM modules incorporated after specific layers to refine feature representations by emphasizing important information.

The proposed methodology extracts the feature from the various stages such as in Unet based model. Each stage in the model employs the feature extracted and combines them with
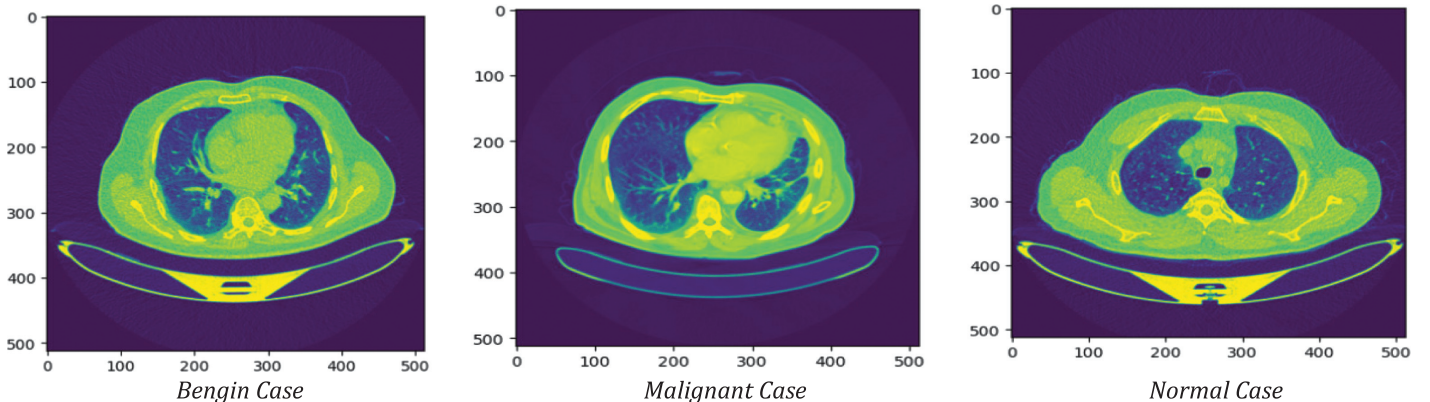


*Bengin Case*      *Malignant Case*      *Normal Case*

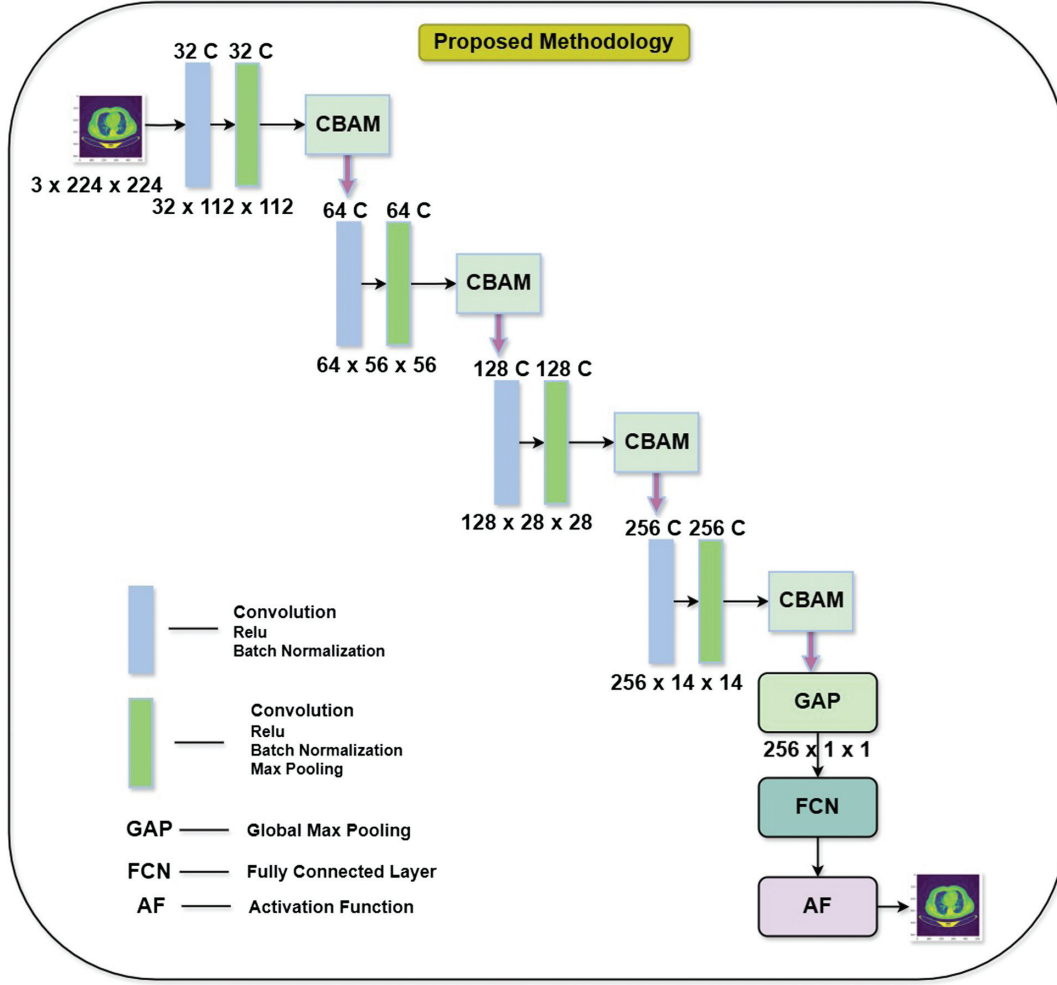**Fig. 1.** Sample Images of dataset.

**Fig. 2.** Unet with CBAM architecture.

the attention mechanism. Here CBAM based attention is used at every stage, which consists of attention at two levels. At the first level, it is of channel attention, and the next level is of the spatial attention throughout at every stage of the Unet architecture. The architecture also employs efficient downsampling through a combination of convolution and pooling layers, which gradually reduces the spatial dimensions of feature maps while preserving essential hierarchical structures. Additionally, after each convolution step before normalization, helping to stabilize the training process, speed up convergence, and reduce internal covariate.

## A. CBAM ATTENTION MECHANISM

Convolutional Block Attention Module (CBAM) is a powerful tool for enhancing the performance of CNNs tasks, including the prediction of lung cancer. CBAM integrates [7] both spatial and channel-wise attention mechanisms to adaptively highlight informative features within the input data. The CBAM module consists of two attention blocks: the CA block and the SA block as shown in Fig. 3.

***1. CHANNEL ATTENTION (CA) BLOCK.*** This block captures the interdependencies between feature channels by calculating channel-wise attention maps. It computes the importance of each feature

channel by utilizing global average pooling and max-pooling operations followed by a series of FC layers. Attention weights obtained from the CA block are then applied to each feature map across all spatial locations.

Let $X$ be the feature map of input size $C \times H \times W$, where $C$ #channels, $H$- height, and $W$-width. In output of the CA block is computed as follows: First, the average and max pooling operations are applied along spatial dimensions to obtain global average and max pooled feature vectors, respectively in equation (1–4):

$$M_{avg} = AvgPool(X) \tag{1}$$

$$M_{max} = MaxPool(X) \tag{2}$$

$$A_{avg} = Sigmoid(W_{avg}.(FC_{avg}(M_{avg}))) \tag{3}$$

$$A_{max} = Sigmoid(W_{max}.(FC_{max}(M_{max}))) \tag{4}$$

Here, $W_{avg}$ and $W_{max}$ are the learnable weights of the fully connected layers, and $FC_{avg}$ and $FC_{max}$ represent the fully connected layers. Finally, the channel-wise attention maps are combined using element-wise addition or multiplication to produce the output feature map as defined in equation (5):
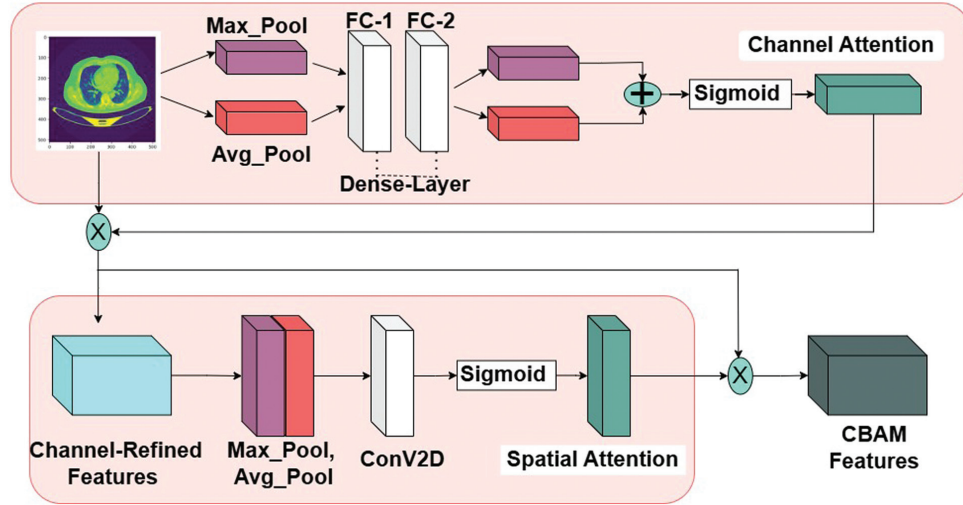
**Fig. 3.**  Convolutional Block Attention module (CBAM) architecture.

$$Y_{CA} = X \odot (A_{avg} + A_{max}) \tag{5}$$

***2. SPATIAL ATTENTION (SA) BLOCK.***    Unlike the CA block, the SA block focuses on capturing spatial dependencies within feature maps. It generates spatial attention maps by considering correlations between neighboring spatial locations. This is achieved through a combination of max pooling and average pooling operations along both spatial dimensions, followed by fully connected layers. The resulting attention weights highlight relevant spatial regions.

By combining the outputs of the CA and SA blocks through element-wise addition or multiplication, the CBAM module produces enhanced feature maps that emphasize both informative channels and spatial regions. Similarly, let $X$ be the input as $C \times H \times W$, The output of the SA block is computed as follows: First, the average and max pooling operations are applied along channel dimensions to obtain channel-wise average and max-pooled feature maps, respectively as defined in equation (6–8):

$$M_{avg} = AvgPool(X) \tag{6}$$

$$M_{max} = MaxPool(X) \tag{7}$$

$$M_{cat} = [M_{avg}, M_{max}] \tag{8}$$

Next, the concatenated feature map is processed through a convolutional layer and then a sigmoid activation function to generate the spatial attention map as defined in equation (9):

$$A_{spatial} = Sigmoid(M_{Conv} * (Conv(M_{cat}))) \tag{9}$$

Here, *Wcon* represents the learnable weights of the convolutional layer. Finally, the spatial attention map is applied to the input feature map using element-wise multiplication to produce the output feature map as defined in equation (10):

$$Y_{SA} = X \odot A_{spatial} \tag{10}$$

## V.  RESULTS AND DISCUSSION

In this Section, the detailed experimental results are explored in the Tables I–V. The dataset with experiments results of VGG-16, VGG-19, ResNet-50 and EfficientNet are explored with hyper parameters such as learning late, epochs and performance metrics. The results of proposed method such Unet with CBAM are perform better than other existing methods.

**Table I.**    Performance of VGG-16 and VGG-19 models

| Model | VGG-16 | | | | | VGG-19 | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Parameter | Learning Rate(L) = 0.0001 | | | | | Learning Rate(L) = 0.0001 | | | | |
| Epochs | 2 | 4 | 6 | 8 | 10 | 2 | 4 | 6 | 8 | 10 |
| Accuracy | 94.26 | 90.22 | 92.48 | **92.80** | 92.40 | 88.48 | 82.40 | 86.40 | **90.40** | 86.20 |
| Precision | 92.40 | 90.20 | 93.40 | **94.60** | 90.10 | 86.40 | 82.20 | 83.20 | **91.20** | 82.40 |
| Recall | 92.10 | 89.48 | 90.10 | **94.80** | 90.20 | 86.60 | 87.20 | 82.40 | **90.40** | 92.30 |
| F1-Score | 92.25 | 89.84 | 91.75 | **94.70** | 90.15 | 86.50 | 84.70 | 82.80 | **90.80** | 87.35 |

**Discussion:** Two models, VGG-16 and VGG-19, were compared across different numbers of epochs. Both models were trained with L of 0.0001. VGG-16 generally outperformed VGG-19 across all metrics and epochs. The highest accuracy achieved by VGG-16 was 94.80% at 8 epochs, while for VGG-19 it was 90.40% at 8 epochs. Metrics consistently higher for VGG-16 compared to VGG-19. Performance generally improved with increasing epochs, but declined after a certain point, indicating the possibility of overfitting. Optimal performance varied for each model, with different epochs yielding the best results.

**Table II.**    Performance of ResNet-50 and ResNet-101 models

| Model | ResNet-50 | | | | | ResNet-101 | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Parameter | Learning Rate(L) = 0.0001 | | | | | Learning Rate(L) = 0.0001 | | | | |
| Epochs | 2 | 4 | 6 | 8 | 10 | 2 | 4 | 6 | 8 | 10 |
| Accuracy | 86.48 | 84.20 | 80.30 | **88.20** | 85.30 | **88.40** | 86.20 | 84.30 | 84.20 | 82.40 |
| Precision | 84.50 | 86.80 | 88.40 | **88.20** | 84.40 | **90.20** | 94.40 | 96.60 | 92.20 | 90.50 |
| Recall | 84.20 | 84.50 | 86.40 | **88.80** | 84.20 | **88.40** | 88.20 | 90.40 | 88.60 | 90.30 |
| F1-Score | 84.35 | 85.65 | 87.40 | **88.50** | 84.3 | **89.30** | 91.30 | 93.50 | 90.40 | 90.40 |

**Discussion:** Compared ResNet-50 and ResNet-101 models with L of 0.0001 across various epochs. The evaluated performance metrics with ResNet-101 generally outperformed ResNet-50, achieving higher accuracy and more stable performance across epochs. ResNet-101's maximum accuracy was 88.40% at 2 epochs, with precision, recall, and F1-score ranging from 90.20% to 93.50%. ResNet-50 achieved a maximum accuracy of 88.20% at 8 epochs, with precision, recall, and F1-score ranging from 84.50% to 88.50%. Overall, ResNet-101 demonstrated slightly better performance and stability, making it potentially preferable for certain applications.

**Table III.**    Performance of EfficientNet model

| Model | EfficientNet | | | | |
|---|---|---|---|---|---|
| Parameter | Learning Rate(L) = 0.0001 | | | | |
| Epochs | 2 | 4 | 6 | 8 | 10 |
| Accuracy | 94.60 | 94.50 | 94.40 | 94.60 | **95.20** |
| Precision | 92.20 | 94.80 | 94.60 | 94.30 | **94.80** |
| Recall | 94.60 | 92.60 | 92.40 | 93.20 | **95.30** |
| F1-Score | 93.40 | 93.70 | 93.50 | 94.25 | **95.50** |

**Discussion:** the EfficientNet model performs best around 10 epochs, achieving the highest accuracy (95.20%) and F1-score (94.80%). Early training (2–4 epochs) improves precision and F1-score but causes inconsistencies in accuracy and recall. Mid-training (6–8 epochs) shows the most significant improvements across all metrics, marking it as the optimal training period. Extended training to 10 epochs may cause overfitting, evidenced by minor drops in accuracy and precision, although recall and F1-score remain strong. Therefore, monitoring and adjusting the training duration is crucial to maintain optimal model performance.

**Table IV.**    Performance of proposed method

| Model | Unet with CBAM(Proposed Method) | | | | |
|---|---|---|---|---|---|
| Parameter | Learning Rate(L) = 0.0001 | | | | |
| Epochs | 2 | 4 | 6 | 8 | 10 |
| Accuracy | 96.40 | 96.20 | 96.20 | 94.80 | **96.80** |
| Precision | 94.60 | 94.80 | 96.40 | 94.60 | **96.40** |
| Recall | 92.40 | 95.30 | 94.20 | 95.20 | **96.60** |
| F1-Score | 93.50 | 95.50 | 95.30 | 94.50 | **96.50** |

**Discussion:** The performance of the Unet with CBAM model, L of 0.0001 across various epochs (2, 4, 6, 8, and 10), reveals significant insights. Initially, the model performs well with an accuracy of 96.40% at 2 epochs, which slightly decreases to 96.20% at 4 epochs, indicating minor variability. Precision improves from 94.60% to 94.80%, and recall increases significantly from 92.40% to 95.30%, resulting in an improved F1-score from 93.50% to 95.50%. In the mid-training phase, the model shows marked improvements, with accuracy jumping to 96.20% at 6 epochs and further to 96.80% at 8 epochs. Precision and recall also peak at 96.40% and 96.60%, in that order, leading to an F1-score of 96.50%. Extended training to 10 epochs maintains these high levels of performance without signs of overfitting, as all metrics remain stable. Overall, training the Unet with CBAMmodel for 8 to 10 epochs yields the best results, achieving an optimal balance of metrics, and maintaining robust performance across extended training.

## A. COMPARISON STUDY

The comparison of various DL models with the proposed Unet with CBAM model highlights the superior performance of the CBAM approach. VGG-16 and VGG-19 models show strong precision

**Table V.**    Comparison of various deep learning models with proposed model

| Model | Accuracy | Precision | Recall | F1-Score |
|---|---|---|---|---|
| VGG-16 | 92.80 | 94.60 | 94.80 | 94.70 |
| VGG-19 | 90.40 | 91.20 | 90.40 | 90.80 |
| Res-50 | 88.20 | 88.20 | 88.80 | 88.50 |
| ResNet-101 | 88.40 | 90.20 | 88.40 | 89.30 |
| EfficientNet | 95.20 | 94.80 | 95.30 | 95.50 |
| **Proposed Method (Unet with CBAM)** | **96.80** | **96.40** | **96.60** | **96.50** |

(94.60% and 90.40%, respectively) but have lower recall and F1-scores. ResNet-50 and ResNet-101 demonstrate even lower performance metrics, with ResNet-50 having an accuracy of 88.20% and ResNet-101 slightly better at 88.40%. EfficientNet performs notably well, achieving 95.20% accuracy and high scores in precision, recall, and F1-score (94.80%, 95.30%, and 95.50%, respectively). However, the proposed Unet with CBAM model surpasses all these models with an impressive accuracy, precision, recall and F1-score 96.80%, 96.40%, 96.60%, 96.50% showcasing its effectiveness.

## VI. CONCLUSION

In this study, the early detection of lung cancer, a critical aspect of patient diagnosis and treatment, was significantly enhanced by the proposed methodology incorporating Unet with CBAM. By integrating channel and spatial attention modules, the proposed approach effectively addressed the limitations of existing deep learning models, such as VGG-16, VGG-19, Inception V3, and ResNet-50, in recognizing malignant lung cells. The model achieved improved performance by supplying integrated features from each module to a meta-learner, resulting in higher recognition accuracy, precision, sensitivity, F-score, and specificity. These findings demonstrated the potential of CBAM-enhanced deep learning in improving early lung cancer detection and diagnosis, ultimately contributing to better patient outcomes.

## CONFLICT OF INTEREST STATEMENT

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

# REFERENCES

[1] J. D. Minna, J. A. Roth, and A. F. Gazdar, "Focus on lung cancer," Cancer cell, vol. 1, no. 1, pp. 49–52, 2002.

[2] D. S. Krishna, G. Srinivas, and P. P. Reddy, "A deep parallel hybrid fusion model for disaster tweet classification on twitter data," Decis Anal. J., vol. 11, p. 100453, 2024.

[3] H. F. Al-Yasriy, M. S. Al-Husieny, F. Y. Mohsen, E. A. Khalil, and Z. S. Hassan, "Diagnosis of lung cancer based on CT scans using CNN," IOP Conf. Ser. Mater. Sci. Eng., vol. 928, p. 022035, 2020.

[4] H. Alyasriy and M. Al-Huseiny, "The IQ-OTH/NCCD lung cancer dataset," Mendeley Data, vol. 4, 2023. Doi: 10.17632/bhmdr45bh2.4.

[5] A. Younis, L. Qiang, C. O. Nyatega, M. J. Adamu, and H. B. Kawuwa, "Brain tumor analysis using deep learning and VGG-16 ensembling learning approaches," Appl. Sci., vol. 12, no. 14, p. 7282, 2022.

[6] M. B. Sahaai, G. R. Jothilakshmi, D. Ravikumar, R. Prasath, and S. Singh, "ResNet-50 based deep neural network using transfer learning for brain tumor classification," AIP Conf. Proc., vol. 2463, no. 1, p. 020014, 2022.

[7] Z. Fu, H. Chen, M. He, and L. Liu, "An enhanced U-Network by combining PPM and CBAM for medical image segmentation," IEEE Access, vol. 12, pp. 107098–107112, 2024.

[8] A. Ijaz, B. Raza, I. Kiran, A. Waheed, A. Raza, H. Shah, and S. Aftan, "Modality specific CBAM-VGGNet model for the classification of breast histopathology images via transfer learning," IEEE Access, vol. 11, pp. 15750–15762, 2023.

[9] S. U. Rehman and V. Gruhn, "A sequential VGG16+ CNN based automated approach with adaptive input for efficient detection of knee osteoarthritis stages," IEEE Access, vol. 12, pp. 62407–62415, 2024.

[10] S. K. Dasari, S. Gorla, and P. R. PVGD, "A stacking ensemble approach for identification of informative tweets on twitter data," Int. J. Inf. Technol., vol. 15, no. 5, pp. 2651–2662, 2023.

[11] Y. Suo, Z. He, and Y. Liu, "Deep learning CS-ResNet-101 model for diabetic retinopathy classification," Biomed. Signal Process. Control, vol. 97, p. 106661, 2024.

[12] W. Meng, S. Liu, and H. Wang, "AFC-Unet: attention-fused full-scale CNN-transformer Unet for medical image segmentation," Biomed. Signal Process. Control, vol. 99, p. 106839, 2025.

[13] Y. Wan, D. Zhou, and C. Wang, "UMF-Net: a Unet-based multi-branch feature fusion network for colon polyp segmentation," Biomed. Signal Process. Control, vol. 99, p. 106851, 2025.

[14] T. Hussain and H. Shouno, "MAGRes-Unet: improved medical image segmentation through a deep learning paradigm of multi-attention gated residual U-Net," IEEE Access, vol. 12, pp. 40290–40310, 2024.

[15] M. A. Ruíz-Ortega, C. Karabağ, E. Roman-Rangel, and C. C. Aldasoro-Reyes, "DRD-Unet, a Unet-like architecture for multi-class breast cancer semantic segmentation," IEEE Access, vol. 12, pp. 40412–40424, 2024.

[16] N. Iriawan, A. A. Pravitasari, U. S. Nuraini, N. I. Nirmalasari, T. Azmi, M. Nasrudin, ... and W. Ferriastuti, "YOLO-Unet architecture for detecting and segmenting the localized MRI brain tumor image," Appl. Comput. Intelligence Soft Comput., vol. 2024, no. 1, p. 3819801, 2024.

[17] G. Li, B. Xi, Y. He, T. Zheng, Y. Li, C. Xue, and J. Chanussot, "Diamond-Unet: a novel semantic segmentation network based on U-Net network and transformer for deep space rock images," IEEE Geosci. Remote Sens. Lett., vol. 21, pp. 1–5, 2024.

[18] X. Wang, Z. Hua, and J. Li, "Attention based dual UNET network for infrared and visible image fusion," Multimedia Tools Appl., vol. 83, no. 25, pp. 1–22, 2024.

[19] A. A. AlArfaj, A. Altamimi, T. Aljrees, S. Basheer, M. Umer, M. A. Samad, ... and I. Ashraf, "Multi-step preprocessing with unet segmentation and transfer learning model for pepper bell leaf disease detection," IEEE Access, vol. 11, pp. 132254–132267, 2023.

[20] B. M. Abuhayi, Y. A. Bezabh, and A. M. Ayalew, "Lumbar disease classification using an Involutional neural based VGG Nets (INVGG)," IEEE Access, vol. 12, pp. 27518–27529, 2024.

[21] S. U. Rehman and V. Gruhn, "A sequential VGG16+ CNN based automated approach with adaptive input for efficient detection of knee osteoarthritis stages," IEEE Access, vol. 12, pp. 62407–62415, 2024.

[22] P. K. Sahoo, M. K. Panda, U. Panigrahi, G. Panda, P. Jain, M. S. Islam, and M. T. Islam, "An improved VGG-19 network induced enhanced feature pooling for precise moving object detection in complex video scenes," IEEE Access, vol. 12, pp. 45847–45864, 2024.

[23] V. Rajasekar, S. Krishnamoorthi, M. Saracevic, D. Pepic, M. Zajmovic, and H. Zogic, "Ensemble machine learning methods to predict the balancing of ayurvedic constituents in the human body," Comput. Sci., vol. 23, pp. 117–132, 2022.

[24] D. Ardila, A. P. Kiraly, S. Bharadwaj, B. Choi, J. J. Reicher, L. Peng, ... and S. Shetty, "End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography," Nat. Med., vol. 25, no. 6, pp. 954–961, 2019.

[25] H. F. Kareem, M. S. Al-Husieny, F. Y. Mohsen, E. A. Khalil, and Z. S. Hassan, "Evaluation of SVM performance in the detection of lung cancer in marked CT scan dataset," Indonesian J. Electr. Eng. Comput. Sci., vol. 21, no. 3, pp. 1731–1738, 2021.

[26] V. Rajasekar, M. P. Vaishnnave, S. Premkumar, V. Sarveshwaran, and V. Rangaraaj, "Lung cancer disease prediction with CT scan and histopathological images feature analysis using deep learning techniques," Results Eng., vol. 18, p. 101111, 2023.

[27] P. R. Radhika, R. A. Nair, and G. Veena, "A comparative study of lung cancer detection using machine learning algorithms," in *2019 IEEE International Conference on Electrical, Computer and Communication Technologies (ICECCT)*, Coimbatore, India, 2019, pp. 1–4.

[28] K. Pradhan and P. Chawla, "Medical Internet of things using machine learning algorithms for lung cancer detection," J. Manage. Anal., vol. 7, no. 4, pp. 591–623, 2020.

[29] H. O. Ahmed and A. K. Nandi, "High performance breast cancer diagnosis from mammograms using mixture of experts with efficient-net features (MoEffNet)," IEEE Access, vol. 12, pp. 133703–133725, 2024.

[30] L. Jiang, S. Zhu, and N. Sun, "An improved lightweight variant of EfficientNetV2 coupled with sensor fusion and transfer learning techniques for motor fault diagnosis," IEEE Access, vol. 12, pp. 84470–84487, 2024.

[31] M. Khomidov and J. H. Lee, "The novel EfficientNet architecture-based system and algorithm to predict complex human emotions," Algorithms, vol. 17, no. 7, p. 285, 2024.