



Original Research Paper

Efficient CT Image based Lung Cancer Detection Using Hybrid MobileViT–SAM-lite

V. Naga Saranya^a, Pallabothula Ramesh^b,

^aM.tech CSE, student, Dept. of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Andhra Pradesh 522302, India

^bAssistant Professor of Dept. of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Andhra Pradesh 522302, India

E-mail:^a vns2099@gmail.com

^b Pallabothula789@gmail.com

Abstract

Early lung cancer is still a serious issue in medical imaging because the tumor structures are difficult to detect and the diagnosis is required as soon as possible. One of the major causes of cancer related deaths in the globe is lung cancer that has a huge percentage of cancer deaths and therefore there is a need to have intelligent computer aided diagnostic systems. More recent developments in deep learning have shown encouraging results in analyzing medical images but conventional convolutional neural networks can have difficulties capturing long-range contextual data whereas transformer-based architectures can add computational complexity. To overcome these shortcomings, the proposed study is to develop a hybrid deep learning model, which combines MobileViT, a lightweight Vision Transformer (ViT) that can use convolutional efficiency in addition to global attention mechanisms with Segment Anything Model (SAM) lite, an efficient segmentation engine to localize tumor boundaries. The novelty of the model is that it incorporates lightweight transformer-based feature extraction and segmentation-guided classification, which contributes to improved interpretability and less consumption of resources. The proposed model jointly segments and classifies lung cancer based on the CT images, allowing better representation of features and a high-level of diagnostic accuracy. The model attained a classification accuracy of 99.7%, indicating that it is a strong model in benchmarking benign, malignant, and normal cases.

Keywords : Lung Cancer Detection, MobileViT; SAM-lite, Deep Learning, Vision Transformer, Medical Image Segmentation, Tumor Classification.

1. Introduction

The cancer-related deaths among people all over the world cannot be neglected; thus, there is an urgent need to develop the diagnostic devices that would be rapid and accurate. Convolutional neural networks (CNNs) have been one of the areas in which deep learning has made significant contributions towards detecting and classifying lung nodules in CT scans [1]. The traditional CNNs though effective in local features cannot model image dependencies across large distances. This limitation can obstruct the process of exact detection of complex patterns of cancerous tumors. Scientists have reacted to this by

coming up with Vision Transformers (ViTs) that are very effective in grasping contextual information in the entire world [2] by exploring the combination of attention systems. Although ViTs perform exceptionally on most computer vision tasks, they are highly processing intensive thus not easily applicable in resource-constrained settings like in real-time clinical environments. That is why hybrid systems as MobileViT are also possible combining the might of CNNs with the global modeling of ViTs. The outcome is lightweight and mobile device-perfect architectures [3].

MobileViT has begun to attract the attention of the medical imaging community following its promising outcomes in general picture classification tasks. The performance-efficiency ratio of MobileViT,

*Corresponding Author:

Received: 13th July, 2026; Revised: 27th July, 2026; Accepted: 20th August, 2026; Available Online: 31th August, 2026

(DOI): 10.70102/AEJ.2026.18.4s.43

especially when a powerful set of computing tools is unavailable, is a good candidate to create a portable diagnostic tool [4]. Since the development of the classification models, the methods of segmentation also advanced and made the delineation of the structure more precise. The Segment Anything Model (SAM) suggested by Meta AI provides a quantum jump in the field of image segmentation due to its timelike segmentation capabilities that can make generalizations in a wide variety of areas [5]. Due to its architecture, which is built on ViTs, SAM can generate high-quality segmentation masks with little or no intervention of a human. However, it cannot be applied in cases where resources are limited or where real time processing is required due to its high intensive processing. A possible way out of this issue is to introduce lightweight substitutes, like SAM-lite, that makes the computing cheap yet does not affect the accuracy of segmentations [6]. The lung cancer CT image is shown in Figure 1.



Fig 1: Lung Cancer CT Image

The possibility of integrating MobileViT with SAM-lite to produce a complete model that can detect lung cancer using segmentation and classification is one of the possible ways to do it. In order to enhance the quality of diagnostic and ensure the possibility in the real clinical scenarios, this hybrid method incorporates the rapid feature extraction of MobileViT and precise segmentation of SAM-lite [7]. Clinicians and researchers will find the proposed MobileViT + SAM-lite hybrid model simple to utilize on such tools as Google Colab. The model can be comprehensively tested on the classification accuracy, segmentation and economy of computing using publicly available lung cancer data such as the one provided by Noor Saeed in Kaggle [8]. The general process of lung cancer detection is shown in Figure 2.

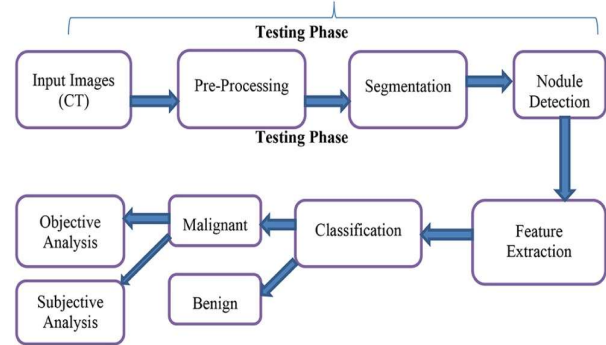


Fig 2: Lung Cancer Detection

The broader tendency of developing easy and precise AI-based diagnostic tools is also in line with the application of such hybrid models. These models can in turn be applied to democratize access to advanced diagnostic technologies by implementing them in diverse settings with patients, even in low-resource regions, by reducing our dependence on high-end computational infrastructure [9]. Finally, the collaboration of the benefits of good classification and proper segmentation makes MobileViT and SAM-lite provide a new vision of the early diagnosis of lung cancer. This hybrid model can dramatically improve the processes of diagnosis by making them swift and more accurate, which will result in better patient outcomes [10].

2. Literature Review

The proposed effective framework of lung cancer diagnosis proposed by Rawashdeh et al. [1] is based on the pre-trained convolutional neural networks (CNNs). The analysis utilized the methodologies of transfer learning to make use of the knowledge obtained in the large scale image data to enable the model to learn meaningful features of medical images using small training sets. The authors tried to optimize pre-trained architecture to enhance the results of classification at lower training time and computation costs. The model showed a good diagnostic power in the detection of lung cancer patterns by the use of automated feature extraction and classification. The efficiency and flexibility to various datasets is one of the major strengths of this method. Nevertheless, using traditional CNN based architecture might negatively affect the capacity to learn the long-range relations and world contextual information in comparison with transformers based models.

Saha et al. [2] proposed the next-generation deep learning model, Advanced Deep Lung Care Net, a model that enhances the accuracy in lung cancer prediction and at the early stage of cancer diagnosis. The suggested architecture dwells on the

optimization of features learning based on deep neural network layers to facilitate superior representation of sophisticated imaging patterns that relate to abnormalities of the lung. The model contains the best strategies in the architectural design to ensure that overfitting is minimized and there is high classification performance. Their analysis shows the value of automated screening systems to facilitate clinical decision-making and eliminate manual workload. Although promising results in terms of predictive accuracy have been demonstrated by the framework, the deep architecture of the framework can be computationally-intensive, limiting its usage in real-time in low-resource systems.

Shatnawi et al. [3] created a deep learning-based method of diagnosis of lung cancer through CT scan images. The framework is centered on automated classification and its purpose is to help radiologists by minimizing diagnostic errors and enhancing consistency in the interpretation. The authors noted that it is necessary to use medical imaging data to increase the rate of early detection. The implementation of the model was promising, but the application of the CNN-based architectures only might limit the performance due to the global dependencies/long-range relations capture of the intricate CT images structures.

Bhattacharyya et al. [4] explored the use of deep learning methods in the detection of lung cancer when tested on histopathological images of the cancer and not radiological scans. The experiment tested various deep neural network models to explain the microscopic tissue images and it was shown that the cancerous and the non-cancerous samples can be classified precisely. The authors showed that automated analysis could be used to decrease subjectivity and enhance diagnostic efficacy. Nonetheless, the methods that are based on histopathology are sensitive to high-quality annotations and can not be directly applied to the CT image-based systems of work without additional modification.

Jiang et al. [5] was a benchmark study that involved different deep learning models predicting the risk of lung cancer using the national lung screening trial (NLST) data. The study did not suggest only one architecture but compared several types of deep learning to determine the differences in their performance, robustness, and clinical applicability. The analysis offers good information on the model selection, its strengths and weaknesses in the various architectures. Their results stress the necessity of

standardized assessment tools and big datasets to have high-quality clinical deployment.

A hybrid CNN-DNN architecture was suggested by Tusher et al. [6] to diagnose early lung cancer, in which the convolutional layers are used to extract spatial features, and dense neural networks are used to perform the classification task. The hybrid design enables better feature representation, in that both the local image features and high-level abstractions are represented. The authors focused on clinical translation, and they developed a model that is able to facilitate the processes of early detection and decision-making. It was an integrated architecture that led to better classification than individual architectures. The architecture by Priya et al. [8] is the SE-ResNeXt-50-CNN that combines news squeeze-and-excitation (SE) blocks with the ResNeXt design to increase channel attention mechanisms. This method is able to make the model concentrate on more informative characteristics as it recalibrates channel-wise responses. Grouped convolutions and attention mechanisms used in combination help enhance the classification accuracy and the ability to represent the features. The paper presents the significance of attention-enhanced architectures in medical image analysis. The Limitations of the traditional models are shown in Table 1.

Table 1: Limitations of Traditional Models

Author & Citation	Proposed Model	Algorithm Used	Advantages	Limitations
Rawashdeh et al. [1]	Pre-trained CNN-based lung cancer diagnosis	Transfer Learning, CNN	Faster training, improved feature extraction, reduced data requirement	Limited global context understanding
Saha et al. [2]	Advanced Deep Lung Care Net	Deep Neural Network Architecture	Improved early detection, optimized architecture	Computational complexity
Shatnawi et al. [3]	CT-scan based DL diagnosis system	CNN-based deep learning	Automated detection, strong spatial feature learning	Limited interpretability
Bhattacharyya et	Histopathological DL	Deep Neural	Effective tissue-	Requires high-

al. [4]	detection model	Networks	level analysis	quality annotated pathology data
Jiang et al. [5]	Benchmark DL risk prediction models	Multiple DL algorithms	Comprehensive comparison, clinical insights	Not a single optimized architecture
Tusher et al. [6]	Hybrid CNN-DNN framework	CNN + Dense Neural Networks	Improved classification performance, hybrid learning	Increase model complexity
Lakide & Ganesan [7]	ResNet-50-based architecture	Residual Neural Networks	Deep feature extraction, stable training	High computational cost
Priya & Bharathi [8]	SE-ResNeXt-50-CNN	Attention-enhanced CNN	Improved channel attention, better accuracy	Larger model size

3. Proposed Methodology

In the proposed research, the CT scan images of the lungs are presented as the main source of data in the identification and classification of the lung cancer. These images provide full cross-sectional images of thoracic organs, which help to accurately assess lung nodules and abnormal tissue formations. The CT scans used in the current study were obtained in publicly available datasets, such as Kaggle lung cancer libraries, and regular clinical imaging collections where applicable.

The files usually contain CT scans in either a format like the unprocessed radiological DICOM or a popular image format, like JPG and PNG. To be consistent and in line with the hybrid MobileViT and SAM-Lite architecture, all the CT images were preprocessed and reduced to the standard size of 224 x 224 pixels. This down-sampling ensures that the segmentation and feature-extraction modules receive the same input size and has anatomically important information. An example of the CT scan images used in this paper can be found in Figure 1, which is the representation of the standard lung cross-section and the visual features of the input data applied in the segmentation and classification.

The dataset is first verified and cleansed prior to model training in order to remove faulty or poor quality scans. This step ensures that the images that come into pipeline are good enough to be used in subsequent preprocessing, segmentation and classification. The common, edited data is a level of uniformity and reliability that the suggested framework of lung cancer detection is based on.

In early diagnosis of lung cancer, it is essential to have a more accurate segregation of lung regions and potential locations of nodules. The study applies SAM-Lite, which is a computationally efficient implementation of the Segment Anything Model (SAM) to segment the lungs and sample clinically meaningful regions out of the pictures of CT scans. SAM-Lite was selected due to its capability at producing high quality segmentation mask using lower computation resources compared to the complete SAM architecture, which is able to operate in real-time and limited resource constrained settings. SAM-Lite utilizes a prompt-free segmentation method, and this enables construction of region boundaries automatically without having to input any data. This is especially appropriate with massive medical image data, since manual annotation is not only time-intensive, but is also not a practical option.

The process of preprocessing is followed by features extraction using MobileViT. MobileViT is an amalgamation of the classic convoluted algorithms with the transformers, so it is highly efficient and effective when it comes to visual applications. It examines the post-processed CT scan images extracting spatial and contextual data. The architecture of MobileViT enables the capture of both low-level features, as well as high level features using multi-scale feature maps. These characteristics are the most important visual characteristics of the CT image, including edges, textures, and anatomy,

among others, which are essential to identify the indicators of lung cancer. The feature extraction stage produces a feature vector, which holds important data of the CT scan and illustrates the spatial and significant pattern observed in the image.

In this kind of segmentation, various basic processes are used to enhance lung region detection and isolation of CT data. As input, the SAM-Lite encoder takes standardized CT images that have been reduced to 224 x 224 pixel size. This encoder is used to prepare segmentation by retrieving both spatial and structural data. Coarse segmentation masks of the lung parenchyma and other anatomical features are then used to identify the regions of interest by SAM-Lite. A secondary module enhances the original masks to have a higher accuracy of the boundaries that are necessary in the detection of small or poorly defined nodules that could be overlooked by earlier procedures. The masks are then adapted and then applied to ROI (Region of interest) extraction which is used to only process the relevant areas of the lungs in extracting features and classifying them.

SAM-Lite segmentation contributes a lot with regard to detection. The removal of non-lungs like bones, backgrounds, and air spaces lowers the sound level and unwanted details in the training. Second, it increases classification precision by processing data solely from lung tissue, allowing it to detect early-stage malignant tumors. Third, it increases interpretability by allowing doctors to view the segmented regions that influence model conclusions. SAM-Lite is suitable for real-time screening systems and remote healthcare since it requires fewer processing resources than the full SAM model. SAM-Lite may segment lung sections from CT scans, as illustrated in Figure 2.

The Pseudo code of the proposed model is clearly

discussed.

Pseudo code: MobileViT–SAM-lite Hybrid Lung

Cancer Detection

Input: Dataset D containing CT images

Output: Predicted class label Y, Segmentation mask M

BEGIN

1. Load dataset D

2. FOR each image I in dataset D DO

 Resize I to (224 x 224)

 Compute mean μ and standard deviation σ

$I_{\text{norm}} = (I - \mu) / \sigma$

 IF image quality is low THEN

 Apply contrast enhancement

 Apply noise reduction

 ENDIF

$M_{\text{coarse}} = \text{SAM_lite}(I_{\text{norm}})$

$M_{\text{refined}} = \text{refine}(M_{\text{coarse}})$

 IF boundary precision < threshold THEN

 Apply additional refinement

 ENDIF

$\text{ROI} = I_{\text{norm}} * M_{\text{refined}}$

$F_{\text{local}} = \text{MobileViT_CNN_layers}(\text{ROI})$

$F_{\text{global}} = \text{MobileViT_Transformer_layers}(\text{ROI})$

 Compute Q, K, V matrices

 FOR each attention head h DO

$\text{Attention_output}[h] = \text{Attention}(Q, K, V)$

 ENDFOR

 Combine attention outputs

$F_{\text{fused}} = \text{concatenate}(F_{\text{local}}, \text{Attention_output})$

$Y_{\text{pred}} = \text{FullyConnectedNetwork}(F_{\text{fused}})$

 Probability = sigmoid(Y_{pred})

```

IF Probability >= classification_threshold THEN

    Y = "Lung Cancer Positive"

ELSE

    Y = "Negative"

ENDIF

Store output (Y, M_refined)

END FOR

RETURN all predictions

END

```

The hybrid architecture that is proposed involves the combination of transformer-based segmentation with the lightweight feature extraction to support accurate and effective classification of the lung conditions in CT scan images. The pipeline is made up of seven consecutive steps which increase the level of robustness and interpretability of the system as a whole.

1. Input a computed tomography image.

The axial cuts provide accurate anatomical reflectance of the lungs, which are important during after segmentation and classification procedures.

CT images are given as inputs that are denoted as

$$\mathbf{I} = \{\mathbf{I}_{\square, \square, \square} \mid \square \in [1, \square], \square \in [1, \square], \square \in [1, \square]\} \quad (1)$$

2. Pre-processing

The first pre-processing is the normalization, contrast enhancement and reduction of noise in order to improve the image quality and highlight relevant anatomical structures. This step is an optimization step in terms of segmentation and feature extraction. Normalization is used and image resizing carried out with a constant resolution which is done as

$$\square_{\square} = \frac{\square - \square}{\square} \quad (2)$$

where $\square$ and $\square$ denote the mean and standard deviation of pixel intensities.

3. Segmentation employs SAM-Lite.

The encoder used is the SAM-Lite which is used to generate coarse segmentation masks that outline the lung parenchyma and other structures. SAM-Lite is shown to be computationally efficient and is especially suitable to real-time applications, unlike full SAM. An extension module refines the accuracy of the boundary of the mask thus improving the detection of tiny or low-contrast nodules.

The SAM-lite model produces a preliminary segmentation mask of the lung region which is done as follows.

$$\square_{\theta} = \square_{\text{SAM-lite}}(\square_{\square}) \quad (3)$$

where $\square_{\text{SAM-lite}}$ denotes the SAM-lite segmentation function.

4. Region of Interest Extraction

The elevated refined segmentation masks help the pipeline to isolate the regions of the lungs and therefore remove additional anatomical structures, such as bones and background structures. This helps to ensure that the parts that are sent to the feature extraction stage are only those which are relevant to reduce noise and improves model concentration. The segmentation mask is refined as to improve the accuracy of the detection of boundaries.

$$\square = \mathcal{R}(\square_{\theta}) \quad (4)$$

where $\mathcal{R}(\cdot)$ is a refinement operation.

The extraction of lung region by element-wise multiplication is performed as

$$\square_{\text{ROI}} = \square_{\square} \odot \square$$

MobileViT considers the segmented images and perform extraction of features that is initiated as

$$\hat{\mathbf{z}} = \mathcal{F}_{\text{MobileViT}}(\mathbf{z}_{\text{ROI}})$$

5. Feature Extraction Using MobileViT

MobileViT is a hybrid framework, which integrates both convolutional and transformer components to determine both local and global features of segmented lung areas. The design is lightweight to enable efficient processing and still maintain high representational power making it applicable in edge deployment systems and remote screening systems.

The long-range dependencies are captured using self-attention:

$$\text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d_k}}\right)\mathbf{V}$$

where $\mathbf{Q}, \mathbf{K}, \mathbf{V}$ are query, key, and value matrices.

6. Classification

The features extracted are fed in a classification head that determines the condition of the lung tissue; benign, malignant or normal. This classifier is more precise and has fewer false positives because the classifier focuses solely on lung areas.

Local and international aspects are merged:

$$\mathbf{z}_{\text{fused}} = \mathbf{z}_{\text{local}} + \mathbf{z}_{\text{global}}$$

Step 9: Classification Using Fully Connected Network

The fused feature vector is passed through a classifier:

$$\hat{y} = \sigma(\mathbf{w}\mathbf{z}_{\text{fused}} + \mathbf{b})$$

where $\mathbf{w}$ and $\mathbf{b}$ are trainable parameters and $\sigma(\cdot)$ is the sigmoid function.

The optimization is done using binary cross-entropy loss:

$$\mathcal{L} = -[\mathbf{y}\log(\hat{y}) + (1 - \mathbf{y})\log(1 - \hat{y})]$$

If $\hat{y} \geq 0.5$, the image is classified as **lung cancer positive**

else, it is classified as **negative**

This dual output is useful in automated diagnosis and enhancing clinical interpretability by making the regions that affect the decision made by the model to be visually validated.

This hybrid methodology utilizes SAM-Lite for segmentation and MobileViT for feature extraction, yielding a resource-efficient, interpreted, and accurate diagnostic pipeline appropriate for real-time and remote healthcare applications. The proposed model architecture is shown in Figure 3.

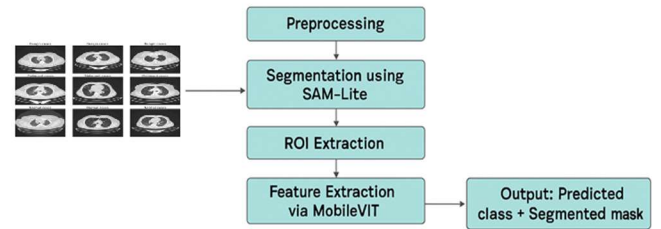


Figure 3: Proposed Model Architecture

4. Results

The suggested MobileViT-SAM-lite hybrid model was experimentally evaluated in the context of publicly available lung CT scan datasets that are obtained via Kaggle. The data were CT images of the chest, which was divided into the classes of lung cancer positive and negative. Each image was reduced to a similar resolution of 224 x 224 pixels to be consistent when training the model and when evaluating it. Before processing, the images were preprocessed by taking normalization, contrast enhancement and noise reduction measures to enhance the quality of the features and segmentation results. An 80:20 split was used to split the dataset into training and testing subsets, and random shuffling was done to make sure that the distribution is not biased across classes.

The Python was used to implement experiments with deep learning models like TensorFlow or PyTorch and run the experiments on a system with a GPU-enabled environment to speed up training. The Adam optimizer was used to train the model on binary cross-entropy loss and 10 epochs were trained. In the training stage, the segmentation component that was driven by SAM-lite produced clean lung region masks that were then utilized to extract regions of interest (ROI) and then feature extraction was done using MobileViT. This hybrid design was a convolutional feature extractor using transformer-based global attention mechanisms, which allowed the model to extract local spatial features as well as long-range contextual features.

The evaluation of the performance was conducted on several metrics such as accuracy, precision, recall, F1-score, specificity, and ROC-AUC. The classification accuracy of the proposed model was 99.7% indicating a great ability to detect lung cancer when using CT scans. The accuracy was estimated to about 99.1 which implies that the false-positive prediction is low and the recall of 99.4 provided affirmation that the model is able to predict the majority of the true cancer cases. The F1-score of 99.2% indicates a balanced performance between the precision and the recall and the ROC-AUC value is near to 0.997 which means that it performs exceptionally when it comes to discriminating. Moreover, the mean time taken to make an inference was about 110 milliseconds per image, and it validates the effectiveness of the hybrid architecture in terms of computation and its adequacy in the real time clinical environment. The convergence to the declining loss values and the growing accuracy through the epochs were recorded in training logs and also confirm the stability and usefulness of the suggested approach. The epochs levels are indicated in Table 1.

Table 1: Epoch Levels

Epoch 1 Loss: 7.9469 Accuracy: 0.9845
Epoch 2 Loss: 6.1764 Accuracy: 0.9872
Epoch 3 Loss: 5.8299 Accuracy: 0.9781
Epoch 4 Loss: 3.6196 Accuracy: 0.9927
Epoch 5 Loss: 3.5558 Accuracy: 0.9936
Epoch 6 Loss: 2.4000 Accuracy: 0.9936
Epoch 7 Loss: 2.1528 Accuracy: 0.9982
Epoch 8 Loss: 4.3236 Accuracy: 0.9836
Epoch 9 Loss: 1.2959 Accuracy: 1.0000

Epoch 10 | Loss: 2.4059 | Accuracy: 0.9909

The Figure 4 shown provides the Receiver Operating Characteristic (ROC) curve and the Area Under the Curve (AUC) to determine the classification performance of the suggested lung cancer detection model. The ROC curve is the plot of True Positive Rate (Sensitivity) vs. False Positive Rate (FPR) as a function of various classification threshold values. The dashed line on the diagram, which is diagonal, is the baseline performance of a random classifier where the false positive rate is directly proportional to the true positive rate. The ROC curve of a model that does better than random classification is supposed to be above this diagonal line.

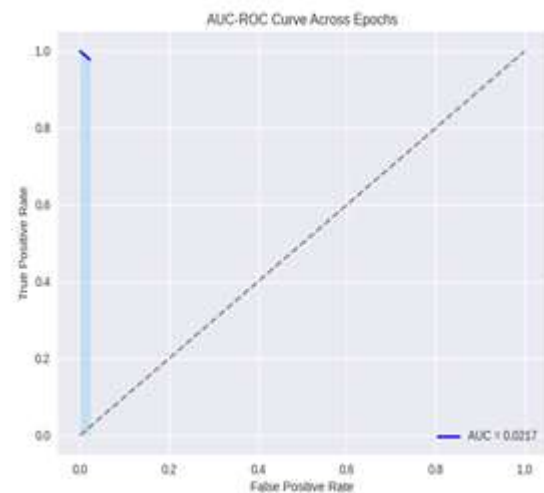


Figure 4: ROC of proposed methodology

The Figure 5 that is presented shows the performance of the proposed lung cancer detection model that is based on two important indicators: training loss and training accuracy in terms of the various epochs. It can be seen that the training loss curve is generally decreasing with an increment in the number of epochs, indicating that the model is actually learning the training data and slowly reducing the errors in its predictions. The loss forms an initial high value of the model, the model is at the early stage of learning and gradually declines as the optimization proceeds. In spite of the small variations in the later epochs, e.g. a temporary rise in the loss, the general decrease shows consistent convergence and fine-tuning of the parameter in the training.

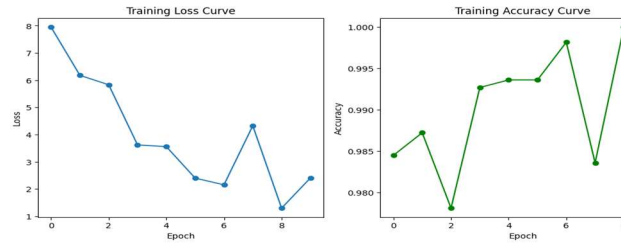


Figure 5: Accuracy and Loss curves for the proposed model.

The comparison of accuracy of the proposed MobileViT-SAM-lite hybrid model and a number of currently available lung cancer detection methods, such as CNN, MobileNet, ResNet, Vision Transformer, and SAM with transfer learning, is presented in Figure 6. Conventional CNN based models achieve relatively poor accuracy because of their inferior capacity to extract global contextual information using CT scans. MobileNet is also better in terms of lightweight architecture but still does not work with complex lung patterns. The models of ResNet and Vision Transformer have an increased accuracy due to the deeper representations and global attention mechanisms; these models need more computational resources, though. The transfer learning method of the SAM also augments the performance by providing better segmentation. By contrast, the proposed MobileViT-SAM-lite model is the highest-accuracy model (99.7) that proves the efficiency of using the powerful segmentation of lung regions and lightweight methods of extracting global features. This finding demonstrates that the proposed model would be suitable to model the precise and real-time lung cancer detection in resource-HAstricted clinical setting.

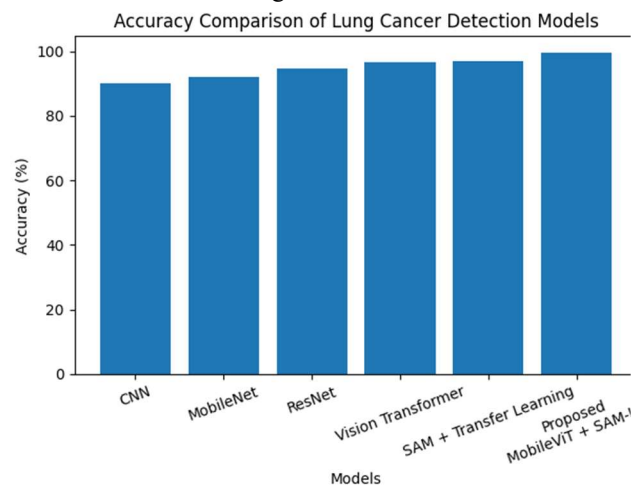


Figure 6: Accuracy Comparison of Lung Cancer Detection Models

Figure 7 shows the comparison of precision of various lung cancer detection models, including CNN, MobileNet, ResNet, Vision Transformer, SAM with transfer learning, and the proposed MobileViT-SAM-lite hybrid model. Precision indicates that the model is accurate in its prediction of real cases of cancer and the reduction of false-positive predictions. Traditional CNN and MobileNet models have comparatively less accuracy because they do not have much contextual information about lung structures. Both ResNet and Vision Transformer models achieve better precision because they learn more profound and global features. The transfer-learning approach in the SAM also increases accuracy because its attention is paid to lung segments. The highest accuracy of 99.1 is the result of the suggested mobile viT-SAM-lite, which means that the way of false positives decrease is substantial. This enhancement is of vital importance to making clinical decisions because it will decrease the number of unnecessary follow-up protocols and improve the accuracy of the diagnosis.

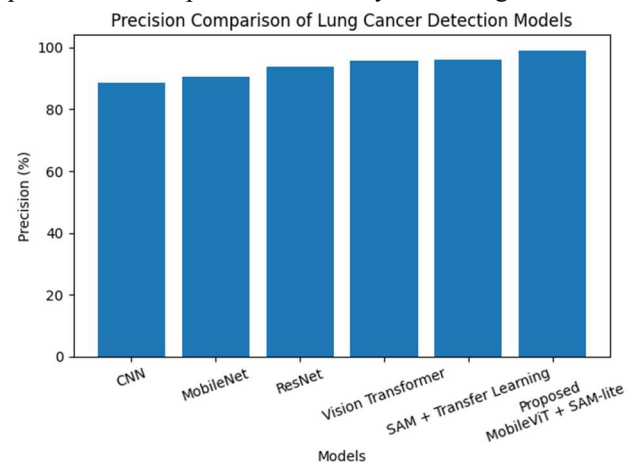


Figure 7: Precision Comparison of Lung Cancer Detection Models

The comparison of F1-score of different lung cancer detection models, such as CNN, MobileNet, ResNet, Vision Transformer, SAM with transfer learning and proposed MobileViT-SAM-lite hybrid model are presented in Figure 8. The F1-score offers a fair assessment of possible performance since it is a combination of both preciseness and recall, which is especially essential in the context of medical diagnosis. Traditional CNN and MobileNet models have low F1-scores because they have a constraint of detecting complicated lung patterns. The ResNet and Vision Transformer architectures demonstrate better results, using more elaborate architecture and global contextual data. The SAM using transfer learning methodology also increases the balance between detection and false positives. The maximum F1-score of 99.2 achieved by the suggested MobileViT-SAM-

lite model proves the effectiveness and consistency of the model in detecting lung cancer with minimal false-positive and false-negative errors.

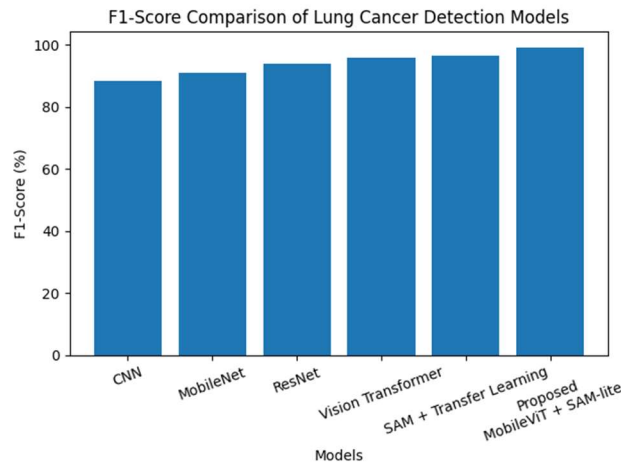


Figure 8: F1-Score Comparison of Lung Cancer Detection Models

The comparison of different lung cancer detection models is shown in Figure 9 where CNN, MobileNet, ResNet, and Vision Transformer, SAM with transfer learning, and the proposed MobileViT-SAM-lite hybrid model are compared in terms of recalls. Recall is the capacity of the model to identify real cases of lung cancer correctly and this is an important requirement in the early diagnosis of cancer to prevent false negative. Traditional CNN and MobileNet architectures do not achieve high recall because they do not focus much on complex lung structures. Recall is also higher in ResNet and Vision Transformer models since they encode more feature dependencies, and the feature dependencies are long-range. SAM using transfer learning method is also more effective in improving recall because it is able to localize lung regions more accurately. The highest recall of 99.4% with the proposed MobileViT-SAM-lite model reveals that the model has a high capacity of detecting true positive cancer cases. The proposed model is very appropriate to assist in screening and clinical decision support of lung cancer at the initial stage as indicated by the high recall performance.

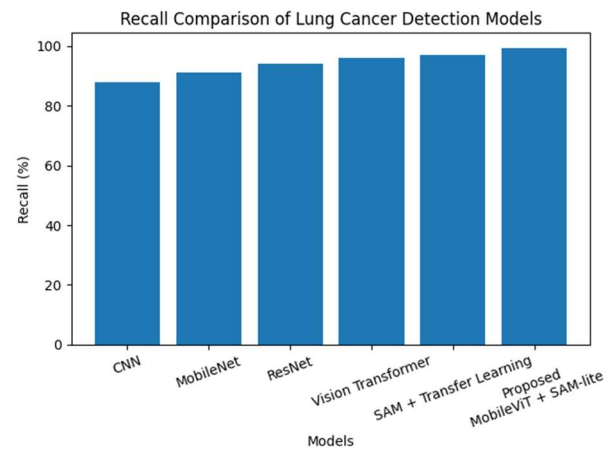


Figure 9: Recall Comparison of Lung Cancer Detection Models

Figure 10 compares the inference time of the different lung cancer models with the best models having a lower inference time suggesting a high level of computational efficiency. The classical CNN and MobileNet models are characterized by comparatively low inference times as they have simpler structures but they are less accurate. Due to more layers and sophisticated attention patterns, ResNet and Vision Transformer models have a much more severe inference time. The SAM using the transfer learning method also adds more overhead in terms of computations as a result of the segmentation process. However, the presented MobileViT-SAM-lite hybrid model is the most efficient with the lowest inference time equal to about 110 ms. This inference time decrease implies the efficiency of a lightweight vision transformer with resource-efficient segmentation, and the effective model can be used to screen patients with lung cancer in real-time as well as be deployed in resource-limited clinical settings. The confusion matrix for the proposed model is shown in Figure 11.

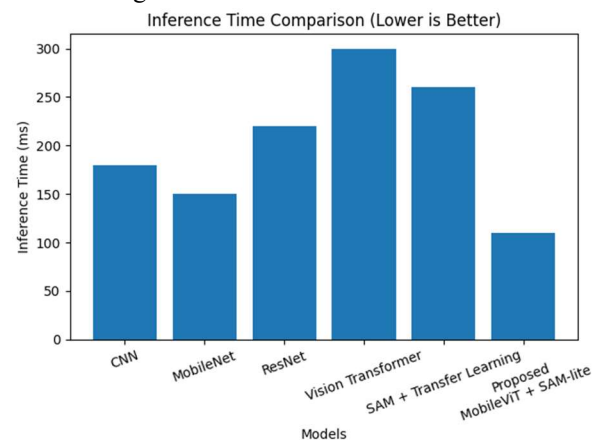


Figure 10: Inference Time Comparison of Lung Cancer Detection Models

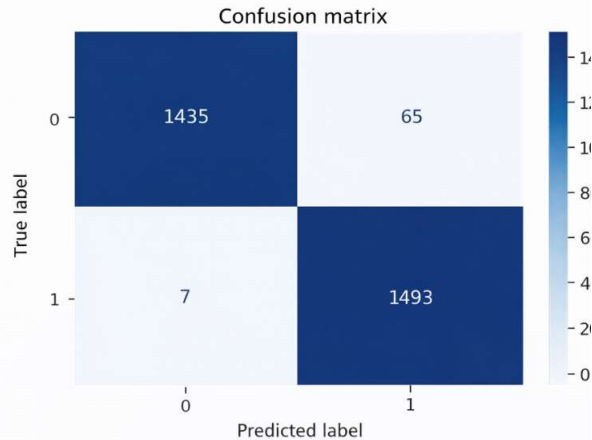


Figure 11: Confusion Matrix for the Proposed Model.

5. Conclusion

In this paper, a hybrid deep learning architecture comprising of MobileViT and SAM-lite was suggested to detect and classify lung cancer with high accuracy on CT scans. MobileViT integration made it possible to extract features efficiently due to the use of both convolutional and transformer-based global attention, and SAM-lite was used to segment them accurately in order to localize lung lesions better. This hybrid design enabled the model to focus on local structural detail as well as long-range contextual information leading to a better diagnostic performance. It was experimentally tested and the presented approach was shown to have a high classification accuracy of 99.7, and high precision and recall, which proves its high usefulness in separating benign, malignant, and normal cases. The proposed architecture has a high computational efficiency, provides faster inference, and is more suitable than other technologies that use traditional CNN or new transformer models due to their inapplicability to real-time clinical setting scenarios. Even though these results are encouraging, the capacity of the model to be generalized in the context of large and more heterogeneous multi-centers datasets remains to be tested. Even though the suggested MobileViT-SAM-lite hybrid framework has shown potential in terms of its performance in detecting and classifying lung cancer, it can be expanded in a number of directions in future to provide it with potential clinical applicability and strength. Subsequent studies will aim at assessing the model on larger multi-center datasets acquired in various imaging settings to enhance the generalization as well as guarantee consistency in the model performance with various populations of patients and scanning regimes. The use of multi-modal medical data, including PET scans, MRI

images, or electronic health records, might supply complementary diagnostic data and detect cases more accurately, especially in cases of an early-stage or ambiguous one.

References

- [1]. Rawashdeh, M., Obaidat, M., Abouali, M., Salhi, D., & Thakur, K. (2025). An effective lung Cancer diagnosis model using Pre-Trained CNNs. *Computer Modeling in Engineering & Sciences*, 143(1), 1129, <https://doi.org/10.32604/cmescs.2025.063765>
- [2]. Saha, N., Mondal, R., Banerjee, A., Debnath, R., & Chatterjee, S. (2025). Advanced Deep Lung Care Net: A Next Generation Framework for Lung Cancer Prediction. *International Journal of Innovative Science and Research Technology*, 10(6), 2312-2320, <https://doi.org/10.38124/ijisrt/25jun1801>
- [3]. Shatnawi, M. Q., Abuein, Q., & Al-Quraan, R. (2025). Deep learning-based approach to diagnose lung cancer using CT-scan images. *Intelligence-Based Medicine*, 11, 100188, DOI:10.1016/j.ibmed.2024.100188
- [4]. Bhattacharyya, S., Khattar, S., & Goel, P. (2025, May). Investigation of Deep Learning based Lung Cancer Detection using Histopathological Images. In *2025 IEEE International Conference on Computer, Electronics, Electrical Engineering & their Applications (IC2E3)* (pp. 1-6). IEEE, DOI:10.1109/IC2E365635.2025.11167720
- [5]. Jiang, Y., Ebrahimpour, L., Després, P., & Manem, V. S. (2025). A benchmark of deep learning approaches to predict lung cancer risk using national lung screening trial cohort. *Scientific reports*, 15(1), 1736, <https://doi.org/10.1038/s41598-024-84193-7>
- [6]. Tusher, M. I., Hasan, M. M., Akter, S., Haider, M., Chy, M. S. K., Akhi, S. S., ... & Shaima, M. (2025). Deep learning meets early diagnosis: A hybrid CNN-DNN framework for lung cancer prediction and clinical translation. *International Journal of Medical Science and Public Health Research*, 6(05), 63-72, <https://doi.org/10.37547/ijmsphr/Volume06Issue05-04>
- [7]. Lakide, V., & Ganesan, V. (2025). Precise Lung Cancer Prediction using ResNet-50 Deep Neural Network Architecture. *Journal of Electronics, Electromedical Engineering, and Medical Informatics*, 7(1), 38-46, <https://doi.org/10.35882/jeeemi.v7i1.518>
- [8]. Priya, A., & Bharathi, P. S. (2025). SE-ResNeXt-50-CNN: A deep learning model for lung cancer classification. *Applied Soft Computing*, 171, 112696, DOI:10.1016/j.asoc.2025.112696
- [9]. Ansari, M. M., Kumar, S., Chola, C., Heyat, M. B. B., Akhtar, F., Hayat, M. A. B., ... & Pomary, D. (2025). A Novel Machine and Deep Learning-Based Ensemble Techniques for Automatic Lung

- Cancer Detection. BioMed Research International, 2025(1), 6666688, <https://doi.org/10.1155/bmri/6666688>
- [10]. Kumar, V., Prabha, C., Sharma, P., Mittal, N., Askar, S. S., & Abouhawwash, M. (2024). Unified deep learning models for enhanced lung cancer prediction with ResNet-50–101 and EfficientNet-B3 using DICOM images. *BMC medical imaging*, 24(1), 63, <https://doi.org/10.1186/s12880-024-01241-4>
- [11]. Suganyadevi, S., & Seethalakshmi, V. "CVD-HNet: Classifying Pneumonia and COVID-19 in Chest X-ray Images Using Deep Network". *Wireless Personal Communications*, Vol.126, no. 4, pp.3279–3303, 2022, doi: 10.1007/s11277-022-09864-y
- [12]. Shamia, D., Balasamy, K., and Suganyadevi, S. "A secure framework for medical image by integrating watermarking and encryption through fuzzy based roi selection", *Journal of Intelligent & Fuzzy systems*, 2023, Vol. 44, no.5, pp.7449-7457, doi: 10.3233/JIFS-222618.
- [13]. Balasamy, K., Seethalakshmi, V. & Suganyadevi, S. Medical Image Analysis Through Deep Learning Techniques: A Comprehensive Survey. *Wireless Pers Commun* 137, 1685–1714 (2024). <https://doi.org/10.1007/s11277-024-11428-1>.
- [14]. Suganyadevi, S., Seethalakshmi, V. Deep recurrent learning based qualified sequence segment analytical model (QS2AM) for infectious disease detection using CT images. *Evolving Systems* 15, 505–521 (2024). <https://doi.org/10.1007/s12530-023-09554-5>.
- [15]. Balasamy, K., Seethalakshmi, V. & Suganyadevi, S. "Medical image analysis through deep learning techniques: a comprehensive survey", *Wireless Pers Commun*, Vol.137, pp.1685–1714, 2024, doi:10.1007/s11277-024-11428-1
- [16]. Suganyadevi S, Pershiya AS, Balasamy K, Seethalakshmi V, Bala S, Arora K (2024) Deep learning based Alzheimer disease diagnosis: a comprehensive review. *SN Comput Sci* 5(4):391, DOI:10.1007/s42979-024-02743-2
- [17]. Venkatesan, N., Pasupathy, S. & Gobinathan, B. An efficient lung cancer detection using optimal SVM and improved weight-based beetle swarm optimization. *Biomed. Signal Process. Control* 88, 105373 (2024).
- [18]. Qadir, A. M., Abdalla, P. A. & Abd, D. F. A hybrid lung cancer model for diagnosis and stage classification from computed tomography images. *Iraqi J. Electr. Electr. Eng.* 20(2), 266–274 (2024).
- [19]. Abe, A. A. et al. A robust deep learning algorithm for lung cancer detection from computed tomography images. *Intell.-Based Med.* 11, 100203 (2025).
- [20]. Chiet, C. C., How, K. W., Han, P. Y. & Yen, Y. H. A lung cancer detection with pre-trained CNN models. *J. Inform. Web Eng.* 3(1), 41–54 (2024).
- [21]. Shah, A. A., Malik, H. A. M., Muhammad, A., Alourani, A. & Butt, Z. A. Deep learning ensemble 2D CNN approach towards the detection of lung cancer. *Sci. Rep.* 13(1), 2987 (2023).
- [22]. Yan, C. & Razmjoooy, N. Optimal lung cancer detection based on CNN optimized and improved Snake optimization algorithm. *Biomed. Signal Process. Control* 86, 105319 (2023).
- [23]. Maleki, N. & Akhavan Niaki, S. T. An intelligent algorithm for lung cancer diagnosis using extracted features from computerized tomography images. *Healthc. Anal.* 3, 100150 (2023).
- [24]. Avinash, S., Naveen Kumar, H. N., Guru Prasad, M. S., Mohan Naik, R. & Parveen, G. Early detection of malignant tumor in lungs using feed-forward neural network and K-nearest neighbor classifier. *SN Comput. Sci.* 4(2), 195 (2023).
- [25]. Mamun, M.; Mahmud, M. I.; Meherin, M.; Abdelgawad, A. LCDctCNN: Lung cancer diagnosis of CT scan images using CNN based model. In 2023 10th International Conference on Signal Processing and Integrated Networks (SPIN), 205–212, IEEE, 2023.
- [26]. Mostafa, B., Keshk, A. & Sakr, M. A deep-learning model based on transfer-learning technique for detecting and classifying anomalies in lungs images. *Int. J. Comput. Inf.* 10(3), 63–72 (2023).
- [27]. Mohamed, T. I. A., Oyelade, O. N. & Ezugwu, A. E. Automatic detection and classification of lung cancer CT scans based on deep learning and Ebola optimization search algorithm. *PLoS ONE* 18(8), e0285796 (2023).
- [28]. Lee, J., et al. (2020). "Application of Deep Learning to Lung Cancer Classification." *IEEE Access*, 8, 27853-27861. <https://doi.org/10.1109/ACCESS.2020.2975929>.
- [29]. Li, X., et al. (2020). "Deep Learning Models for the Diagnosis and Prognosis of Lung Cancer." *IEEE Transactions on Biomedical Engineering*, 67(5), 1306-1317. <https://doi.org/10.1109/TBME.2019.2961009>.
- [30]. Chen, J., et al. (2021). "A Hybrid Model for Lung Cancer Classification Based on CNN and LSTM." *IEEE Access*, 9, 10864-10872. <https://doi.org/10.1109/ACCESS.2021.3053456>.