



## NoduViT: An Attention-Driven DeepLab–CustomViT Hybrid Model for Segmentation-Informed Multiclass Pulmonary Patch Classification

Fachruddin<sup>1</sup>, Yovi Pratama<sup>2\*</sup>, Errissya Rasywir<sup>3</sup>, Yessi Hartiwi<sup>4</sup>, Jasmir<sup>5</sup>, Hastie Audytra<sup>6</sup>, Marrylinteri Istoningtyas<sup>7</sup>

<sup>1,4,5</sup>Information System Department, Computer Science Faculty, Dinamika Bangsa University, Jambi, Indonesia

<sup>2,3,7</sup>Informatics Engineering Department, Computer Science Faculty, Dinamika Bangsa University, Jambi, Indonesia

<sup>6</sup>Information System Department, Computer Science Faculty, Nahdlatul Ulama Sunan Giri Bojonegoro University, Bojonegoro, Indonesia

\*[yovi.pratama@gmail.com](mailto:yovi.pratama@gmail.com)

### Abstract

*Lung cancer detection using CT imaging requires accurate lung segmentation and nodule classification. This study evaluates SegFormer configurations and hybrid CNN–Transformer models using the LIDC-IDRI dataset. Initial experiments showed that pretrained and from-scratch SegFormer models failed to detect target classes, while fine-tuning achieved IoU scores of 0.6628 and 0.6244 for the right and left lungs, respectively, with corresponding F1-scores of 0.7972 and 0.7688. However, nodule segmentation remained challenging (IoU 0), with limited improvement using a Nodule Loader (IoU 0.1487, F1 0.2589). Compared with recent models reporting Dice scores above 94%, SegFormer demonstrated lower segmentation performance. For classification, standalone ViT, ResNet, and DeepLab achieved accuracies of 72%, 61%, and 82%, respectively, whereas hybrid ResNet+ViT and DeepLab+ViT achieved 98% and 99.54%. Despite these improvements, class imbalance reduced minority-class nodule recall and weighted F1-scores. Overall, fine-tuned SegFormer effectively segments major lung structures but struggles with nodules, while hybrid CNN–Transformer classifiers achieve high accuracy yet require improved class balancing for reliable clinical generalization.*

**Keywords:** classification; CNN; nodules; segmentation; ViT

*How to Cite:* F. Fachruddin, “NoduViT: An Attention-Driven DeepLab–CustomViT Hybrid Model for Segmentation-Informed Multiclass Pulmonary Patch Classification”, *J. RESTI (Rekayasa Sist. Teknol. Inf.)*, vol. 10, no. 5, pp. 1526–1537, Oct. 2026, doi: [10.29207/resti.v10i5.7124](https://doi.org/10.29207/resti.v10i5.7124).

*Permalink/DOI:* <https://doi.org/10.29207/resti.v10i5.7124>

*Received:* August 26, 2025

*Accepted:* June 1, 2026

*Available Online:* October 17, 2026

*This is an open-access article under the [CC BY 4.0 License](#)*

*Published by [Ikatan Ahli Informatika Indonesia](#)*

### 1. Introduction

Lung cancer remains the leading cause of cancer-related deaths worldwide, and early detection is crucial for improving patient outcomes. Pulmonary nodules, which are small masses visible in the lungs on CT scans, are key indicators of early-stage lung cancer [1-3]. Accurate segmentation and classification of these nodules are essential for assessing malignancy, guiding treatment decisions, and reducing unnecessary biopsies [4-8]. However, automated analysis of lung nodules is highly challenging due to their small size, variable shapes, and similarity in intensity to surrounding tissues [1], [9-12]. Manual assessment by radiologists is time-consuming and prone to errors, emphasizing the need for robust deep learning-based solutions. While many prior studies have focused on single-object nodule segmentation, they often ignore the anatomical context

provided by surrounding lung structures [10], [13-16]. Multi-class segmentation that includes both lungs and nodules can offer richer spatial information, enhancing classification accuracy [1], [12], [17], [18]. In recent years, significant advancements have been made in deep learning-based detection and segmentation of pulmonary nodules. Various methods have been proposed, ranging from cascaded networks [10], [19], attention-based AWEU-Net [20], to enhanced 3D V-Net models [10].

While these studies report impressive results achieving IoU scores over 90% and classification accuracies above 95%, most of them suffer from notable limitations [12], [13], [21], [22]. The vast majority of existing approaches are limited to single-object segmentation, primarily focusing on nodules without addressing the anatomical structure of the lungs [18],

[23-25]. Although previous studies have integrated segmentation and classification, segmentation outputs are often limited to preprocessing rather than representation learning. The potential of combining lung anatomy and nodule information within a segmentation-informed hybrid CNN–ViT framework for multiclass patch classification remains underexplored [23], [26-28]. Some studies are also constrained by non-generalizable clinical datasets or computationally expensive 3D models that require substantial computational resources [23], [29]. Interpretability and clinical validation remain challenges. To address these gaps, this study proposes NoduViT, a two-stage segmentation-informed hybrid CNN–ViT framework for multiclass pulmonary patch classification. The framework segments lung structures and nodules from CT slices before classifying extracted patches. Its contributions include evaluating SegFormer configurations, comparing standalone and hybrid classifiers, and assessing class-imbalance effects on nodule recognition.

## 2. Methods

The overall experimental scheme of NoduViT is illustrated in Figure 1. The pipeline begins with annotated CT slices containing four classes: background, right lung, left lung, and pulmonary nodules. Images undergo patch extraction, resizing, normalization, augmentation, and train–test partitioning.

Segmentation is performed using SegFormer variants, ResNet, DeepLab, and DeiT, followed by hybrid ResNet+DeiT and DeepLab+DeiT architectures to integrate convolutional features with attention-based global representations. Performance is evaluated using pixel-level metrics, including IoU, accuracy, precision, recall, and balanced accuracy, alongside classification metrics such as macro F1, weighted F1, and class-specific F1-scores. This pipeline provides a comprehensive evaluation of NoduViT for multiclass pulmonary nodule patch classification.

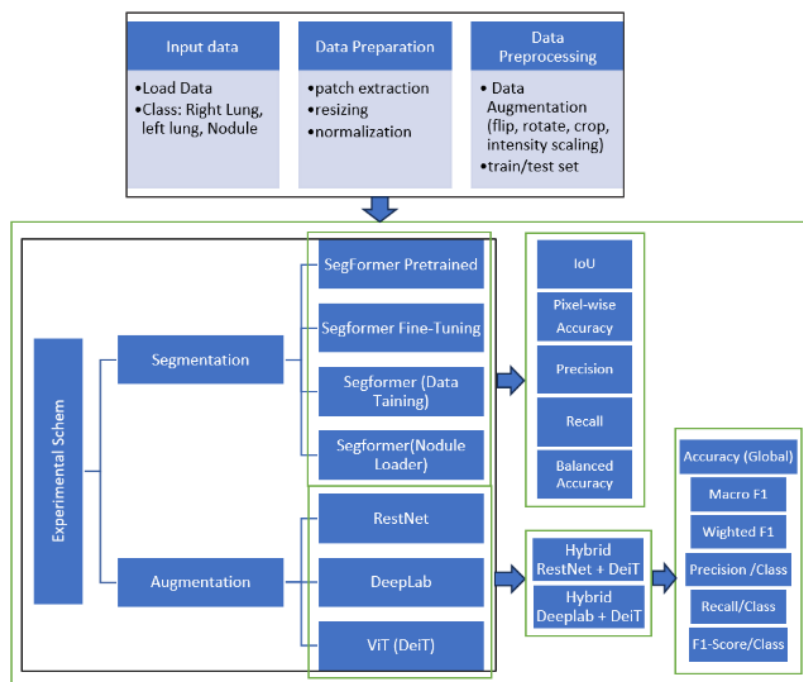


Figure 1. Experimental scheme of NoduViT: workflow of an Attention-Driven DeepLab-CustomViT Hybrid Model for Multiclass Patch Classification of Pulmonary Nodules.

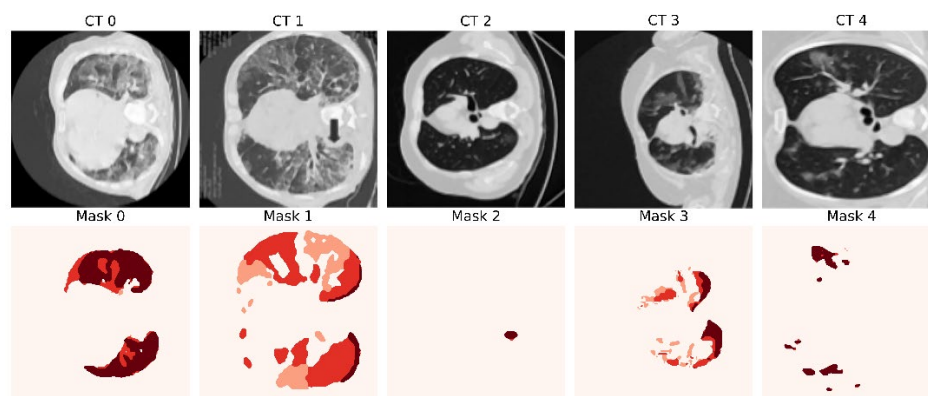


Figure 2. Pulmonary CT scan data snippet and mask visualization with three label types.

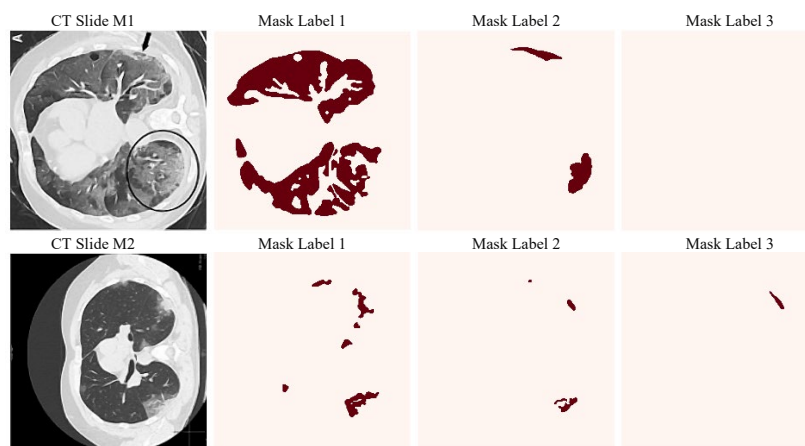


Figure 3. Pulmonary CT scan snippet with label masks illustrated separately.

## 2.1 Dataset Preparation

The CT scan images used in this study were sourced from the LIDC-IDRI (Lung Image Database Consortium and Image Database Resource Initiative) dataset, a widely used public benchmark for lung nodule detection and classification. The dataset consists of thoracic CT scans annotated by multiple expert radiologists and is publicly accessible via The Cancer Imaging Archive (TCIA) [30]. The preparation of CT scan data plays a critical role in ensuring high-quality inputs for the segmentation and classification tasks. Such CT-based lung cancer analysis is widely adopted in recent deep learning studies [8], [10], [14]. Figure 2 presents representative CT slices and corresponding masks of the right lung, left lung, and pulmonary nodules. These voxel-level annotations serve as ground truth for supervised learning (Figure 3), following standard clinical CT orientation [21]. Figure 2 shows representative CT slices and ground-truth masks of the right lung, left lung, and pulmonary nodules from LIDC-IDRI, following standard multiclass segmentation practices [10], [14], where Label 0 = Background, Label 1 = Right Lung, Label 2 = Left Lung, and Label 3 = Lung Nodule.

Figure 3 illustrates CT slices with separated segmentation masks for the right lung (Label 1), left lung (Label 2), and pulmonary nodules (Label 3). As summarized in Table 1, lung positions follow patient-view orientation, while nodules appear as small, sparsely distributed regions. These annotations serve as ground truth for segmentation evaluation and multiclass patch classification in NoduViT.

Table 1. Label Interpretation

| Interpretation | Clarification                                                                                                             |
|----------------|---------------------------------------------------------------------------------------------------------------------------|
| Right Lung     | Appears on the left side of the CT image due to patient-view orientation; label is derived from LIDC-IDRI mask (Label 1). |
| Left Lung      | Appears on the right side of the image; the anatomical view sometimes partially obscured by the heart (Label 2).          |
| Lung Nodule    | Small, localized regions annotated by LIDC-IDRI radiologists (Label 3); may not appear on every slice.                    |

Table 1 summarizes the spatial orientation of already-available ground-truth labels, not manual annotation performed by the authors. In the data preparation stage, several steps were carried out to ensure that the CT scans and their corresponding masks were properly processed and ready for multiclass patch classification:

CT scan slices were loaded along with ground-truth voxel-level annotations provided by the LIDC-IDRI dataset. These annotations were created by expert thoracic radiologists and not manually labeled in this study [8], [10].

The segmentation labels were defined into four classes, as presented in Table 1. Label 0 represents the background, while Label 1 corresponds to the right lung, visible on the left side of the CT image and anatomically located on the patient's right side. Label 2 represents the left lung, visible on the right side of the CT image and often partially obscured by the heart. Finally, Label 3 identifies lung nodules, which appear as small localized regions and may not be visible in every CT slice.

Consistent with findings from previous work on nodule detection, the dataset exhibits a strong class imbalance [10], [14], [22], with Background: 24,394,464 voxels, Right Lung: 1,196,461 voxels, Left Lung: 589,210 voxels, and Nodule: 34,265 voxels. Nodules represent <1% of all voxels, a known challenge in pulmonary CT research. This distribution demonstrates a strong class imbalance, where nodules occupy only a small fraction compared to lung regions and background.

The label distribution is reported at the voxel level, where each voxel represents a pixel in the CT volume. This distribution differs from the classification dataset, which is based on patch-level samples rather than individual voxels. As a result, the number of samples used for classification does not directly correspond to the voxel counts in the segmentation masks. The extremely small proportion of nodule voxels (<1%) indicates a severe class imbalance, which significantly affects both segmentation and classification performance.

Patch-based extraction is performed to capture localized structures following standard practices in lung CT classification and segmentation [8], [10], [15].

All patches were normalized and resized to  $512 \times 512$  pixels to maintain consistent input dimensions across models.

Data were prepared into batch formats that follow PyTorch's tensor convention, with image batches shaped as `torch.Size ([4, 3, 512, 512])` and corresponding masks shaped as `[4, 512, 512]`, similar to established practices in medical image deep learning pipelines [6], [22].

This organization ensures consistent input dimensions and efficient batch processing during model training and evaluation.

Finalized image-mask pairs were prepared for subsequent preprocessing steps, such as data augmentation and training-testing partitioning. In summary, the data preparation stage standardized CT scans and their masks into a structured format suitable for deep learning.

This ensured consistency, highlighted the imbalance between anatomical regions and nodules, and established a robust foundation for further preprocessing and augmentation steps in the experimental pipeline.

## 2.2 Segmentation and Classification

To address the complexity of pulmonary nodule analysis, this study adopts a two-stage framework encompassing segmentation and classification. Multiple segmentation backbones were employed in this study based on their established performance and feature representation capabilities. SegFormer, a transformer-based architecture for efficient semantic segmentation, was implemented using pretrained, fine-tuned, and nodule-focused variants [23], [28].

ResNet served as a CNN baseline due to its residual connections and strong spatial feature extraction capabilities[6]. DeepLab was selected for its ability to capture multi-scale contextual information across lesions of varying sizes [14], [26]. Additionally, Vision Transformer (ViT) and DeiT were employed to capture long-range dependencies through global self-attention, with DeiT incorporating data-efficient training and distillation tokens to enhance performance [28].

For the classification stage, hybrid architectures were employed to fuse convolutional and transformer-based features. Hybrid ResNet + DeiT combines ResNet's local spatial feature extraction with DeiT's global attention to enhance contextual understanding [19], [22]. Meanwhile, Hybrid DeepLab + DeiT integrates DeepLab's multi-scale semantic features with transformer-based attention, forming the core of the proposed NoduViT model [19], [22].

## 2.3 Evaluation Metrics

Model performance was evaluated at two levels. Segmentation-level metrics included IoU, pixel accuracy, precision, recall, and balanced accuracy to assess segmentation quality and class imbalance.

Classification-level metrics included accuracy, macro and weighted F1-scores, and per-class precision, recall, and F1-score to evaluate overall and class-specific performance in multiclass patch classification. These metrics are commonly used in medical image segmentation and classification evaluation [26], [29].

## 2.4 Training Configuration

Table 2 summarizes the training configurations for standalone, hybrid, and segmentation models. All models used the Adam optimizer with learning rates adjusted according to model complexity and stability.

Table 2. Training Configuration of Proposed Models (Optimizer: Adam)

| Model Variant               | Task           | Data Strategy       | Backbone / Initialization            | LR   | Epochs | Batch | Loss Function          | Input Size |
|-----------------------------|----------------|---------------------|--------------------------------------|------|--------|-------|------------------------|------------|
| ResNet                      | Classification | Patch-level samples | Pretrained ResNet                    | 1e-4 | 10     | 4     | Cross-Entropy          | 512×512    |
| DeepLab Classifier          | Classification | Patch-level samples | DeepLabV3-ResNet50 backbone          | 1e-4 | 10     | 4     | Cross-Entropy          | 512×512    |
| Vision Transformer (DeiT)   | Classification | Patch-level samples | Pretrained DeiT                      | 2e-5 | 5      | 4     | Cross-Entropy          | 512×512    |
| ResNet + DeiT               | Hybrid         | Patch-level samples | CNN + DeiT fusion                    | 1e-4 | 10     | 4     | Cross-Entropy          | 512×512    |
| DeepLab + DeiT              | Hybrid         | Patch-level samples | DeepLab + DeiT fusion                | 1e-4 | 10     | 4     | Cross-Entropy          | 512×512    |
| ResNet18 + Custom ViT       | Hybrid         | Patch-level samples | ResNet18 + lightweight ViT           | 2e-5 | 10     | 32    | Cross-Entropy          | 224×224    |
| Hybrid DeepLab + Custom ViT | Hybrid         | Patch-level samples | DeepLabV3-ResNet50 + lightweight ViT | 2e-5 | 10     | 32    | Cross-Entropy          | 224×224    |
| SegFormer (Pretrained)      | Segmentation   | All slices          | ADE20K pretrained                    | 1e-4 | 20     | 4     | Cross-Entropy          | 512×512    |
| SegFormer (Data-Trained)    | Segmentation   | All slices          | Dataset-based initialization         | 1e-4 | 20     | 4     | Cross-Entropy          | 512×512    |
| SegFormer (Nodule Loader)   | Segmentation   | Nodule-only slices  | SegFormer-B0                         | 1e-4 | 20     | 4     | Weighted Cross-Entropy | 512×512    |

Classification models were trained on image patches, while segmentation models used full CT slices or nodule-focused samples. Cross-Entropy loss was applied in most experiments, whereas the Nodule Loader employed Weighted Cross-Entropy to address class imbalance.

Transformer-based and lightweight hybrid models used smaller learning rates and larger batch sizes for stable optimization, while SegFormer configurations required 20 epochs. Pretrained, data-trained, fine-tuned, and nodule-focused SegFormer variants were evaluated to assess the effects of initialization and sampling strategies on pulmonary nodule segmentation.

Overall, Table 2 demonstrates that the proposed framework was evaluated under diverse yet controlled training configurations, enabling a fair comparison between standalone CNN, transformer, and hybrid CNN-ViT architectures. The variation in data strategy, initialization, and optimization settings also supports the analysis of how class imbalance and small-lesion characteristics influence segmentation and classification performance.

All experiments were implemented using Python and the PyTorch deep learning framework in a GPU-based computing environment. The same implementation environment and evaluation protocol were applied consistently across all model variants to ensure fair comparison and reproducibility.

### 3. Results and Discussions

This section reports the performance of the proposed NoduViT model on segmentation and multiclass patch classification. Results are compared with baseline architectures, and findings are discussed in terms of accuracy, robustness, and clinical relevance.

#### 3.1 Visualization Analysis of Segmentation Results

Figure 4 presents representative CT scan slices alongside their predicted masks and overlay visualizations to illustrate the performance of the proposed segmentation model. The visual results in Figure 4 show that the model is able to capture the overall lung structures effectively. However, the delineation of nodular regions remains inconsistent, particularly in small or low-contrast areas.

As shown in Figure 4, the model is able to capture the general boundaries of lung structures with reasonable accuracy. However, discrepancies are observed in regions with low contrast and around vessel-adjacent areas, where the model tends to produce fragmented or imprecise segmentation. In particular, small nodular regions are not consistently detected, indicating a limitation in capturing fine-grained features. This behavior suggests that while the model is effective for large anatomical structures, it remains less reliable for detecting small lesions.

Figure 5 illustrates several failure cases where the model is unable to accurately detect nodular regions.

These failures are primarily associated with small lesion sizes, low contrast between nodules and surrounding tissues, and proximity to vascular structures. In such cases, the model either misses the nodules entirely (false negatives) or produces incomplete segmentation.

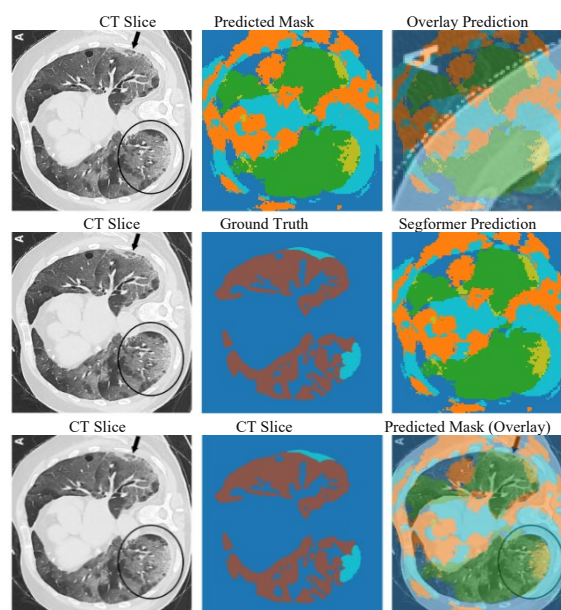
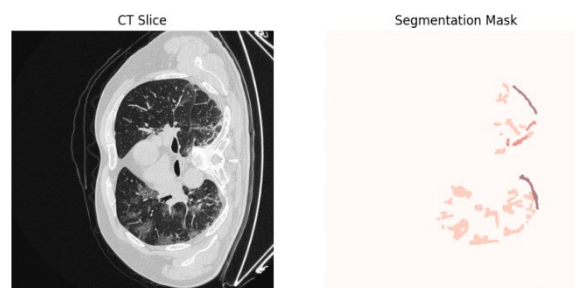


Figure 4. CT scan snippets with predicted masks and overlay visualizations



(a)



(b)

Figure 5. Failure cases in nodule segmentation and classification.

This observation further confirms the challenge of detecting small pulmonary nodules using the current segmentation approach. To further explain these failure

cases, the distribution of nodule voxels across CT slices was analyzed, as presented in Table 3.

Table 3 presents the distribution of nodule voxels across selected CT slices. As shown, even the most nodule-dense slice (Slice 032) contains only 5,561 nodule voxels, which represents a very small proportion of the total image area. This indicates a severe class imbalance at the pixel level, where nodule regions occupy only a minor fraction compared to lung structures and background. Such imbalance makes the learning process biased toward dominant classes and increases the likelihood of false-negative predictions, particularly for small and low-contrast nodules. In addition to voxel-level imbalance, class imbalance is also observed at the patch/sample level, as shown in Table 3.

Table 3. Distribution of Nodule Pixels Across CT Slices

| Slice Index | Nodule Voxels (Label 3) |
|-------------|-------------------------|
| 032         | 5561                    |
| 099         | 4847                    |
| 001         | 3994                    |
| 079         | 3364                    |
| 077         | 1519                    |
| 031         | 1511                    |
| 075         | 1496                    |
| 019         | 1485                    |
| 054         | 1106                    |
| 027         | 1047                    |

Table 4. Class Distribution in LIDC-IDRI Dataset

| Class      | Samples | Percentage |
|------------|---------|------------|
| Right Lung | 187     | 68.75%     |
| Left Lung  | 77      | 28.31%     |
| Nodule     | 8       | 2.94%      |

Table 4 further confirms the class imbalance at the sample level, where nodule patches represent only 2.94% of the dataset. This imbalance directly impacts both segmentation and classification performance, as models tend to prioritize dominant classes such as lung regions. As a result, high overall accuracy can be achieved while still failing to reliably detect nodules.

### 3.2 Segmentation Result of Standalone Model

To evaluate the baseline performance of a standalone segmentation backbone, we first employed the pretrained SegFormer model. Tables 5 and 6 summarize the confusion matrix-based metrics, Intersection-over-Union (IoU), and classification measures obtained from the model.

Table 5. Confusion Matrix-Based Metrics and IoU Results of SegFormer (Pretrained Model)

| Label | TP | TN     | FP    | FN    | IoU |
|-------|----|--------|-------|-------|-----|
| 1     | 0  | 200466 | 0     | 61678 | 0   |
| 2     | 0  | 256749 | 0     | 5395  | 0   |
| 3     | 0  | 228288 | 33856 | 0     | 0   |

Table 6. Classification Metrics of SegFormer (Pretrained Model)

| Label | Acc    | Prec | Recall | F1-Score | Balanced Accuracy |
|-------|--------|------|--------|----------|-------------------|
| 1     | 0.7647 | 0    | 0      | 0        | 0.5               |
| 2     | 0.9794 | 0    | 0      | 0        | 0.5               |
| 3     | 0.8708 | 0    | 0      | 0        | 0.4354            |

The results in Tables 5 and 6 indicate that although SegFormer achieves strong overall accuracy, its ability to detect all classes remains limited, as shown by zero precision and recall for certain labels. Low IoU scores further highlight difficulties in segmenting small nodular regions. These findings suggest that while SegFormer preserves global lung structures, segmentation-informed and imbalance-aware approaches are needed to improve nodule sensitivity and reduce false negatives.

Further experiments trained SegFormer directly on the dataset to evaluate its performance beyond pretrained initialization. The results are presented in Tables 7 and 8, including confusion matrix metrics, IoU, and classification outcomes.

Table 7. Confusion Matrix-Based Metrics and IoU Results of Segformer (Data Training)

| Label | TP | TN      | FP     | FN    | IoU |
|-------|----|---------|--------|-------|-----|
| 1     | 0  | 257,491 | 0      | 4,653 | 0   |
| 2     | 0  | 0       | 0      | 0     | 0   |
| 3     | 0  | 228,288 | 33,856 | 0     | 0   |

Table 8. Classification Metrics of Segformer (Data Training)

| Label | Acc    | Prec | Recall | F1-Score | Balanced Accuracy |
|-------|--------|------|--------|----------|-------------------|
| 1     | 0.9823 | 0    | 0      | 0        | 0.5               |
| 2     | 0      | 0    | 0      | 0        | 0                 |
| 3     | 0.8708 | 0    | 0      | 0        | 0.4354            |

The data training results (Tables 7 and 8) show minimal improvement compared to the pretrained configuration. While accuracy for Label 1 appears high, the precision, recall, and IoU remain at zero across all classes, reflecting a lack of effective positive detection.

Notably, Label 2 could not be classified at all, underscoring the difficulty of capturing nodular patterns solely with a standalone SegFormer model. These findings indicate that standalone transformer segmentation remains insufficient for reliable small-nodule detection under severe class imbalance.

To improve performance, the SegFormer model was fine-tuned on the pulmonary CT dataset. The evaluation results are summarized in Tables 9 and 10, presenting confusion matrix-based metrics with IoU as well as detailed classification measures.

Table 9 Confusion Matrix-Based Metrics and IoU Results of (Fine-Tuning)

| Label | TP     | TN      | FP     | FN     | IoU      |
|-------|--------|---------|--------|--------|----------|
| 1     | 42,747 | 984,081 | 9,189  | 12,559 | 0.662796 |
| 2     | 36,303 | 990,440 | 11,361 | 10,472 | 0.62445  |
| 3     | 11,404 | 971,496 | 25,676 | 40,000 | 0.14795  |

Table 10. Classification Metrics of SegFormer (Fine-tuning)

| Label | Accu racy | Preci sion | Recall | F1-Score | Balanced Accuracy |
|-------|-----------|------------|--------|----------|-------------------|
| 1     | 0.9793    | 0.8231     | 0.7729 | 0.7972   | 0.8818            |
| 2     | 0.9792    | 0.7616     | 0.7761 | 0.7688   | 0.8824            |
| 3     | 0.9374    | 0.3076     | 0.2219 | 0.2578   | 0.5981            |

The fine-tuned model shows clear improvements over pretrained and data-trained configurations. Labels 1 and 2 achieved IoU scores above 0.62 and balanced

accuracy around 0.88, indicating effective segmentation of major lung structures. However, Label 3 (nodules) remained challenging due to its small size, sparse distribution, and low contrast. Despite improvements in lung segmentation, nodule detection remained inconsistent. Mitigation strategies, including patch-based learning, class-imbalance analysis, and increased nodule-containing slices through a Nodule Loader, yielded only limited gains. These findings confirm that fine-tuning improves major lung segmentation, while reliable nodule localization remains challenging due to severe voxel-level imbalance and small-lesion characteristics.

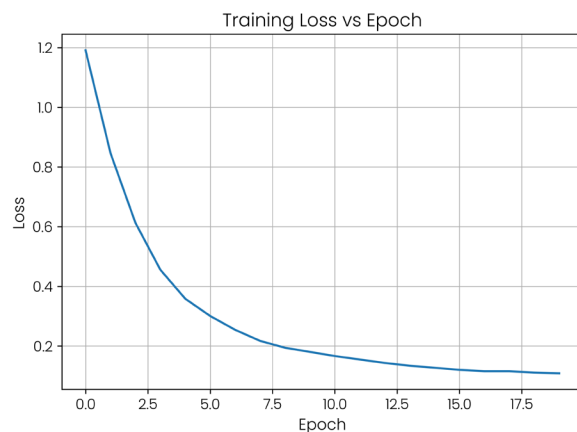


Figure 6. Training loss of Segformer of (Fine-tuning)

As shown in Figure 6, the training loss decreased consistently from an initial value of approximately 1.2 to below 0.15 after 20 epochs, indicating stable convergence of the model. The sharp decline during the first few epochs reflects effective early learning, while the subsequent plateau demonstrates that the network gradually reached optimal parameter adjustment. The absence of significant fluctuations suggests that the optimization process was stable. Nevertheless, since this plot only represents training loss, validation loss and additional evaluation metrics (IoU, F1-score) are required to confirm the model's generalization capability and to ensure that overfitting does not occur.

Finally, a nodule-specific loader was applied to the SegFormer model to emphasize nodule-related regions during training. The outcomes are presented in Tables 11 and 12, detailing confusion matrix-based IoU metrics and classification performance across labels.

Table 11. Confusion Matrix-Based Metrics and IoU Results of Segformer (Nodule Loader)

| Label | TP   | TN        | FP     | FN     | IoU    |
|-------|------|-----------|--------|--------|--------|
| 1     | 6663 | 981,005   | 26,873 | 34,035 | 0.0986 |
| 2     | 8909 | 988,652   | 16,709 | 34,306 | 0.1487 |
| 3     | 0    | 1,043,098 | 0      | 5478   | 0      |

Table 12. Classification Metrics of Segformer (Nodule Loader)

| Label | Acc    | Prec   | Recall | F1-Score | Balanced Accuracy |
|-------|--------|--------|--------|----------|-------------------|
| 1     | 0.9419 | 0.1987 | 0.1637 | 0.1795   | 0.5685            |
| 2     | 0.9513 | 0.3478 | 0.2062 | 0.2589   | 0.5948            |
| 3     | 0.9948 | 0      | 0      | 0        | 0.5               |

The Nodule Loader configuration showed limited improvement under severe class imbalance, with IoU values of 0.0986 and 0.1487 for Labels 1 and 2, respectively. However, precision, recall, and F1-scores remained low, while Label 3 was undetected. Adjustments to learning rate, batch size, loss weighting, and nodule-containing slice sampling yielded only marginal improvements, highlighting the difficulty of small-lesion segmentation using standalone transformer models.

Table 13 compares four SegFormer configurations: pretrained, data training, fine-tuning, and Nodule Loader. Fine-tuning achieved the best overall segmentation performance, with IoU scores of 0.6628 and 0.6244 and F1-scores of 0.7972 and 0.7688 for Labels 1 and 2, respectively. Label 3 achieved lower performance (IoU 0.1480, F1-score 0.2578), but fine-tuning was the only configuration detecting nodular regions to some extent. In contrast, Nodule Loader provided marginal gains for Labels 1 and 2 but failed to detect Label 3. These findings highlight fine-tuning as the most effective SegFormer strategy, although reliable nodule segmentation remains challenging due to severe voxel-level class imbalance.

Table 13. Summary of SegFormer Configurations

| Configuration | IoU                                                    | F1-Score                                               | Overall                                                                                       |
|---------------|--------------------------------------------------------|--------------------------------------------------------|-----------------------------------------------------------------------------------------------|
| Pretrained    | 0 (all labels)                                         | 0 (all labels)                                         | Failed to detect any class                                                                    |
| Data Training | 0 (all labels)                                         | 0 (all labels)                                         | No effective segmentation                                                                     |
| Fine-Tuning   | 0.6628 (Label 1) / 0.6244 (Label 2) / 0.1480 (Label 3) | 0.7972 (Label 1) / 0.7688 (Label 2) / 0.2578 (Label 3) | Best overall segmentation result; strong for major lung structures but still weak for nodules |
| Nodule Loader | 0.0986 (Label 1) / 0.1487 (Label 2) / 0 (Label 3)      | 0.1795 (Label 1) / 0.2589 (Label 2) / 0 (Label 3)      | Limited improvement under severe class imbalance                                              |

### 3.3 Classification Result of Standalone Model

To establish a baseline performance for classification, the ResNet model was first evaluated using global metrics. The results are summarized in Table 14, which reports overall accuracy, Macro F1, and Weighted F1 scores. As shown in Table 14, ResNet achieved an overall accuracy of 0.61 with Macro and Weighted F1 scores of 0.54 and 0.59, respectively.

Table 14. Overall Classification Metrics of ResNet

| Metric      | Value |
|-------------|-------|
| Accuracy    | 0.61  |
| Macro F1    | 0.54  |
| Weighted F1 | 0.59  |

These results indicate moderate performance, suggesting room for improvement, particularly in handling class imbalance. To gain deeper insights into ResNet's classification performance across individual

categories, class-wise metrics were computed. The outcomes are presented in Table 15.

Table 15. Class-wise Classification Metrics of ResNet

| Class      | Precision | Recall | F1-Score | Support |
|------------|-----------|--------|----------|---------|
| Right Lung | 0.70      | 0.76   | 0.73     | 182     |
| Left Lung  | 0.33      | 0.26   | 0.29     | 82      |
| Nodule     | 0.56      | 0.62   | 0.59     | 8       |

From Table 15, it can be observed that the model performs best on the right lung class (F1-score = 0.73), while performance on the left lung class is considerably lower (F1-score = 0.29). Nodule detection shows moderate results (F1-score = 0.59) but is limited by low sample support. These findings highlight ResNet's bias toward dominant classes and the challenge of accurately classifying minority classes such as nodules. The DeepLab model was next evaluated to compare its classification capability with ResNet. The overall performance metrics are reported in Table 15.

Table 16. Overall Classification Metrics of DeepLab

| Metric      | Value |
|-------------|-------|
| Accuracy    | 0.82  |
| Macro F1    | 0.84  |
| Weighted F1 | 0.82  |

As shown in Table 16, DeepLab achieved substantially better results than ResNet, with an accuracy of 0.82 and Macro/Weighted F1 scores of 0.84 and 0.82, respectively. This indicates a strong and balanced classification ability across classes. To further analyze class-specific performance, precision, recall, and F1-scores were calculated for each category, as presented in Table 17.

Table 17. Class-wise Classification Metrics of DeepLab

| Class      | Precision | Recall | F1-Score | Support |
|------------|-----------|--------|----------|---------|
| Right Lung | 0.95      | 0.77   | 0.85     | 927     |
| Left Lung  | 0.63      | 0.92   | 0.75     | 392     |
| Nodule     | 1.00      | 0.86   | 0.93     | 37      |

According to Table 17, DeepLab achieved its highest performance in nodule detection (F1-score = 0.93), demonstrating strong sensitivity despite the limited support of 37 samples. The right lung class also showed high precision (0.95) with an F1-score of 0.85, while the left lung class yielded a lower precision (0.63) but very high recall (0.92), resulting in an F1-score of 0.75.

These results suggest that DeepLab provides more balanced multiclass classification performance compared to ResNet under the evaluated experimental setting. The performance of the Vision Transformer (DeiT) model was then assessed to investigate its effectiveness in multiclass patch classification. The overall results are provided in Table 18.

Table 18. Overall Classification Metrics of ViT (DeiT)

| Metric      | Value |
|-------------|-------|
| Accuracy    | 0.72  |
| Macro F1    | 0.59  |
| Weighted F1 | 0.64  |

As shown in Table 18, ViT (DeiT) achieved an overall accuracy of 0.72, with a Macro F1 of 0.59 and a

Weighted F1 of 0.64. These values suggest moderate performance, indicating that while the model captures global context effectively, its class balance remains suboptimal. A detailed breakdown of class-specific performance for ViT (DeiT) is presented in Table 19.

Table 19. Class-wise Classification Metrics of ViT (DeiT)

| Class      | Precision | Recall | F1-Score | Support |
|------------|-----------|--------|----------|---------|
| Right Lung | 0.72      | 0.99   | 0.84     | 187     |
| Left Lung  | 0.67      | 0.08   | 0.14     | 77      |
| Nodule     | 0.86      | 0.75   | 0.80     | 8       |

According to Table 19, the right lung class achieved the highest recall (0.99) with an F1-score of 0.84, reflecting strong sensitivity. The nodule class also showed promising results with an F1-score of 0.80, despite its limited support. However, the left lung class performed poorly (F1-score = 0.14), with recall as low as 0.08, suggesting significant difficulty in capturing this region. These outcomes highlight both the potential and the limitations of ViT (DeiT) when applied independently for multiclass lung patch classification.

To evaluate the effectiveness of the proposed NodeViT: A Segmentation-Informed Hybrid CNN-ViT Framework for Multiclass Pulmonary Patch Classification, we first present the classification results obtained from the hybrid ResNet and ViT (DeiT) configuration. The performance was assessed across 10 epochs, with metrics focused on training loss reduction and accuracy improvement. As shown in Table 20, the Hybrid ResNet + ViT model demonstrates a steady improvement throughout training. The loss decreases from 28.74 at epoch 1 to 3.40 at epoch 10, while the training accuracy gradually increases to 0.9677.

Table 20. Classification Results of Hybrid ResNet and ViT (DeiT)

| Epoch | Loss    | Accuracy |
|-------|---------|----------|
| 1     | 28.7361 | 0.6304   |
| 2     | 22.4741 | 0.6977   |
| 3     | 21.2037 | 0.7014   |
| 4     | 17.2539 | 0.7770   |
| 5     | 12.7311 | 0.8470   |
| 6     | 6.6787  | 0.9226   |
| 7     | 5.1563  | 0.9493   |
| 8     | 4.2683  | 0.9631   |
| 9     | 3.8905  | 0.9650   |
| 10    | 3.4030  | 0.9677   |

These results indicate that the hybrid architecture can effectively learn discriminative multiclass patch representations during training.

As shown in Table 21, the Hybrid DeepLab + ViT model converges faster than the Hybrid ResNet + ViT model. The training loss decreases from 24.44 at epoch 1 to 1.49 at epoch 10, while the training accuracy reaches a peak of 0.9935 at epoch 7 and remains stable above 0.98 during the final epochs.

These findings suggest that DeepLab-derived spatial representations contribute to faster optimization and more stable convergence during training. However, this high training accuracy should be interpreted together with the class-wise evaluation results in Tables 22 and

23, since overall accuracy alone may overestimate performance in imbalanced datasets.

Table 21. Classification Results of Hybrid DeepLab and ViT (DeiT)

| Epoch | Loss    | Accuracy |
|-------|---------|----------|
| 1     | 24.4380 | 0.6799   |
| 2     | 18.2039 | 0.7408   |
| 3     | 7.1450  | 0.9253   |
| 4     | 4.5185  | 0.9511   |
| 5     | 1.8306  | 0.9852   |
| 6     | 1.2263  | 0.9899   |
| 7     | 0.8375  | 0.9935   |
| 8     | 1.2275  | 0.9926   |
| 9     | 1.5617  | 0.9852   |
| 10    | 1.4862  | 0.9954   |

As shown in Table 22, the Hybrid DeepLab + ViT model achieved an overall test accuracy of 0.65, with macro and weighted F1-scores of 0.55 and 0.63, respectively.

Table 22. Classification Performance of Hybrid DeepLab + ViT

| Class        | Precision | Recall | F1-Score | Support |
|--------------|-----------|--------|----------|---------|
| Right Lung   | 0.71      | 0.81   | 0.75     | 178     |
| Left Lung    | 0.46      | 0.33   | 0.39     | 87      |
| Nodule       | 0.60      | 0.43   | 0.50     | 7       |
| Accuracy     |           |        | 0.65     | 272     |
| Macro Avg    | 0.59      | 0.52   | 0.55     | 272     |
| Weighted Avg | 0.62      | 0.65   | 0.63     | 272     |

The model achieved stable performance for the Right Lung class, with recall of 0.81 and F1-score of 0.75. However, the Left Lung and Nodule classes remained more difficult to recognize, particularly the Nodule class, which achieved recall of only 0.43 with support of seven samples. These results indicate that severe class imbalance and the small size of nodules continue to limit reliable minority-class recognition.

As shown in Table 23, the Hybrid ResNet + ViT model achieved an overall test accuracy of 0.68, with macro and weighted F1-scores of 0.67. The model achieved its best performance on the Nodule class, reaching precision of 1.00, recall of 0.71, and F1-score of 0.83. Nevertheless, this result should be interpreted

cautiously because the nodule class contains only seven test samples. The Left Lung class remained the most challenging category, indicating that anatomical asymmetry and visual similarity across lung regions still affect classification performance.

Table 23. Classification Performance of Hybrid ResNet + ViT

| Class        | Precision | Recall | F1-Score | Support |
|--------------|-----------|--------|----------|---------|
| Right Lung   | 0.76      | 0.81   | 0.78     | 187     |
| Left Lung    | 0.43      | 0.36   | 0.39     | 77      |
| Nodule       | 1.00      | 0.71   | 0.83     | 7       |
| Accuracy     |           |        | 0.68     | 271     |
| Macro Avg    | 0.73      | 0.63   | 0.67     | 271     |
| Weighted Avg | 0.67      | 0.68   | 0.67     | 271     |

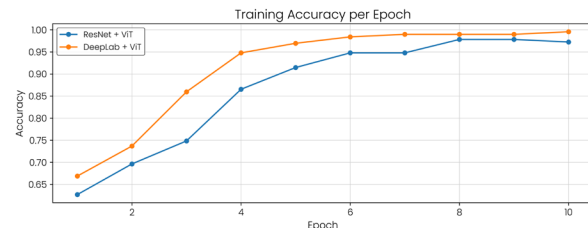


Figure 7. Comparison of Training Accuracy per Epoch of Hybrid Model (ResNet+ ViT and Deeplab+ViT)

Figure 7 compares the training accuracy trends of the proposed hybrid models. Both models demonstrate consistent improvement across epochs, indicating successful convergence during training. The DeepLab + ViT model reaches high accuracy earlier and maintains stable performance above 0.98 during the final epochs, whereas the ResNet + ViT model improves more gradually and reaches 0.9677 at epoch 10. These results suggest that DeepLab-derived spatial features contribute to faster optimization and more stable convergence behavior.

As shown in Table 24, the standalone and hybrid models exhibit different performance characteristics. DeepLab achieved the most balanced standalone classification performance under the evaluated experimental setting, likely due to its ability to capture multiscale contextual information.

Table 24. Comparison of Classification Results of Standalone and Hybrid Models

| Model                | Best Training Accuracy | Macro F1 | Weighted F1 | Details                                                               |
|----------------------|------------------------|----------|-------------|-----------------------------------------------------------------------|
| ViT Only             | 0.72                   | 0.59     | 0.64        | Left lung is difficult to recognize                                   |
| ResNet               | 0.61                   | 0.54     | 0.59        | Similar to ViT, but weaker                                            |
| DeepLab              | 0.82                   | 0.84     | 0.82        | Most balanced across classes                                          |
| Hybrid ResNet + ViT  | 0.9677                 | 0.67     | 0.67        | Better nodule-class recognition                                       |
| Hybrid DeepLab + ViT | 0.9954                 | 0.63     | 0.59        | Faster convergence but minority-class recognition remains challenging |

The hybrid models achieved substantially higher training accuracy, with Hybrid DeepLab + ViT reaching 0.9954 and Hybrid ResNet + ViT reaching 0.9677. However, the macro and weighted F1-scores indicate that minority-class recognition remains challenging under severe class imbalance conditions. These findings demonstrate that overall accuracy alone is insufficient for evaluating model reliability in multiclass medical imaging tasks.

Figure 8 presents the training loss curves of the hybrid models across 10 epochs. Both models show a consistent reduction in loss values, indicating effective optimization during training. The DeepLab + ViT model demonstrates faster convergence, reducing the loss from 24.44 to 1.49, while the ResNet + ViT model decreases from 28.74 to 3.40. These findings further confirm that DeepLab-derived spatial representations

provide more efficient optimization behavior during training.

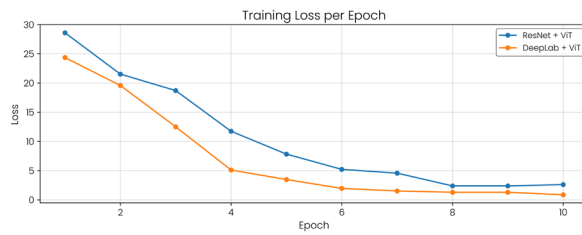


Figure 8. Comparison of Training Loss per Epoch of Hybrid Model (ResNet+ ViT and Deeplab+ViT)

Table 25. Per-class Performance of Hybrid Models

| Model       | Class  | Precision | Recall | F1   |
|-------------|--------|-----------|--------|------|
| ViT         | Nodule | 0.86      | 0.75   | 0.80 |
| ResNet+ViT  | Nodule | 1.00      | 0.71   | 0.83 |
| DeepLab+ViT | Nodule | 0.60      | 0.43   | 0.50 |

As shown in Table 25, the ResNet+ViT model achieved the strongest nodule-class recognition performance among the evaluated hybrid models, with precision, recall, and F1-score values of 1.00, 0.71, and 0.83, respectively.

In contrast, although DeepLab+ViT demonstrated faster convergence and higher training accuracy, its nodule-class performance remained lower, with an F1-score of 0.50. These findings indicate that minority-class recognition remains challenging under severe class imbalance conditions.

Overall, the experimental findings reveal different strengths between the two hybrid architectures. DeepLab+ViT achieved superior training convergence and higher overall training accuracy, whereas ResNet+ViT demonstrated stronger minority-class robustness during test-set evaluation, particularly for the nodule class.

### 3.5 Benchmarking of the Segmentation and Classification in research

To validate the robustness of the proposed method, a comparative benchmarking analysis was conducted against several state-of-the-art studies. Table 25 summarizes recent segmentation models from different references, including both conventional CNN-based approaches and transformer-enhanced architectures. These benchmarks serve as a baseline to assess the effectiveness of our fine-tuned SegFormer model.

The results in Table 26 show that CNN-based and enhanced U-Net-style architectures achieve strong performance for nodule-only segmentation tasks. In contrast, although the fine-tuned SegFormer achieved competitive IoU values for major lung structures, reliable nodule segmentation remained difficult.

These findings indicate that transformer-based segmentation models still face challenges in detecting small pulmonary nodules under severe class imbalance conditions. To further demonstrate the effectiveness of the proposed hybrid models, we compared them with

recently reported classification approaches in lung cancer research.

Table 26. Benchmarking of Segmentation Results

| Method & Reference                 | Dataset          | Segmentation Target           | Evaluation Metrics                                      |
|------------------------------------|------------------|-------------------------------|---------------------------------------------------------|
| Improved V-Net with Attention [31] | LUNA16, LNDb     | Nodule only                   | Dice $\approx$ 94.9%, Sensitivity $\approx$ 81–92%      |
| C+EffxNet [19]                     | Hospital dataset | Nodule only                   | Accuracy $\sim$ 98%                                     |
| ResNodNet [15]                     | Public CT        | Nodule only                   | Accuracy $\sim$ 98.6%                                   |
| SegFormer (fine-tuned, our model)  | LIDC-IDRI Subset | Right Lung, Left Lung, Nodule | IoU: 0.66 (Right Lung), 0.62 (Left Lung), 0.15 (Nodule) |

As shown in Table 27, the proposed hybrid architectures achieved competitive classification performance, with Hybrid DeepLab + Custom ViT reaching the highest best training accuracy of 0.9954. However, severe class imbalance may inflate overall accuracy, making class-wise precision, recall, and F1-score essential for evaluating nodule recognition. Despite strong patch-classification capability, lower minority-class performance and remaining false negatives highlight ongoing challenges. Comparisons with prior studies should also be interpreted cautiously, as most focus on single-object nodule analysis, whereas this framework integrates multiclass lung and nodule segmentation with classification.

Table 27. Benchmarking of Classification Results

| Method & Reference                      | Dataset                                 | Task                 | Evaluation                     |
|-----------------------------------------|-----------------------------------------|----------------------|--------------------------------|
| TPOT + SVM                              | CT Images                               | Classification       | Accuracy $\approx$ 91.8%       |
| Hybrid ML [22]                          |                                         |                      |                                |
| C+EffxNet [19]                          | Hospital dataset                        | Classification       | Accuracy $\approx$ 98.0%       |
| ResNodNet [15]                          | Public CT                               | Classification       | Accuracy $\approx$ 98.6%       |
| Hybrid DeepLab + Custom ViT (our model) | LIDC-IDRI Subset (Achebak et al., 2020) | Patch Classification | Best Accuracy $\approx$ 99.54% |
| Hybrid ResNet + Custom ViT (our model)  | LIDC-IDRI Subset                        | Patch Classification | Best Accuracy $\approx$ 96.77% |

These findings demonstrate the potential of segmentation-informed hybrid CNN-ViT classification, although further validation using balanced datasets, nodule-focused augmentation, and clinically representative cohorts is needed. Compared with specialized nodule segmentation models reporting Dice scores above 90% [10], [18], [19], SegFormer showed lower nodule segmentation performance due to small lesion size and class imbalance. Nevertheless, hybrid classifiers achieved competitive performance by combining convolutional and transformer-based representations.

#### 4. Conclusions

This study evaluated SegFormer configurations and hybrid deep learning models for lung segmentation and patch-based classification using the LIDC-IDRI dataset. Fine-tuned SegFormer achieved the best segmentation performance for major lung structures, with IoU values of 0.6628 (Right Lung) and 0.6244 (Left Lung), while nodule segmentation remained limited (IoU 0.1479). The Nodule Loader provided only marginal improvement, and transformer-only segmentation underperformed compared with prior CNN-based and hybrid approaches. For classification, standalone ViT and ResNet showed moderate performance, while DeepLab achieved more balanced results. Hybrid ResNet+ViT and DeepLab+ViT achieved best training accuracies of 0.9677 and 0.9954, respectively. However, minority-class nodule recognition remained challenging despite high overall accuracy. NoduViT demonstrates promising multiclass patch-classification capability through hybrid feature fusion, although limitations persist in nodule segmentation, class imbalance, and its non-end-to-end pipeline. Future work should prioritize lesion-focused segmentation, class balancing, and tighter integration of segmentation and classification.

#### Declarations

##### AI Usage Statement

The authors used ChatGPT (OpenAI) for language refinement and editorial assistance during manuscript preparation. The authors reviewed and verified the manuscript and retain full responsibility for its scientific content, experimental analyses, interpretation of results, and conclusions.

##### Author Contributions

Fachruddin: investigation, supervision; Yovi Pratama: conceptualization, methodology, software; Errissya Rasywir: methodology, formal analysis, validation; Yessi Hartiwi: project administration; Jasmir: data curation, validation; Hastie Audytra: project administration, formal analysis; Marrylinteri Istoningtyas: project administration, formal analysis.

##### Funding Statement

This research was supported by the Directorate of Higher Education, Ministry of Education, Culture, Research, and Technology of the Republic of Indonesia under the Fundamental Research Grant No. 036/LL10/DT.05.00/PL/2025.

##### Conflict of Interest Statement

The authors declare no conflict of interest.

##### Ethical Clearance

This study utilized the publicly available, de-identified Lung Image Database Consortium and Image Database Resource Initiative (LIDC-IDRI) dataset, accessed through The Cancer Imaging Archive (TCIA). No new

human participants were recruited, no additional clinical data were collected, and no animal experiments were conducted. The research involved secondary analysis of publicly available, de-identified medical imaging data.

##### Data and Code Availability

The dataset analyzed in this study is publicly available through The Cancer Imaging Archive (TCIA) under the Lung Image Database Consortium and Image Database Resource Initiative (LIDC-IDRI) collection. The implementation code and related experimental materials are available from the corresponding author upon reasonable request.

#### References

- [1] N. Ağuloğlu, A. Aksu, D. S. Unat, and Ö. Selim Unat, "The value of PET/CT radiomic texture analysis of primary mass and mediastinal lymph node on survival in patients with non-small cell lung cancer," *Revista Española de Medicina Nuclear e Imagen Molecular (English Edition)*, vol. 43, no. 5, p. 500027, Sep. 2024, doi: <https://doi.org/10.1016/j.remnie.2024.500027>.
- [2] A. Ayeni, O. Evbuomwan, and M.-D.-T. W. Vangu, "The Role of [18F]FDG PET/CT in Monitoring of Therapy Response in Lung Cancer," *Seminars in Nuclear Medicine*, vol. 55, no. 2, pp. 175–189, Mar. 2025, doi: <https://doi.org/10.1053/j.semnuclmed.2025.02.002>.
- [3] B. Cangut, R. Akinlusi, A. Mohseny, N. Ghesani, and M. Ghesani, "Evolving Paradigms in Lung Cancer: Latest Trends in Diagnosis, Management, and Radiopharmaceuticals," *Seminars in Nuclear Medicine*, vol. 55, no. 2, pp. 264–276, Mar. 2025, doi: <https://doi.org/10.1053/j.semnuclmed.2025.02.005>.
- [4] P. Agustiansyah et al., "Automated Precancerous Lesion Screening Using an Instance Segmentation Technique for Improving Accuracy," *Sensors*, vol. 22, no. 15, p. 5489, Jul. 2022, doi: <https://doi.org/10.3390/s22155489>.
- [5] A. W. Arum, S. Nurmaini, D. P. Rini, P. Agustiansyah, and M. N. Rachmatullah, "Segmentation of Squamous Columnar Junction on VIA Images using U-Net Architecture," *ComEngApp*, vol. 10, no. 3, pp. 209–219, Oct. 2021, doi: <https://doi.org/10.18495/comengapp.v10i3.387>.
- [6] N. Duklan, S. Kumar, H. Maheshwari, R. Singh, S. D. Sharma, and S. Swami, "CNN Architectures for Image Classification: A Comparative Study Using ResNet50V2, ResNet152V2, InceptionV3, Xception, and MobileNetV2," *SSRG-IJECE*, vol. 11, no. 9, pp. 11–21, Sep. 2024, doi: <https://doi.org/10.14445/23488549/IJECE-V11I9P102>.
- [7] H. Irshad, A. Veillard, L. Roux, and D. Racoceanu, "Methods for Nuclei Detection, Segmentation, and Classification in Digital Histopathology: A Review—Current Status and Future Potential," *IEEE Reviews in Biomedical Engineering*, vol. 7, pp. 97–114, 2014, doi: <https://doi.org/10.1109/RBME.2013.2295804>.
- [8] G. Geethu Lakshmi and P. Nagaraj, "Lung cancer detection and classification using Optimized CNN features and Squeeze-Inception-ResNeXt model," *Computational Biology and Chemistry*, p. 108437, Mar. 2025, doi: <https://doi.org/10.1016/j.compbiolchem.2025.108437>.
- [9] Z. Hou et al., "Overestimation of contralateral hilar lymph node metastasis in non-metastatic non-small cell lung cancer and its predictive model: HAM," *Radiotherapy and Oncology*, vol. 201, p. 110575, Dec. 2024, doi: <https://doi.org/10.1016/j.radonc.2024.110575>.
- [10] S. S. Nair, V. N. M. Devi, and S. Bhasi, "Lung Cancer Detection from CT Images: Modified Adaptive Threshold Segmentation with Support Vector Machines and Artificial Neural Network Classifier," *Current Medical Imaging*, vol. 20, Jan. 2024, doi: <https://doi.org/10.2174/1573405620666230714110914>.

- [11] J. Park *et al.*, "Development of a multi-modal learning-based lymph node metastasis prediction model for lung cancer," *Clinical Imaging*, vol. 114, p. 110254, Oct. 2024, doi: <https://doi.org/10.1016/j.clinimag.2024.110254>.
- [12] S. Vinjamuri and V. Pant, "Demystifying the Role of Immuno PET-CT in Non-Small Cell Lung Cancer: Clinical Value and Research Trends," *Seminars in Nuclear Medicine*, vol. 55, no. 2, pp. 212–220, Mar. 2025, doi: <https://doi.org/10.1053/j.semnucmed.2025.02.004>.
- [13] Q. Wan, L. Bai, Z. Wang, and B. Ji, "Diagnostic performance of FAPI PET/CT for the detection of lymph node metastases in lung cancer patients: a meta-analysis," *Academic Radiology*, Oct. 2024, doi: <https://doi.org/10.1016/j.acra.2024.09.026>.
- [14] N. Velmurugan, R. Rajeswari, S. Naganjaneyulu, and A. Anupama, "Rat swarm political optimizer based deep learning approach for lung lobe segmentation and lung cancer detection using CT images," *Biomedical Signal Processing and Control*, vol. 105, p. 107612, Jul. 2025, doi: <https://doi.org/10.1016/j.bspc.2025.107612>.
- [15] L. R. Vadala, M. Das, C. Raga Madhuri, and S. Merugula, "SpiLenet based detection and severity level classification of lung cancer using CT images," *Computers and Electrical Engineering*, vol. 123, p. 110036, Apr. 2025, doi: <https://doi.org/10.1016/j.compeleceng.2024.110036>.
- [16] E. Rasywir *et al.*, "Robust SegFormer-Based Cervical Neoplastic Lesion Segmentation Using Lesion-Specific Augmentation and Anatomical Label Optimization," *IEEE Access*, vol. 13, pp. 205915–205931, 2025, doi: <https://doi.org/10.1109/ACCESS.2025.3639572>.
- [17] F. Öztürk Ergür, A. Öztürk, Ö. Özdağ, E. Tatcı, Ö. Özmen, and A. Yılmaz, "When to Consider Invasive Lymph Node Staging in Non-Small-Cell Lung Cancer? A Novel Scoring System Utilising Metabolic Parameters in 18F-FDG PET/CT," *Archivos de Bronconeumología*, vol. 60, pp. S4–S12, Oct. 2024, doi: <https://doi.org/10.1016/j.arbres.2024.05.020>.
- [18] Y. H. Yang, H. E. Kim, B. J. Park, J. G. Lee, D. J. Kim, and C. Y. Lee, "Patterns and preoperative risk factors of occult lymph node metastasis in clinical stage I lung cancer," *Lung Cancer*, vol. 202, p. 108461, Apr. 2025, doi: <https://doi.org/10.1016/j.lungcan.2025.108461>.
- [19] P. Shyamala Bharathi and C. Shalini, "Advanced hybrid attention-based deep learning network with heuristic algorithm for adaptive CT and PET image fusion in lung cancer detection," *Medical Engineering & Physics*, vol. 126, p. 104138, Apr. 2024, doi: <https://doi.org/10.1016/j.medengphy.2024.104138>.
- [20] F. Li, W. Han, and H. Sun, "Serum miR-155-5p assists the diagnostic sensitivity and accuracy of low-dose spiral CT imaging in early lung cancer," *Molecular and Cellular Probes*, vol. 78, p. 101994, Dec. 2024, doi: <https://doi.org/10.1016/j.mcp.2024.101994>.
- [21] C. Richlitzki *et al.*, "Advances of PET/CT in Target Delineation of Lung Cancer Before Radiation Therapy," *Seminars in Nuclear Medicine*, vol. 55, no. 2, pp. 190–201, Mar. 2025, doi: <https://doi.org/10.1053/j.semnucmed.2025.02.013>.
- [22] N. N. Murthy and K. Thippeswamy, "TPOT with SVM hybrid machine learning model for lung cancer classification using CT image," *Biomedical Signal Processing and Control*, vol. 104, p. 107465, Jun. 2025, doi: <https://doi.org/10.1016/j.bspc.2024.107465>.
- [23] C. Raghuram and M. Thenmozhi, "Short Review on Contrastive Learning-based Segmentation Techniques for Medical Image Processing," in *2023 International Conference in Advances in Power, Signal, and Information Technology (APSIT)*, Jun. 2023, pp. 290–296, doi: <https://doi.org/10.1109/APSIT58554.2023.10201707>.
- [24] B. F. Smith, B. L. Scott, and P. J. Michaels, "Metastatic seminoma presenting in an external iliac lymph node in an asymptomatic elderly patient undergoing screening in a multi-cancer early detection study: Report of a case and review of the literature," *Human Pathology Reports*, p. 40, p. 300772, Jun. 2025, doi: <https://doi.org/10.1016/j.hpr.2025.300772>.
- [25] Y. Li *et al.*, "Radiomics analysis of 18F-FDG PET/CT for visceral pleural invasion in non-small cell lung cancer with pleural attachment," *Clinical Radiology*, p. 106867, Mar. 2025, doi: <https://doi.org/10.1016/j.crad.2025.106867>.
- [26] A. M. Alhassan, "Driving Training-Based Optimization-Multitask Fuzzy C-Means (DTBO-MFCM) Image Segmentation and Robust Deep Learning Algorithm for Multicenter Breast Histopathological Images," *IEEE Access*, vol. 11, pp. 136350–136360, 2023, doi: <https://doi.org/10.1109/ACCESS.2023.3335667>.
- [27] B. D. Basaran, W. Zhang, M. Qiao, B. Kainz, P. M. Matthews, and W. Bai, "LesionMix: A Lesion-Level Data Augmentation Method for Medical Image Segmentation," Aug. 17, 2023, *arXiv*: 2308.09026, doi: <https://doi.org/10.48550/arXiv.2308.09026>.
- [28] M. Khalil, A. Khalil, and A. Ngom, "A Comprehensive Study of Vision Transformers in Image Classification Tasks," Dec. 05, 2023, *arXiv*: arXiv:2312.01232, doi: <https://doi.org/10.48550/arXiv.2312.01232>.
- [29] C. Chodvadiya, N. Mahla, K. G. Singh, and K. Sharad Jadhav, "FESS Loss: Feature-Enhanced Spatial Segmentation Loss for Optimizing Medical Image Analysis," in *2024 IEEE International Symposium on Biomedical Imaging (ISBI)*, May 2024, pp. 1–5, doi: <https://doi.org/10.1109/ISBI56570.2024.10635209>.
- [30] O. Leleu *et al.*, "Lung cancer screening by low-dose CT scan in France: final results of the DEP KP80 study after three rounds," *eBioMedicine*, vol. 109, p. 105396, Nov. 2024, doi: <https://doi.org/10.1016/j.ebiom.2024.105396>.
- [31] X. Ma, H. Song, X. Jia, and Z. Wang, "An improved V-Net lung nodule segmentation model based on pixel threshold separation and attention mechanism," *Sci Rep*, vol. 14, no. 1, p. 4743, Feb. 2024, doi: <https://doi.org/10.1038/s41598-024-55178-3>.