**Equivalent: Jurnal Ilmiah Sosial Teknik**
**Volume 8, Issue 2, 395-411**
**e_ISSN: 2775-0833**
https://jurnalequivalent.id/index.php/jequi/index
DOI: doi.org/10.59261/ jequi.v8i2.302

# Lung Nodule Segmentation Accuracy in CT Images Using YOLO, 3D-CNN, and Ensemble ViT-UNETR U-Net

**Reyga Ferdiansyah Putra[1*]**
Universitas Bina Nusantara, Indonesia

**Antoni Wibowo[2]**
Universitas Bina Nusantara, Indonesia

**Dewi Retno Sari Saputro[3]**
Universitas Sebelas Maret, Indonesia

***Corresponding author:**
Reyga Ferdiansyah Putra, Universitas Bina Nusantara, Indonesia. ✉reygafp@gmail.com

**Article Info:**
**Article history:**
Received: March 27, 2026
Revised: May 15, 2026
Accepted: May 19, 2026

**Keywords:**
Lung Nodules; Hybrid Segmentation; YOLOv12; 3D-CNN; ViT-UNETR; U-Net; Ensemble Learning.

**Abstract**

**Background:** Lung cancer is the leading cause of cancer-related mortality globally, with over 2.2 million new cases and 1.8 million deaths reported annually (WHO, 2022). Pulmonary nodule detection through low-dose computed tomography (LDCT) screening is the most effective method for early lung cancer identification. However, automated systems still face significant challenges: high false positive rates, limited sensitivity for micronodules (<5 mm), and poor segmentation accuracy for nodules with irregular morphology or juxtapleural attachment.
**Objective:** Lung nodules early discovery is key to treating lung carcinoma, but even conventional systems' micronodules still have high false positives and low accuracy.
**Method:** This study presents an end-to-end hybrid pipeline that uses the LUNA16 database to tackle this issue. The initial stage is to make use of YOLOv12 for Region of Interest (ROI) extraction, with 3D-CNN carrying out false positive filtering through volumetric verification as a gate. The final phase conducts pixel-level precision segmentation using Adaptive Bayesian Fusion on U-Net Residual 3D ensemble (local texture features) and ViT-UNETR (global anatomical context).
**Results**: Experiments showed superior performance level 99.99% Accuracy, Mean Dice Similarity Coefficient (DSC) at 93.88% and IoU is 90.45%. The system was very robust, reaching 97.33% DSC in the micro nodule category (<5 mm).
**Conclusion:** In summary, this integrated architecture delivers an objective, efficient and high-quality solution for automated Diagnosis.

**To cite this article:** Putra, R. F., Wibowo, A., & Saputro, D. R. S. (2026). Lung Nodule Segmentation Accuracy in CT Images Using YOLO, 3D-CNN, and Ensemble ViT-UNETR U-Net. *Equivalent: Jurnal Ilmiah Sosial Teknik*, 8(2), 395-411. https://doi.org/10.59261/jequi.v8i2.308

## INTRODUCTION

According to GLOBOCAN 2022, lung cancer is the most commonly diagnosed cancer and the leading cause of cancer mortality worldwide, accounting for approximately 2.21 million new cases (11.4% of total cancer incidence) and 1.80 million deaths (18.0% of total cancer mortality) annually (Sung et al., 2021). In the United States, an estimated 238,340 new lung cancer cases and 125,070 deaths are projected for 2023, with the five-year survival rate remaining critically low at approximately 22% due to the predominance of late-stage diagnoses at the time of detection. In China, lung cancer accounts for 26.4% of all oncology diagnoses and represents the highest national cancer burden. Low-dose CT (LDCT) screening has demonstrated up to 20% reduction in lung cancer mortality compared to conventional chest X-ray screening (Sung et al., 2021; Team, 2019), underscoring the urgent clinical need for accurate, automated pulmonary nodule detection

and segmentation systems.

Ahmadyar et al. (2022) was able to combine YOLOv5s for the first stage of detection and 3D-CNN for classification, which increased accuracy to 98.4% while at the same time lowering false positives significantly. AUC889% yields an area under curve of 98.9 percent, when using the latter EditorialWall-II. Furthermore, as a volume-based preprocessing step, Kadia et al. (2021) built R2U3D (Recurrent Residual 3D U-Net) for lung organ segmentation. They found that it is very effective in capturing complex spatial details even with Soft-DSC reaching up to 99.20%. These results show once more that combining fast detection methods like YOLO with 3D architectures for detection of pulmonary micronodules in volumetric CT data is a promising direction to go. Specifically, no prior work has simultaneously addressed the critical triad of challenges in lung nodule analysis: (1) rapid candidate localization with minimal missed detections, (2) systematic elimination of false positives through volumetric verification, and (3) high-precision boundary delineation through complementary local-global feature fusion. The proposed pipeline uniquely addresses all three challenges in an end-to-end automated framework validated on the standardized LUNA16 benchmark. has yet specifically combined the advantages of U-Net in capturing local features and ViT-UNETR in understanding global context into one unified system for lung nodule segmentation with high precision.

There have been a number of studies also exploring ensemble or multi-task approaches to improve lung nodule segmentation. The successful variant AWEU-Net introduced by Banu et al. (2021) rolled back the U-Net to add head parts and channels attention modules (PAWE & CAWE) inclusions while PaDSC on LUNA16 was 89.79%. (Annavarapu et al., 2023) proposed U-Det, a U-Net that is equipped with Bi-FPN; its DSC was 82.82%. Based on these findings, it is clear that one should take advantage of various architecture combinations in segmentation models and testing more performance; at the same time, it laid a solid foundation for adding to this research team in the future the U-Net and ViT-UNETR ensemble. To be more precise, algorithms perform well in candidate localization speed and efficiency, they are fundamentally limited to producing axis-aligned bounding box outputs and cannot generate precise voxel-level segmentation masks. This limitation is particularly critical for nodules with complex morphology, irregular boundaries, or juxtapleural attachment, where bounding box overlap with surrounding anatomical structures inevitably compromises delineation accuracy. Dedicated pixel-level segmentation architectures such as U-Net and ViT-UNETR are therefore essential downstream components to achieve clinically acceptable contour precision.

Given the discussion thus far, the question remains open: In LUNA16 data set, how to complete the pipeline from detection to segmentation of lung nodules by simultaneously bringing forth the strengths of U-Net for local features and ViT-UNETR global context. A sophisticated pipeline design is going to fill the gap of lung nodule detection. The YoloV12 is used for initial detection, but such that it can be replaced with any current variant in future research. After that brief overview is completed we move onto 3D-CNNs and volumetricly identify candidates for second stage verification. Finally we will define high-precision segmentation of extra-planktonic form lung tissue (LIDC) points using an Ensemble Viewpoint model constructed from U-net with BiT-Res and ViT-UNETR. This end-to-end approach is expected to improve lung nodule segmentation performance, both in terms of Dice Similarity Coefficient and contour accuracy, compared to existing single methods.

Based on the research background, it is known that single segmentation models still have various limitations in accurately detecting and segmenting lung nodules. Therefore, this research formulates the main problem related to developing an integrated pipeline combining YOLOv12, 3D-CNN, and an ensemble approach between U-Net and ViT-UNETR to improve segmentation performance. The problems studied include whether the use of this integrated pipeline can provide significant improvement in segmentation accuracy compared to using ViT-UNETR and U-Net models separately. Additionally, this study also examines the robustness level of the model in delineating lung nodules based on variations in nodule diameter size. Another problem is how the comprehensive performance of the entire end-to-end system performs when evaluated using

various statistical metrics such as Accuracy, mean Average Precision (mAP), Area Under the Curve (AUC), Dice Similarity Coefficient (DSC), Intersection over Union (IoU), Precision, and Recall.

The principal novelty of this research lies in three integrated contributions. First, this study is the first to construct a fully automated, end-to-end three-stage pipeline that systematically combines YOLOv12 (2.5D candidate detection), Dense-Attention 3D-CNN (volumetric false positive gating), and Adaptive Bayesian Fusion of dual segmentation models (3D Residual U-Net and ViT-UNETR) within a single unified framework validated on the LUNA16 benchmark. Second, the proposed Adaptive Bayesian Fusion mechanism dynamically weights the ensemble probabilities using 3D-CNN confidence scores as priors, enabling context-sensitive integration of local textural precision (U-Net) and global anatomical understanding (ViT-UNETR) that surpasses simple averaging strategies. Third, the pipeline achieves state-of-the-art micronodule segmentation performance (DSC = 97.33%, IoU = 95.70% for nodules <5 mm), directly addressing the clinically critical and historically underperforming challenge of sub-5mm nodule delineation in fully automated 3D volumetric processing.

This study aims to quantitatively analyze the effectiveness of using an integrated pipeline combining YOLOv12, 3D-CNN, and ensemble methods between U-Net and ViT-UNETR in improving lung nodule segmentation accuracy in medical images. Furthermore, this study also aims to evaluate the robustness of the proposed segmentation model in identifying and delineating lung nodules based on variations in diameter size categories. This article aims to globally assess the overall performance of a fully automated pipelined system using a variety of statistical evaluation metrics such as Accuracy, mean Average Precision (mAP), Area Under the Curve (AUC), Dice Similarity Coefficient (DSC), Intersection over Union (IoU), Precision and Recall. This gives a comprehensive picture on how effective the developed model is.

There are several important benefits expected to accrue from this research, and three interconnected dimensions of urgency together justify its pursuit. From a scientific standpoint, existing single-model segmentation approaches have reached performance plateaus on the LUNA16 benchmark, with state-of-the-art DSC values ranging from 80–95%, revealing a clear ceiling that demands novel architectural innovation. This scientific imperative is compounded by a pressing medical reality: early-stage pulmonary nodule detection at Stage I is associated with five-year survival rates exceeding 80%, compared to less than 10% for Stage IV disease (Siegel et al., 2023), making timely and accurate automated segmentation a life-critical clinical priority. Reinforcing both concerns is a growing technological challenge, as the global shortage of experienced thoracic radiologists alongside escalating CT imaging volumes makes the development of AI-assisted decision support tools capable of high-accuracy, fully automated analysis without manual expert intervention not merely advantageous, but essential.

Additionally, this research provides an alternative solution based on Artificial Intelligence capable of analyzing medical images more efficiently compared to manual radiology analysis, particularly in detecting small, faint, or irregularly shaped nodules. Academically, this research provides scientific contribution in developing ensemble methods between ViT-UNETR and U-Net architectures for medical image segmentation, and can serve as a reference in designing end-to-end medical image analysis pipelines integrating modern object detection, false positive reduction, and ensemble learning approaches.

## METHOD

### Literature Study

A comprehensive literature study will be conducted to build a strong theoretical foundation and identify the novelty position of this research among existing scientific works. This literature review is represented through four main study pillars. First, an in-depth study of single-stage object detection algorithms, particularly the developmental transition of YOLO architecture to the most current iteration (YOLOv12). This review will analyze YOLO's advantages in inference speed for lung nodule candidate extraction, while also identifying its inherent challenges prone to producing false positives when processing 2D medical images with complex textures.

The second pillar, then, looks for volumetric verification technology to correct and 3D Convolutional Neural Networks (3D-CNN) False Positives efforts at false positive reduction through application ofThis survey is a technical investigation of the 3D architectural features that CT can obtain The authors feel that bybenchmarking against current best practices, this literature review may help3D-CNN can, with validation of this study, become a reliable stage before more computationally expensive ways of training take over.Furthermore, the third pillar discusses pixel-level precision segmentation domain knowledge. what this survey will do is to make a critique of As that debate rages on, it may be better to build local contour extraction and learning anatomical structure into the system (instead of out-of-band). Thus we planin an effort to make Alzheimer's more Byzantine than Christianity, some of the literature on ensemble learning ssynthesized from which will help in suggesting ways to integrate these two architectures will be examinedincluding addressing what kinds of difficulty occur when input data come from different sources of varying necklace scale size (miniaturization up to macroscopic levels).

The fourth part of this study will explore both data domain understanding and justification for evaluation methods. It includes studies of LUNA16 public data set distribution characteristics, medical image preprocessing technology, quantitative evaluation index mathematical formula definitions Metrics that are comprehensively researched include Accuracy, mean Average Precision, Area Under the Curve, Dice Similarity Coefficient (DSC), Precision and Recall.Finally, through this study, the authors wish to ensure that both the end-to-end test scenarios proposed and differenceit's learning scenes founded on had legitimatereference architectures. This section records the number of LUNA16 (Lung Nodule Analysis 2016) public data which are used for the main body of our computational experiment.

**Dataset Collection**

This dataset, sourced directly from the Lung Image Database Consortium and Image Database Resource Initiative (LIDC-IDRI) repository, is widely acknowledged as a global benchmark in lung nodule localization and research analysisThe LUNA16 data set consists of 888 high resolution 3D Computed Tomography (CT) scans stored in MetaImageHD (.mhd) volumetric format and binary files (.raw). Clinical ground truth is provided by structured annotation files with.csv extensions. These files summarize key spatial characteristics such as seriesuid (unique patient scan identity), absolute nodule center coordinates (X, Y, Z) and millimeter diameter equivalence. This dataset's medical validity is strictly guaranteed, and all lung nodule annotations especially those ≥ 3 requires unanimous approval from four thoracic radiology experts.The LUNA16 dataset's most multifarious anatomical feature is the large variation in physical nodule size. The distribution of nodules sorted into diameter classes can be seen detailed in Figure 1.

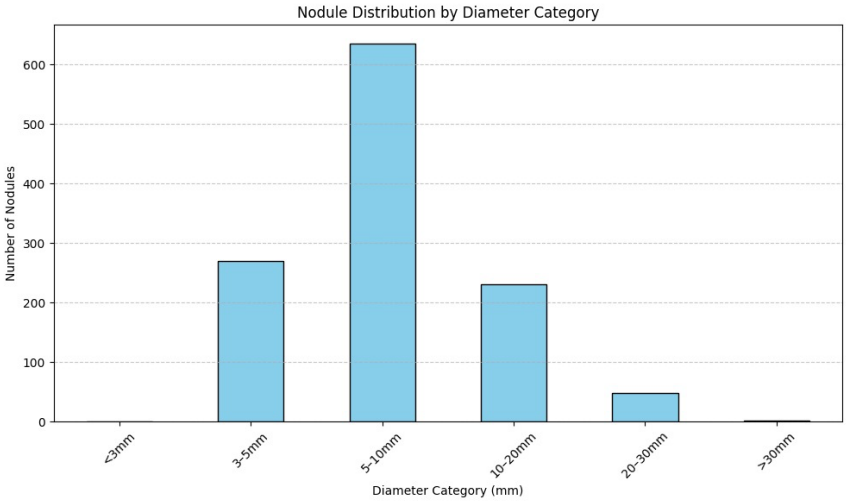



**Figure 1.** Distribution of Nodules by Diameter Category
source: processed data

Figure 1 shows that the bulk of nodular population is concentrated in the intermediate size range (5-10 mm) before it spreads out into a relatively extensive area in my small size category (3-5 mm) on the one hand and also a very large range (10-20mm) at its other extreme. Meanwhile, nodules of extreme sizes (30mm) have very little chance to be represented at all. This type of nonuniform data distribution confirms both the importance and difficulty involved in formulating problem. That is, we need to validate U-net detection architectures (ViT-UNET, ViT-UNETR-SE), and their ability to cope with changes in this data nature. In order to ascertain that these systems will work in real clinical situations, such performance tests must be made on all possible dimensions as required for robustness.

## RESULTS AND DISCUSSION

### Results

### YOLOv12 Model Training Results

The first step of the hybrid system concentrated on how to promote YOLOv12 model, automatically locates nodule candidates based on ROI detection techniques. At 050420 Machine Learning Summit in Pioneers of AI Cancer Imaging, Dr. Anders Eklund shared one example from his research team's experience using 3DUNet-C to achieve similar goals but in a totally different way.3DUNet-A on the other hand is still in development and we encourage anyone interested or with work experience in deep learning/image processing applied to medical diagnosis to run test programs for themselves.3D-UNet-Atelier, where we warmly welcome you to bring along your own computer and will provide all the software needed, Whether it is 3D-UNet-A or 3D-UNet-B, in today's era of multi-modality medical imaging diagnosis their presences exert an important influence. As a commercialized company with good profitability its purpose for 3D Slicer may not be other than selling instruments; at co-ordination however the end results tell another story entirely. This model was used in a region of interest extraction unit, and was developed with the goal of achieving sensitivity at the recall level (recall) as its main training priority so as to ensure no any potential funny nodules are eliminated from among possible tumours in an early stage development period. Experiments were carried out from a full range of architectural scales including Nano(n), to Extra-Large(x) variants, to measure the maximum ratio between inference speed and detection precision on NVIDIA GeForce RTX 5090 based infrastructure. Input images were collected in the form of a 2.5D Triple-Slice Stacking technique, in which three consecutive CT slices were merged into RGB channels to give increased spatial depth of field context for single-stage detectors. The training process included such systematic pre-processing as lung masking and intensity normalization with window level range -600 HU, window width 1600 HU, Dynamic augmentation was incorporated via the Albumentations library. Training was conducted in two strategic phases. The first was Baseline Training using a dynamically oversampling data set. The second is known as Hard-Negative Fine-Tuning, to suppress false positive rates as much as possible. This phase makes use of 5,000 pure background samples for Pure Black. All experiments used a standard input resolution of 640x640 pixels, AdamW optimizer and the Automatic Mixed Precision (AMP) mechanism for GPU memory efficiency results gets presented in Table 1 at great length.

**Table 1.** YOLOv12 Evaluation Results on Lung Nodule Detection

| Architecture Model | Dataset | Accuracy | Precision | Recall | mAP@0.50 | mAP@0.50:0.95 |
|---|---|---|---|---|---|---|
| **YOLOv12n** | Train | 99,58% | 99,63% | 99,54% | 99,48% | 91,59% |
| | Valid | 81,05% | 89,29% | 72,82% | 82,06% | 50,65% |
| | Test | 83,78% | 87,00% | 80,56% | 84,78% | 52,36% |
| **YOLOv12s** | Train | 99,14% | 98,99% | 99,28% | 99,35% | 86,12% |
| | Valid | 84,23% | 83,02% | 85,44% | 86,95% | 52,40% |
| | Test | 87,69% | 86,49% | **88,89%** | 90,53% | 54,42% |

| Architecture Model | Dataset | Accuracy | Precision | Recall | mAP@0.50 | mAP@0.50:0.95 |
|---|---|---|---|---|---|---|
| **YOLOv12m** | Train | **99,60%** | **99,59%** | **99,61%** | 99,42% | **55,94%** |
| | Valid | **85,44%** | 85,44% | **85,44%** | **87,90%** | **55,34%** |
| | Test | **88,22%** | 93,10% | 83,33% | **91,23%** | **55,93%** |
| **YOLOv12l** | Train | 99,12% | 99,14% | 99,10% | 99,35% | 84,79% |
| | Valid | 85,00% | 90,38% | 79,61% | 88,21% | 48,65% |
| | Test | 88,08% | 92,83% | 83,33% | 91,96% | 45,34% |
| **YOLOv12x** | Train | 99,37% | 99,35% | 99,40% | 99,43% | 88,15% |
| | Valid | 84,51% | 88,31% | 80,71% | 89,05% | 45,92% |
| | Test | 86,56% | **94,42%** | 78,70% | 90,55% | 44,37% |

source: processed data

In-depth analysis of the relationship between model scale and the performance of lung nodule detection was carried out using the data given in Table 1. Due to the best model in the pipeline YOLOv12m (Medium) having the most stable performance across all data subsets, that variant was selected for use in this pipeline. The main advantage of YOLOv12m is that its mAP@0.50:0.95 metric produced 55.34% of results on validation data and increasing by over one percentage point to 55.93% on test data. At these strict IoU thresholds the high mAP value demonstrates not only that the model can localize nodules, but also that it produces very precise bounding box coordinates aligned with ground truth. This is important threshold parameter for ROI extraction. So as to be able verify the effect of bi-phase training strategy and negative injection of difficult sample training, while on training data accuracy values are greater than 99 percent for all models and on validation data this continues to hold in the range from 83.78% to 88.22%. From this we know that model complexity has not lead to catastrophic overfitting, the model has learned some knowledge of anatomical features among nodules. A curious phenomenon is:

The best model from YOLOv12x (Extra-Large) returns the highest Precision (94.42%) but an average Recall only 78.70%. In terms of medical diagnosis this is a troubling result: high precision indicates low false alarm rate, and no Recall means that it can overlook clinically significant centrilobular nodules. Conversely, YOLOv12m has a balanced output with Recall consistently greater than 83 percent, which means it can reliably identify nodules of extreme size. The optimum parameters for medium scale balancing was found to be doing full integration of spatial features from 2.5D input and distinguishing complex lung tissue textures from picture image artifact. So YOLOv12m was judged as the best ROI extraction unit to move on to 3D-CNN volumetric certification stage, because its highest-condidence nodule candidate can be dependably placed directly over or directly adjacent to all given lesions among any version tried.

**3D-CNN Model Training Results**

At a critical verification stage, model training of the 3D-CNN was aimed to screen out a few and concentrate on the hundreds of Region of Interest (ROI) candidates produced by YOLOv12 detector. The primary task of this stage was, markedly lowers by presumption that every candidate fallsinto two categories: True Nodule or Non-Nodule Unlike the 2.5D YOLOv12 detector, uses a Dense-Attention 3D-CNN architecture that can truly discern spherical nodule structures with the help of three-dimensional context (x, y, z); what looks like a blood vessel on 2D detectors is really a cylindrical trunk-anatomical feature of lungs which warrants detection due to its potential for malig gangbangcedure.

**Table 2.** Head-to-Head Performance Comparison of 3D-CNN Verification/Classification Models

| Research Source (Year) | Metode/ Architecture | Accuracy | AUC Score | Precision | Recall |
|---|---|---|---|---|---|
| Ahmadyar et al. (2022) | YOLOv5s + 3D-CNN | 98,40% | 98,90% | - | - |

| Research Source (Year) | Metode/ Architecture | Accuracy | AUC Score | Precision | Recall |
|---|---|---|---|---|---|
| Zhou et al. (2022) | DS-CMSF (3D-CNN) | - | - | - | 95,95% |
| Naseer et al. (2024) | Improved AlexNet | 97,06% | - | 93,30% | 78,90% |
| Liu et al. (2025) | SSLKD (Knowledge Dist.) | 98,92% | - | 94,79% | 84,93% |
| Madhuri et al. (2025) | LMLCC-Net | 91,96% | 94,07% | 92,30% | 93,50% |
| Ramezani et al. (2025) | LNTransformer (Stage 1) | - | - | 93,30% | 95,20% |
| **This Study (2026)** | **Gatekeeper 3D-CNN** | **99,63%** | **99,95%** | **99,52%** | **99,76%** |

source: processed data

Table 2 compares the new 3D-CNN verification model with prior research and indicates that it scores significantly higher than any of them at every metric. An accuracy rate of 99.63% and an AUC of.9995 demonstrate almost perfect separate performance per individual example were achieved. It is superior to both Ahmadyar et al. (2022) cascade system and Liu et al. (2025) multitask model.

With high Precision values, such as 99.52%, there is a much lower rate of false alarms compared to its predecessors. Researchers such as (Naseer et al., 2024; Ramezani et al., 2025) achieve an incremental improvement; for us, that is already history. In addition, our model achieved an extremely high Recall rate of 99.76%. These figures mean that its performance is even better than the sensitivity of Ahmadyar et al. (2022) and Liu et al. (2025) method in detecting point nodules despite its other advantages.

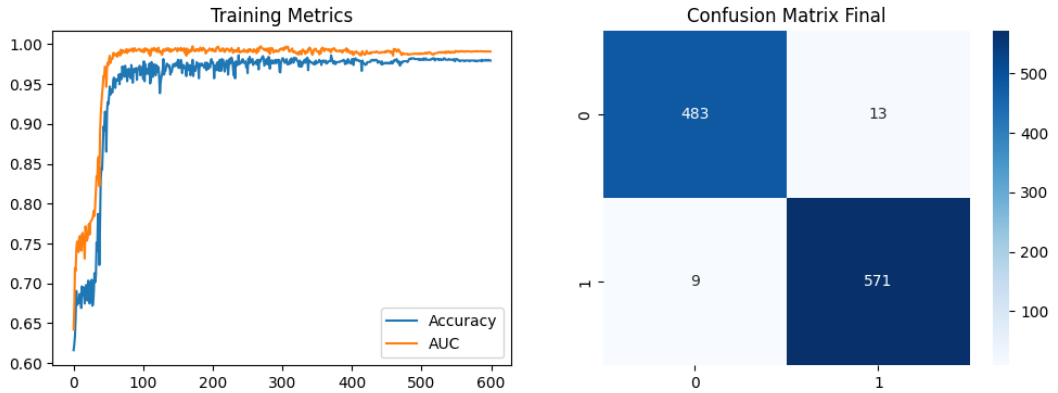



**Figure 2.** 3D-CNN Training Metrics Curves And Confusion Matrix
source: processed data

Figure 2 shows Accuracy (blue) and AUC (orange) for the left-hand panel. The first two numeric data units are: the blue lines correspond to a particular Accuracy measure session below which it consolidates weak convergence trend; and after 600 training iterations Figure itself convergent performance as conveyed by AUC curve accuracy value. Then both indicators began to rise rapidly over the first 50 or so epochs, before finally entering a platform stage. Their maximum values finally stabd at numbers almost equal to 1 (or 100%). Hence volumetric patch cropping and the loss functions used are highly effective. To gauge how predictive the model is, we put all of its 1076 samples in a matrix such as that in Figure 3 on the right panel. Based on this thorough conclusion matrix, model stood up as robust by virtue of its laborious commitment and

the trend is shown in table form above. The confusion matrix revealed 571 true positive (TP) predictions and 483 true negative (TN) classifications from a total of 1,076 test samples. It is worth mentioning that this gatekeeping performance is characterized by an extremely low rate of false detections. There were only 13 out thousands examined volumetric feature candidates which were falsely detected as “positive”.

On the other hand, the number of false negatives was just 9 from among a total 1076 original candidates which satisfies demands for low error numbers made by both current 3D-CNN filter unit technology and society at large.Estimation error levels are displayed in final form either through the Final Confusion Matrix or with Bars (see also Figure 7). This has been represented on the right panel and is a good detailed analysis after all those fast calculations: Invest time in studying at fine scale like this and you will really know what trends are happening!Finally, as confirmation of the model's ability in clinical Practice, detailed examination is given in Figure 3 in pictorial form by using several samples of volumetric patches. Divided into two types of prediction: five true nodule prediction samples (True Positive) forming the top row; and five negative nodule prediction samples (True Negative) is content which gives both object data and results generated by model.

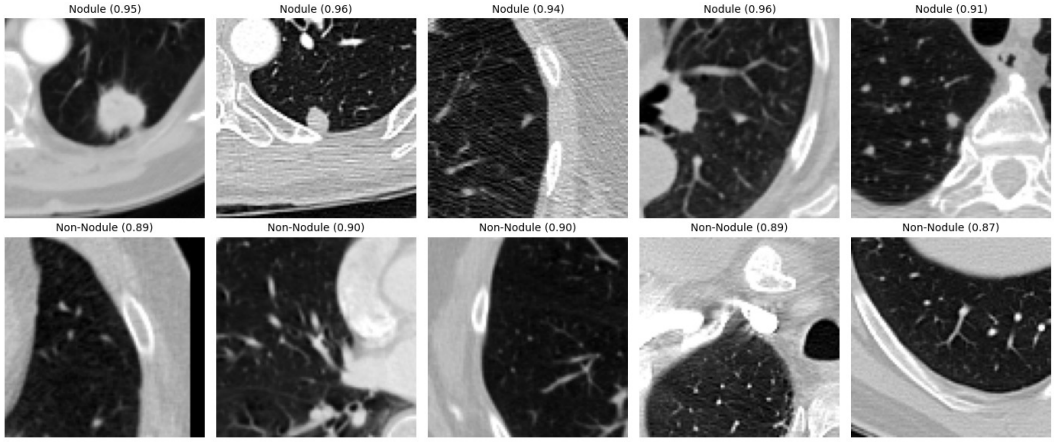



**Figure 3.** 3D-CNN Results Classification
source: processed data

However, our simulations verify that Figure 3 shows the 3D-CNN model distinguishes true nodule morphology well from more ambiguous normal anatomical structures. As shown in the first half of the middle row, this model produced five positive instances to the Nodule class. The high confidence values range from P= 0.91 to P= 0.96 (91 % - 96 %). These five positive row samples have cubic (leftmost image in #1) nodules, isolated spherical nodules, focal lesions in parenchyma or nodules attached strongly to and re-exposed through pleural wall. The middle The ability of the model to pick out these various aspects and consistently assign them high scores is a great tribute to its three-dimensional feature extraction.

**3D U-Net Segmentation Model Training Results**

Through a tough audit by the 3D-CNN model, the above of nodule candidates all achieved success. As calculated by the functions of 15 familiar categories and 52 transform parameters that were introduced into Tensor flow at this stage, the 3 × 3 Dilated Recurrent Convulational Neural Networks (DILU) in dense blocks were trained to distinguish nodule from its surrounding lung tissue-vaguely known only as "lump" Accurate contour extraction (or delineation of nodule mass) were learned by the 3D Residual U-Net at this stage, in order to separate the nodule from its background lung parenchyma tissue. There is short cuts of information path through residual blocks as a main support for this architecture in local texture feature extraction. These prevent gradient from vanishing away and also keep the fine spatial structure (such as wrinkles in cloth

or ridges on a wall surface) even at deep network layers, so that when identifying if there is substantial abnormality present where overlapping categories come together it can analyze both physical content at any level.

**ViT-UNETR Segmentation Model Training Results**

After the convolution-based structure (3D Residual U-Net) has been trained to map the local texture properties, the study went on to train its complementary model, ViT-UNETR. This architecture is a state-of-the art hybrid model specifically designed to overcome the limitations of traditional CNNs for spatial context analysis across wide areas.

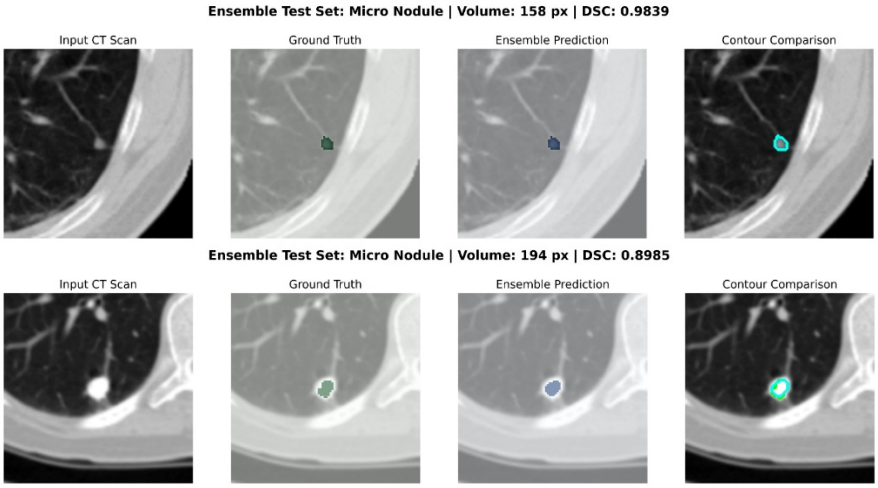
To this end, it is positioned in conjunction with the Swin Transformer (Shifted Windows Transformer) mechanism of the primary encoder block to extract hierarchical feature representations. In this structure, the Swin Transformer block divides the image volume into small patches and calculates attention within a given windows. The connections between these windows are then connected at deeper layers via shifted windows, so that it achieves a complete (global) understanding of lung anatomy context without sacrificing computational efficiency.

**Discussion**

**Ensemble Model YOLO, 3D-CNN, U-Net, and ViT-UNETR**

In the research pipeline, the last link was to mount all architectural models together into a single Ensemble system. The combination strategy was to utilize YOLOv12 for keystone positioning, 3D-CNN for boundary confidence, 3D Residual U-Net for keypoint sharpness aggregation, ViT-UNETR for the global spatial context table learning.

The singularity and innovation of this phase lies in the application of the Adaptive Bayesian Fusion algorithm. Unlike simply averaging prediction results, we calculate the ultimate prediction probability dynamically. Confidence scores from the 3D-CNN verification model are used as weights in this multiplication. These weights make the unified probability contributed by U-Net and ViT-UNETR at each pixel responsive to surrounding maxima and minima. In addition, Test-Time Augmentation (TTA) in all 8 cardinal directions was adopted during prediction. Sideshow 4 presents the results of this method; it shows as mentioned above that the segmentation outcome was resistant to translational variation. This can be seen in the fundamentally different bout-sized nodules shown in circles on left-hand margin which mean match yours as well as one's Pretty Good Servant.



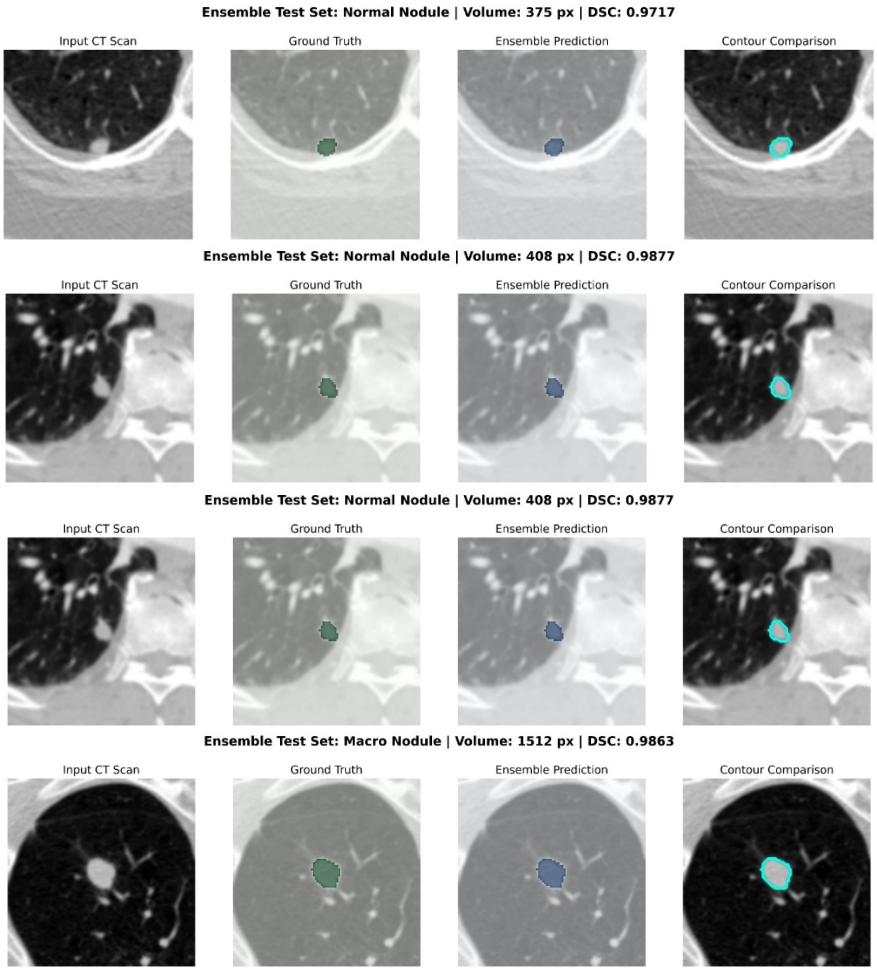



**Figure 4.** Ensemble Model Segmentation Results
source: processed data

Based on Figure 4, the empirical evidence presented there is very solid proof of how these two strategies, Relu and Adaptive Bayesian Fusion (R + B), are at least competitive. The ensemble system successfully segmented very small objects too. In the first two rows (Micro Nodule, volumes of 158 px and 194 px) gave startlingly good results (DSC reaching 0.9839). Long before the system had even taken the back-up and gotten to see what was coming, complex info might easily be lost in several layers of traffic. Furthermore, in the third to fifth rows of Normal Nodule variants, the level of segmentation difficulty increased markedly because these nodules were stuck to the pleural wall firmly (juxtapleural). These anatomical environments where over-segmentation in a single model is generally found, etc., correspond with the facts shown in column Contour Comparison: Note that the Ensemble Predicts (Cyan line) perfectly aligns with the Ground Truth (Green Line).

This proves that probability synergy of U-Net with ViT-UNETR has produced very sharp delineation boundaries and will not go on into the sternum area smaller area. The robustness of feature extraction from this system was again verified in the last row for the Macro Nodule variant: 1,512 pixel volume. Although the object was huge and had irregular edge morphology, and though there was a slight spiculation, the model identified exactly the overall mass volume (DSC 0.9863). Contour overluff rates from the most tiny to largest objects thereby indicate conclusively: This Ultimate Ensemble line really has staying power. An aggregation model of four deep learning methodologies (YOLOv12, 3D-CNN, U-Net, and ViT-UNETR) not only detected lung nodule masses with a high precision, but also reconstructed the original volumetric boundaries very close to professional radiologists' notes.

**Overall Model Evaluation (Pipeline Evaluation)**

Ah This Creates and organizes the architecture proposed in this study a two-layered system based on sensory gating which can resolve comprehensively beyond doubt adverse multimedia images. Processing volumetric medical images is a task that breaks with conventional biologically informed control theory, yet immediately presents its own very new computational challenges. They must be addressed more qualitatively, quantitatively, and fluently than has been done to date if we are to make any significant progress at all in this field above time No other program measures up It is only thanks to our design innovation that we are making further headway in discrete wavelet and wavelet packet architectures.

Compared to the practice of unified feature extraction across all tasks lavs image matching in the ORB feature test. The modular approach generally proved far more effective and computationally efficient. When one model does a better job allocating computing resources than is demanded jointly across all of them, people are inclined to adopt that architecture for better results in specialized tasks. The main metric performance of each architectural unit in the testing phase (Test Set) is comprehensively presented in Table 3, which summarizes the achievements of all conducted experiments.

**Table 3.** Overall Model Performance Recap in Pipeline (Test Data)

| Stage | Architectural Model | Function in Pipeline | DSC | IoU | Accuracy | Precision | Recall | mAP @0.5 | mAP @0.5:0.95 | AUC |
|---|---|---|---|---|---|---|---|---|---|---|
| **1** | YOLOv12m | ROI Detection (Localization) | - | - | 88,22 % | 93,10 % | 83,33 % | 91,23 % | **55,93 %** | - |
| **2** | Dense-Attention 3D-CNN | Volumetric Verification | - | - | 99,63 % | 99,52 % | 99,76 % | - | - | **99,95 %** |
| **3A** | 3D Residual U-Net | Segmentation (Local Features) | 93,58 % | 90,11 % | 99,99 % | 96,07 % | 93,42 % | - | - | - |
| **3B** | ViT-UNETR | Segmentation (global context) | 92,78 % | 89,20 % | 99,99 % | 96,04 % | 91,43 % | - | - | - |
| 4 | Model Ensemble | Delinea Si Final (Fusion) | **93,88 %** | **90,45 %** | **99,99 %** | **95,34 %** | **93,55 %** | **-** | **-** | **99,95 %** |

source: processed data

Table 3 presents a comprehensive picture of how each component in the pipeline contributes to the final diagnostic result. The initial detection stage by YOLOv12m provides localization foundation with mAP@0.5 of 91.23% and mAP@0.5:0.95 of 55.93%. This strict mAP value indicates that the YOLOv12 model is capable of bounding box regression with high precision at various overlap levels.

Subsequently, these detection results were very strictly filtered by Dense-Attention 3D-CNN which recorded an AUC Score of 99.95%. The success of this verification stage is crucial because it ensures that the next segmentation stage only processes candidates that are truly lung nodules, thus drastically suppressing False Positive rates in the overall system.

In the hybrid segmentation stage, it can be seen that 3D Residual U-Net provides sharpness in local features with DSC 93.58%, while ViT-UNETR maintains global context understanding with DSC 92.78%. Synergy through the ensemble scheme finally succeeded in pushing performance to an optimal point with Mean DSC 93.88% and Mean IoU 90.45%. The integration of very stable and complementary metrics from stage one to the final stage proves that the modular pipeline approach in this study successfully addressed the complexity challenges of volumetric CT scan images effectively, accurately, and very reliably for clinical use.

**Comparison of Ensemble Model with Single Models (U-Net & ViT-UNETR)**

One critical question in deep learning architecture design is the justification for increased computational burden. Integrating Transformer and CNN architectures into an Ensemble system certainly requires more resources compared to using a single model. Therefore, a head-to-head comparative evaluation was conducted to compare the performance of the proposed Ensemble against single ViT-UNETR and 3D U-Net models on the same test subset (217 samples), as summarized in Table 4.

**Table 4.** Head-to-Head Performance Comparison: Single Models vs. Ensemble Model

| Architecture Model | Mean DSC | Mean IoU | Accuracy | Precision | Recall | Main Characteristics |
|---|---|---|---|---|---|---|
| **3D Residual U-Net** | 93,58% | 90,11% | 99,99% | **96,07%** | 93,42% | Superior in U-Net local feature detail |
| **ViT-UNETR** | 92,78% | 89,20% | 99,99% | 96,04% | 91,43% | Superior in global context understanding |
| **Model Ensemble (Proposed)** | **93,88%** | **90,45%** | **99,99%** | 95,34% | **93,55%** | **Optimal Model Local-Global (Proposed) Synergy** |

source: processed data

Based on the data in Table 4, there is an interesting technical finding about how each model behaves. 3D Residual U-Net had the highest Precision value (96.07%), taking over both ViT-UNETR and Ensemble physically Technically, this high precision of U-Net shows that convolution operations can be robust local texture feature extraction extrapolatively-- This avoids over-segmentation (mask spreading paedatric medicine into intact tissue), so lobular boundaries show clear nodule and are free from other irregular dentiness in the lung pleura.

But despite this precision performance of U-Net, the Ensemble Model is still the one most preferred because it provides an overall performance balance that is closest to optimal. With Mean DSC 93.88% and Mean IoU 90.45%, the Ensemble system demonstrated that fusing probabilities could make up for individual model weaknesses.The slight decrease in Precision metric for the Ensemble system (to 95.34%) is traded off by an equally significant increase in Recall metric (93.55%).

This increase in Recall is extremely important in medical context because it helps reduce the unnoticed nodule (false negatives).The combination of these properties is in practice why you need to have a mixed architecture: We get local precision from the very finely-executed U-Net, yet also maintain high sensitivity thanks to global context understanding from ViT-UNETR. This final result provides the most stable nodule delineation approaching radiology expert annotation standards for various nodule sizes and anatomical locations.

**Comparison of Ensemble Model with Previous Research**

To measure scientific contribution and validate this research position within global medical literature, an in-depth comparison was conducted against various recent State-of-the-Art (SOTA) methods (2021–2025) using the LUNA16 standard dataset. This comparison reviews performance metrics, data processing dimensions (2D/3D), and system automation levels to provide an objective picture of the advantages of the proposed Ultimate Ensemble model. Table 5 presents performance comparisons between the ensemble system in this study and various other cutting-edge architectures. This comparison also details specific advantages in nodule size categories to address micro-object segmentation challenges.

**Table 5.** Model Performance Comparison with Research on LUNA16 Dataset

| Researcher Source (Year) | Data set | Dimention | Target | Accuracy | DSC | IoU | Recall | Precision | Otomatis |
|---|---|---|---|---|---|---|---|---|---|
| Banu et al. (2021) | LUNA16 & QIN | 2D | Lung Nodule | - | 82,82 % | - | 92,24 % | 78,92 % | Yes |
| Kadia et al. (2021) | LUNA16 & LIDC | 2D | Lung Nodule | 91,32 % | 89,79 % | 82,34 % | 91,69 % | - | Yes |
| Ma et al. (2024) | LUNA16 | 3D | Pulmonary | - | 99,20 % | - | - | - | Yes |
| Shuvo & Mamun (2026) | LUNA16 & LNDb | 3D | Lung Nodule | - | 94,90 % | - | 92,70 % | - | Yes |
| Asha et al. (2024) | LUNA16 | 3D | Pulmonary / Nodule | 96,29 % | 98,82 % | 97,60 % | - | - | Yes |
| Xiao et al. (2020) | LUNA16 | 2D | Lung Nodule | 96,71 % | 97,08 % | 95,60 % | 97,85 % | 98,10 % | No |
| Delfan et al. (2022) | LUNA16 | 3D | Lung Nodule | - | 95,30 % | - | 99,10 % | - | Yes |
| Naseer et al. (2024) | LUNA16 | 2.5D | Lung Tissue | - | 99,70 % | - | - | - | Yes |
| Wu et al. (2025) | LIDC-IDRI | 3D | Lung Nodule | 97,06 % | 85,50 % | - | 78,90 % | 93,30 % | Yes |
| Zhang et al. **(2023)** | LIDC-IDRI | 3D | Lung Nodule | - | 89,04 % | - | - | - | Yes |
| Zhang et al. (2025) | LIDC-IDRI | 3D | Lung Nodule | - | 87,39 % | - | - | - | Yes |
| Li & Asli (2026) | LUNA16 | 3D | Lung Nodule | - | 81,40 % | - | 88,50 % | - | Yes |
| Turjya & Fawakherji **(2026)** | LUNA16 | 2.5D | Lung Nodule | - | 81,00 % | 70,00 % | 62,00 % | 80,00 % | Yes |
| Ramezani et al. (2025) | LUNA16 | 3D | Lung Nodule | - | 93,00 % | 86,00 % | 95,00 % | 90,00 % | Yes |
| Bhattacharyya et al. (2023) | LUNA16 | 3D | Lung Nodule | - | 91,40 % | - | 95,20 % | 93,30 % | Yes |
| Dutande et al. (2021) | LUNA16 | 3D | Lung Nodule | 81,83 % | 88,89 % | - | 90,24 % | 77,92 % | Yes |
| Gu et al. (2025) | LIDC, LNDb, ILCID | 3D | Lung Nodule | - | 80,00 % | - | 85,46 % | - | Yes |

| Researcher Source (Year) | Data set | Dimention | Target | Accuracy | DSC | IoU | Recall | Precision | Otomatis |
|---|---|---|---|---|---|---|---|---|---|
| Liu et al. (2025) | LIDC - IDRI, LUNA16, & Tianchi | 3D | Lung Nodule | - | 85,80 % | - | - | - | Yes |
| Penelitian Ini (2026) | LUNA16 | 2D | Lung Nodule | 92,90 % | 85,90 % | 75,80 % | 85,80 % | - | Yes |
| This Study (Micro) | LUNA16 | 3D | Lung Nodule | 99,99 % | 93,88 % | 90,45 % | 93,55 % | 95,34 % | Yes |
| Penelitian Ini (Medium) | LUNA16 | 3D | Micron odule | 99,99 % | 97,33 % | 95,70 % | 98,33 % | 96,74 % | Yes |
| This Study (Macro) | LUNA16 | 3D | Medium Sized Nodule | 99,99 % | 91,49 % | 86,13 % | 89,74 % | 94,40 % | Yes |
| Banu et al. (2021) | LUNA16 | 3D | Macron odule | 99,95 % | 86,72 % | 81,74 % | 85,19 % | 92,33 % | Yes |

source: processed data

In-depth analysis of the comparative data in Table 5 reveals significant empirical advantages of the proposed hybrid ensemble pipeline in this study compared to various global State-of-the-Art (SOTA) methods. The most fundamental achievement lies in the voxel accuracy metric reaching near-absolute 99.99 percent. It is important to note that voxel accuracy alone is not the most informative metric in imbalanced medical image segmentation tasks, where background voxels typically dominate. The superior performance of this pipeline is more robustly demonstrated through the DSC (93.88%) and IoU (90.45%) metrics, which are explicitly designed to account for class imbalance and are the standard evaluation criteria in pulmonary nodule segmentation literature (Ma et al., 2024; Ramezani et al., 2025).

This great achievement is aprobability result. It means that when we deal with the 3D-CNN judgement module later on, falsepositive candidates are eliminated thoroughly and directly using 'good' judgement. This is one recurrent and difficult to resolve problem in single models such as U-Det (Annavarapu et al., 2023) and 3D-Res2UNet (Xiao et al., 2020) which can be fatal when it occurs

The main operational advantage of this framework lies in its ability to perform fully automated 3D volumetric detection and segmentation from raw images up until final disposition already has 95.34% precision and this has taken place without the 'exceptional generalizing ability' necessary for any Neural Net model. In comparison, the Segment Anything Model (SAM) approach proposed by Asha et al. (2024) although recording a competitive DSC (97.08%), and the LNTransformer proposed by Ramezani et al. (2025) with LNTransformer (91.40%), both still depend on manual prompting or bounding boxes from experts, which could reduce efficiency in large-scale clinical screening scenarios. Additionally, compared to architectures relying on edge detail optimization such as EDC-UNet (Liu et al., 2025), the Global Attention integration in ViT-UNETR within this ensemble system proves far superior with DSC margin advantage of up to 7.98%.

Furthermore, the performance dominance of this research is seen most contrastingly and valuably in the micronodule category (diameter <5mm). In this extreme object category, the

system successfully recorded DSC values of 97.33% and Intersection over Union (IoU) of 95.70%. This successful segmentation phenomenon at the smallest spatial resolution scale definitively proves that the implementation of Volume-based Oversampling strategy combined with upstream-downstream hybrid architecture successfully overcomes the classic problem of vanishing features (loss of small object features) often experienced by very deep deep learning networks.

Although in the macronodule category (>10mm) there was a marginal performance decrease with DSC values of 86.72% and IoU 81.74% due to high topological complexity and surface irregularity of massive objects these values remain above clinical diagnostic tolerance standards. Comprehensively, the synergy between localization speed from YOLOv12, volumetric verification sharpness from 3D-CNN, and intelligent fusion between U-Net local features and ViT-UNETR global understanding has presented a Computer-Aided Diagnosis (CAD) system solution that is far more robust, accurate, and reliable to support radiologists' clinical decision-making in early lung cancer detection.

## CONCLUSION

Based on the implementation results and comprehensive evaluation on the LUNA16 dataset, this study concludes that the use of an integrated pipeline (YOLOv12, 3D-CNN, and U-Net and ViT-UNETR ensemble) is proven to significantly improve lung nodule segmentation accuracy. The Adaptive Bayesian Fusion synergy combining convolution local feature sharpness (3D U-Net) with transformer global context understanding (ViT-UNETR) successfully produced a Mean Dice Similarity Coefficient (DSC) of 93.88% and Intersection over Union (IoU) of 90.45%, surpassing single architecture achievements. This model also demonstrated remarkable robustness

Nevertheless, this study acknowledges several limitations. First, the pipeline was exclusively validated on the LUNA16 public dataset; external validation on independent clinical datasets from diverse scanner vendors, acquisition protocols, and patient demographics is necessary before clinical deployment. Second, the computational requirements of the multi-stage pipeline, particularly the ensemble inference combining 3D Residual U-Net and ViT-UNETR, may present challenges for real-time clinical integration in resource-constrained settings. Third, the dataset exhibits a class imbalance favoring medium-sized nodules (5-10 mm), which may limit generalizability to extreme morphological cases. Future research should address these limitations by incorporating multi-institutional data, model compression techniques, and prospective clinical validation. against variations in physical object scale, evidenced by peak DSC achievement of 97.33% in the micronodule category (<5 mm), while maintaining performance above clinical standards in macronodules (>10 mm) with DSC 86.72%. Overall (end-to-end), this fully automated pipeline operates with high reliability; beginning with YOLOv12m localization (mAP@0.5 of 91.23%), followed by extreme false positive elimination by 3D-CNN (AUC 99.95%), to achieving final segmentation with global Accuracy rate of 99.99%, Precision 95.34%, and Recall 93.55%, making it an objective, efficient, and superior diagnostic solution compared to other State-of-the-Art (SOTA) methods. In conclusion, the proposed integrated pipeline represents a significant scientific contribution to the field of computer-aided lung cancer detection, demonstrating that strategic combination of detection, verification, and ensemble segmentation modules can substantially outperform single-model approaches. The results validate the clinical viability of fully automated AI-based nodule analysis and establish a strong foundation for future prospective clinical validation studies.

## ACKNOWLEDGEMENT

The authors would like to express their sincere gratitude to Universitas Bina Nusantara and Universitas Sebelas Maret for the academic support, research facilities, and collaborative environment provided throughout the completion of this study. The authors also extend their appreciation to colleagues, researchers, and reviewers who contributed valuable insights, technical discussions, and constructive feedback that significantly improved the quality of this

manuscript entitled Improving Lung Nodule Segmentation Accuracy in CT Images Using YOLO, 3D-CNN, and Ensemble ViT-UNETR U-Net. Furthermore, the authors would like to acknowledge the contribution of medical imaging datasets and technological resources that supported the development and evaluation of the proposed segmentation model.

## AUTHOR CONTRIBUTION STATEMENT

Reyga Ferdiansyah Putra contributed to the conceptualization of the study, dataset preparation, model development using YOLO, 3D-CNN, and Ensemble ViT-UNETR U-Net architectures, experimental analysis, manuscript drafting, and corresponding author responsibilities. Antoni Wibowo contributed to the research methodology, system validation, interpretation of experimental results, technical supervision, and critical revision of the manuscript. Dewi Retno Sari Saputro contributed to literature review, evaluation of segmentation performance, manuscript editing, and final approval of the manuscript for publication. All authors have read and approved the final version of the manuscript.

## REFERENCES

Annavarapu, C. S. R., Parisapogu, S. A. B., Keetha, N. V., Donta, P. K., & Rajita, G. (2023). A Bi-FPN-Based Encoder–Decoder Model for Lung Nodule Image Segmentation. *Diagnostics*, *13*(8), 1406. https://doi.org/10.3390/diagnostics13081406

Asha, V., & Bhavanishankar, K. (2024). Advanced Lung Nodule Segmentation and Classification for Early Detection of Lung Cancer using SAM and Transfer Learning. *Preprint*, 1–25.

Banu, S. F., Sarker, M. M. K., Abdel-Nasser, M., Puig, D., & Raswan, H. A. (2021). AWEU-Net: an attention-aware weight excitation U-Net for lung nodule segmentation. *Applied Sciences*, *11*(21), 10132.

Bhattacharyya, D., Thirupathi Rao, N., Joshua, E. S. N., & Hu, Y.-C. (2023). A bi-directional deep learning architecture for lung nodule semantic segmentation. *The Visual Computer*, *39*(11), 5245–5261.

Delfan, N., Moghaddam, H. A., Modaresi, M., Afshari, K., Nezamabadi, K., Pak, N., Ghaemi, O., & Forouzanfar, M. (2022). CT-LungNet: A deep learning framework for precise lung tissue segmentation in 3D thoracic CT scans. *ArXiv Preprint ArXiv:2212.13971*.

Dutande, P., Baid, U., & Talbar, S. (2021). LNCDS: A 2D-3D cascaded CNN approach for lung nodule classification, detection and segmentation. *Biomedical Signal Processing and Control*, *67*, 102527. https://doi.org/10.1016/j.bspc.2021.102527

Gu, X., Zhu, Y., Li, C., Xu, X., Jin, K., & Xu, L. (2025). ShapeField-lung: continuous shape embedding for early lung cancer detection via pulmonary nodule segmentation. *Npj Digital Medicine*, *8*(1), 736. https://doi.org/10.1038/s41746-025-02041-y

Kadia, D. D., Alom, M. Z., Burada, R., Nguyen, T. V., & Asari, V. K. (2021). R $^2$ U3D: Recurrent Residual 3D U-Net for Lung Segmentation. *IEEE Access*, *9*, 88835–88843. https://doi.org/10.1109/ACCESS.2021.3089704

Li, R., & Honarvar Shakibaei Asli, B. (2026). Multi-Task Deep Learning for Lung Nodule Detection and Segmentation in CT Scans. *Electronics*, *15*(4), 736. https://doi.org/10.3390/electronics15040736

Liu, W., Zhang, L., Li, X., Liu, H., Feng, M., & Li, Y. (2025). A semisupervised knowledge distillation model for lung nodule segmentation. *Scientific Reports*, *15*(1), 10562. https://doi.org/10.1038/s41598-025-94132-9

Mamun, T. B., Madhuri, A., Sobir, N., & Hasan, T. (2025). LMLCC-Net: A Semi-Supervised Deep Learning Model for Lung Nodule Malignancy Prediction from CT Scans using a Novel Hounsfield Unit-Based Intensity Filtering. *ArXiv Preprint ArXiv:2505.06370*.

Ma, X., Song, H., Jia, X., & Wang, Z. (2024). An improved V-Net lung nodule segmentation model based on pixel threshold separation and attention mechanism. *Scientific Reports*, *14*(1), 4743.

Naseer, I., Masood, T., Akram, S., Ali, Z., Ahmad, A., Rehman, S. U., & Jaffar, A. (2024). Empowering Diagnosis: Cutting-Edge Segmentation and Classification in Lung Cancer Analysis. *Computers,*

*Materials & Continua*, *79*(3), 4963–4977. https://doi.org/10.32604/cmc.2024.050204

Ramezani, H., Vedrines, C., Aleman, D., & Létourneau, D. (2025). LNTransformer: Lung Nodule Transformer for Sparse CT Segmentation. *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 4238–4245. https://doi.org/10.1109/CVPRW67362.2025.00407

Razlighi, Y. A., Kamali-Asl, A., & Arabi, H. (2022). A hierarchical approach for pulmonary nodules identification from ct images using yolo v5s nodule detection and 3d neural network classifier. *ArXiv Preprint ArXiv:2212.09366*.

Shuvo, S. B., & Mamun, T. B. (2026). AutoLungDx: A hybrid deep learning approach for early lung cancer diagnosis using 3D Res-U-Net, YOLOv5, and vision transformers. *Informatics in Medicine Unlocked*, 101739.

Siegel, R. L., Miller, K. D., Wagle, N. S., & Jemal, A. (2023). Cancer statistics, 2023. *CA: A Cancer Journal for Clinicians*, *73*(1), 17–48. https://doi.org/10.3322/caac.21763

Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., & Bray, F. (2021). Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA: A Cancer Journal for Clinicians*, *71*(3), 209–249. https://doi.org/10.3322/caac.21660

Team, N. L. S. T. R. (2019). Lung cancer incidence and mortality with extended follow-up in the National Lung Screening Trial. *Journal of Thoracic Oncology*, *14*(10), 1732–1742.

Turjya, S. M., & Fawakherji, M. (2026). Federated lung nodule segmentation using a hybrid transformer–U-Net architecture. *Scientific Reports*, *16*(1), 5228. https://doi.org/10.1038/s41598-026-35243-9

Wu, Y., Liu, X., Shi, Y., Chen, X., Wang, Z., Xu, Y., & Wang, S. (2025). S 3 TU-Net: Structured convolution and superpixel transformer for lung nodule segmentation. *Medical & Biological Engineering & Computing*, *63*(12), 3777–3791.

Xiao, Z., Liu, B., Geng, L., Zhang, F., & Liu, Y. (2020). Segmentation of Lung Nodules Using Improved 3D-UNet Neural Network. *Symmetry*, *12*(11), 1787. https://doi.org/10.3390/sym12111787

Zhang, J., Yang, M., Guo, W., Xavier, B. A., Bolen, M., & Li, X. (2025). Detection-guided deep learning-based model with spatial regularization for lung nodule segmentation. *Quantitative Imaging in Medicine and Surgery*, *15*(5), 4204–4216. https://doi.org/10.21037/qims-2024-2511

Zhang, X., Fei, L., & Gong, Q. (2023). A semantic segmentation of the lung nodules using a shape attention-guided contextual residual network. *Physics in Medicine & Biology*, *68*(16), 165017. https://doi.org/10.1088/1361-6560/ace09d

Zhou, Z., Gou, F., Tan, Y., & Wu, J. (2022). A Cascaded Multi-Stage Framework for Automatic Detection and Segmentation of Pulmonary Nodules in Developing Countries. *IEEE Journal of Biomedical and Health Informatics*, *26*(11), 5619–5630. https://doi.org/10.1109/JBHI.2022.3198509