



OPEN Three-dimensional reconstruction of lung tumors from computed tomography scans using adversarial and transductive learning

Zhisen He¹, Leila Jamel², Danyi Huang³, Gaozhe Jiang⁴, Zaffar Ahmed Shaikh^{5,6}, Md.Abeer Aljohani Khan⁷ & Seyed Jalaeddin Mousavirad⁸✉

Lung cancer is a critical health issue, and early detection is crucial for enhancing patient outcomes. This study presents a novel framework for generating three-dimensional (3D) representations of lung tumors from computed tomography (CT) scans, addressing three key challenges in the analysis process. Firstly, we address the precise segmentation of lung tissues, which is complicated by a high proportion of non-lung pixels that skew the classifier. Our method uses a customized generative adversarial network (GAN) enhanced with an off-policy proximal policy optimization (PPO) strategy. This strategy enhances segmentation performance by addressing inherent classifier biases and implementing a reward system to more accurately identify minority samples. Secondly, the framework enhances tumor detection in the segmented areas by employing a specialized GAN trained with an adversarial loss, which helps the generator create tumor regions that match real ones in both shape and internal features, even when contrast is low or boundaries are unclear. Thirdly, after tumor detection, the EfficientNet model extracts essential features for 3D reconstruction. The features are then enhanced by a spatial attention-based transductive long short-term memory (TLSTM) network for better performance. The TLSTM network enhances performance by assigning greater weight to samples near the test point within a transductive learning framework. Tested on the Lung Image Database Consortium Image Collection (LIDC-IDRI) dataset, our methodology achieved Hausdorff distance (HD) and Euclidean distance (ED) metrics of 0.648 and 0.985, respectively, indicating superior performance compared to existing methods. Our research introduces a clinical tool that significantly boosts the capabilities of radiologists in diagnosing and planning treatment for lung cancer. Code is publicly available at <https://github.com/ZhisenHe/3D-representation/>.

Keywords Lung cancer, Three-dimensional tumor reconstruction, Generative adversarial network, Off-policy proximal policy optimization, Transductive learning

Lung cancer poses a significant threat to public health, being one of the most lethal cancers that results in a substantial number of fatalities globally each year. The World Health Organization (WHO) reports approximately 2.09 million new cases annually, culminating in around 1.76 million fatalities¹. However, the early identification of lung cancer increases survival probabilities by up to 55%. Despite advanced imaging approaches such as CT scans, X-radiation (X-rays), and magnetic resonance imaging (MRI), intrinsic constraints still

¹Department of Electrical, Computer, and Systems Engineering, Case Western Reserve University, Cleveland 44106, USA. ²Department of Information Systems, College of Computer and Information Sciences, Princess Nourah bint Abdulrahman University, Riyadh 11671, Saudi Arabia. ³Department of Chemical Engineering, Columbia University, New York City 10027, USA. ⁴Department of Chemistry, The University of Hong Kong, Hong Kong 999077, People's Republic of China. ⁵Department of Computer Science and Information Technology, Benazir Bhutto Shaheed University Lyari, Karachi 75660, Pakistan. ⁶School of Engineering, École Polytechnique Fédérale de Lausanne, Lausanne 1015, Switzerland. ⁷School of Biomedical Engineering, Hainan University, Haikou 570228, China. ⁸Department of Computer and Electrical Engineering, Mid Sweden University, Sundsvall, Sweden. ✉email: seyedjalaeddin.mousavirad@miun.se

affect their utility^{2,3}. CT scans produce precise 3D visuals of lung tissues, but their propensity for high false-positive rates often results in superfluous biopsies and operations^{4–6}. X-rays are common for initial screenings, yet they lack the sensitivity to pinpoint tumors early and fail to accurately distinguish between benign and malignant growths. MRIs provide excellent soft tissue contrast but are less effective for lung imaging due to the substantial air content and motion of the organ during respiration⁷. These limitations highlight the need for more advanced computational techniques, especially those based on deep learning. Such methods can improve detection accuracy and support precise 3D reconstruction of lung tumors. Accurate 3D modeling is crucial for comprehending tumor morphology and informing effective treatment planning.

To date, advanced deep learning techniques have been applied to 3D modeling. Deep learning models are trained on large datasets of carefully labeled images. These models are highly effective at identifying complex patterns that often go unnoticed by human radiologists and traditional machine-learning techniques^{8,9}. They improve the accuracy and reliability of lung cancer diagnosis. They also reduce false positives and limit the need for invasive procedures¹⁰. Deep learning models divide the task of 3D modeling into three subcomponents: lung segmentation, tumor detection, and 3D tumor reconstruction². However, existing deep learning frameworks still face notable limitations in each of these stages. First, lung segmentation models often struggle with severe class imbalance, where non-lung pixels significantly outnumber lung pixels. This imbalance leads to biased classifiers that favor the majority class, thereby reducing segmentation accuracy. Second, despite architectural advances, accurate tumor detection remains difficult, especially when tumors have unclear boundaries or low contrast. This can result in both over-detection and missed detection of important regions. Third, many current 3D reconstruction methods rely on generalized temporal models such as standard LSTMs. These models often fail to capture subtle differences in tumor growth or spatial structure, which leads to low-quality reconstructions and limited clinical usefulness¹⁰. To address the specific challenges in lung segmentation, tumor detection, and 3D reconstruction, this study proposes a unified deep-learning framework that integrates all three modules to generate high-fidelity 3D representations of lung tumors.

Traditional methods for addressing class imbalance in lung segmentation operate at two levels. At the data level, techniques include undersampling the majority class, oversampling the minority class, and generating synthetic samples. At the algorithmic level, loss functions are reweighted to emphasize signals from the minority class. Although these strategies are partially effective, they often lead to overfitting, poor generalizability, and limited adaptability to complex spatial patterns in medical images¹¹. Deep reinforcement learning (DRL) has recently emerged as a promising alternative. It filters noise, sharpens critical features, and has proven effective in various domains. In DRL, the reward function can mitigate class imbalance by assigning higher rewards for the correct classification of minority-class samples and applying stronger penalties for their misclassification. This weighting scheme encourages the learning algorithm to prioritize underrepresented classes, thereby reducing bias toward the majority class and enhancing overall fairness in recognition¹². However, DRL methods often face the bias-variance trade-off, which can lead to unstable training and inconsistent results. In this paper, we adopt PPO to handle the imbalance problem in lung segmentation more effectively. PPO is an on-policy RL algorithm that stabilizes learning by preventing large policy updates through a clipping mechanism. It maintains a balance between exploration and exploitation and is computationally efficient for high-dimensional tasks with continuous variables. To improve performance and data efficiency, we use an off-policy version of PPO. This variant reuses agent-environment interactions stored in a replay buffer, which reduces the cost of training and enhances generalization in complex, high-dimensional environments such as medical image segmentation¹³.

To address the challenges of accurate tumor detection, particularly in regions with unclear boundaries or low contrast, our model incorporates a composite loss function consisting of classification loss, bounding box regression loss, and a novel adversarial loss component. Most conventional methods use only classification and bounding box regression losses for tumor classification and localization¹⁰. In contrast, our method includes adversarial learning to enhance feature-level realism and structural precision. The adversarial loss serves as a powerful regularizer, encouraging the generator to produce bounding boxes that are both geometrically accurate and feature-wise similar to those of real tumor regions. The discriminator compares the feature maps of predicted and actual tumor regions to provide detailed feedback. This feedback improves the ability of the generator to identify tumor-like regions, even when texture contrast is low or boundaries are unclear. This integration of adversarial loss effectively overcomes the limitations of conventional detection methods, which often fail under subtle or ambiguous visual conditions. As a result, our model delivers more reliable tumor localization and greatly enhances the clinical utility of the final 3D reconstructions.

After addressing challenges in lung segmentation and tumor detection, the next challenge lies in improving temporal modeling during 3D tumor reconstruction. We utilize a TLSTM network¹⁴ to address the limitations of existing temporal models in 3D tumor reconstruction. TLSTM adjusts its parameters based on the proximity and unique features of each input sample. Unlike conventional LSTM models that apply the same learning to all data, TLSTM dynamically adapts to the test data distribution. For example, standard LSTM models may not distinguish between slow-growing and fast-growing tumors when progression rates vary. In contrast, TLSTM adapts more effectively to these local differences. We also include a spatial attention mechanism in TLSTM. This enables the model to focus on the most important parts of the input sequence, improving the accuracy and clinical relevance of the 3D reconstructions.

After addressing the three core challenges separately, we now present the overall architecture of the proposed framework. This framework integrates the segmentation, detection, and reconstruction modules into a unified 3D modeling pipeline. The article introduces a triple-tiered strategy for creating 3D representations of lung cancer from CT scans. Initially, we employ a GAN featuring a U-shaped network (U-Net) architecture to segment the CT images. To address the class imbalance problem, we use the off-policy PPO approach. The discriminator applies dilated convolution layers to improve the generator output. It distinguishes real images from artificial ones and emphasizes essential features during training. In the second phase, the segmented

images are processed using another GAN configured with the mask region-based convolutional neural network (Mask R-CNN) architecture. This network accurately identifies and locates target areas within the images for precise tumor detection. To reinforce the accuracy of tumor localization, the generator applies a specific loss function, ensuring alignment between synthetic outputs and actual CT data. The discriminator supports output validation using dilated convolutions for robust feedback. The final phase involves 3D modeling, where a spatial attention-based TLSTM network is paired with a GAN. At this stage, the tumor images from the second step are fed into the trained EfficientNet network to extract features. These features are then passed to the TLSTM and subsequently to the generator, which constructs a 3D model based on the TLSTM outputs.

The effectiveness of our approach is robustly confirmed via extensive evaluation, and scrupulously compared with established methodologies utilizing the LIDC-IDRI dataset and standard performance indicators. This detailed juxtaposition underscores the superior performance and promising potential of our method to enhance the early detection of lung cancer. Table 1 lists the abbreviations and their corresponding definitions used throughout the study.

The critical contributions of our model include:

- Our model employs off-policy PPO to tackle the class imbalance commonly found in traditional lung segmentation methods. Off-policy PPO differs from on-policy methods as it utilizes data from various policies, which boosts the efficiency of the model. This feature is especially useful in medical contexts where data from infrequent but crucial conditions are limited. Additionally, we have incorporated a novel reward system within the off-policy PPO framework. This system focuses on the minority class by rewarding accurate identification and penalizing errors in classifying underrepresented classes, thereby encouraging a more balanced and fair classification process.
- A novel adversarial loss term is integrated into the generator loss function during tumor detection. Conventional methods rely only on classification and bounding box regression losses. In contrast, the adversarial component enhances the feature-level realism of predicted regions. This helps the model detect tumors more effectively, even when contrast is low or boundaries are unclear. This enhancement significantly reduces false detections and improves clinical reliability.
- The TLSTM method represents a substantial innovation in 3D tumor modeling. It enhances model training by weighing the influence of training samples based on their proximity to test samples. This approach significantly improves the precision and consistency in 3D representations, making it a valuable tool for accurate tumor morphology analysis.

The structure of this article is as follows: relevant research on lung cancer is reviewed next, followed by the basic concepts. The proposed methodology is then outlined, after which the findings and analysis of our experiments are presented. Finally, the paper concludes with closing remarks.

Related works

The advent of artificial intelligence (AI) has dramatically transformed various medical fields^{15–17}, enhancing diagnostic performance and treatment results in areas including skin, colon, and heart cancers^{18–23}. Lung cancer, notable for its high prevalence and mortality, requires sophisticated technological solutions for timely detection and accurate treatment planning^{24–27}. Specifically, creating 3D representations of lung tumors is essential to grasp tumor growth complexities and improve treatment methods.

This review meticulously divides the literature on lung cancer detection and analysis into three key sections: lung segmentation, tumor detection, and 3D tumor reconstruction. Each section discusses significant developments and methods that enhance diagnostic and treatment techniques in lung cancer care.

Lung segmentation

In 2023, Gite et al.²⁸ explored the effectiveness of U-shaped network++ (U-Net++) for lung segmentation in X-ray images to enhance tuberculosis (TB) detection. Their approach marks a significant departure from traditional methods by segmenting lungs prior to classification, which is shown to increase TB detection performance. The study compares U-Net++ with three other segmentation models, underscoring the advantages of using advanced segmentation before classification to prevent data leakage and improve diagnostic outcomes. Arvind et al.²⁹ have enhanced the U-Net architecture to develop a lighter, more efficient version for semantic lung segmentation from chest X-ray images. This modified U-Net includes multiple dropouts in all deconvolutional layers, effectively addressing overfitting and surpassing the performance of the original model. The redesigned architecture demonstrates robust generalization across diverse datasets, proving particularly effective in quickly identifying thoracic abnormalities and thereby facilitating prompt and accurate medical diagnoses. Chen et al.³⁰ proposed a novel lung-dense attention network (LDANet) to automatically segment lung parenchyma from CT images. This network features two main mechanisms: residual spatial attention (RSA) and gated channel attention (GCA). RSA emphasizes spatial details and suppresses irrelevant areas, while GCA adaptively adjusts channel weights to highlight critical features. Additionally, LDANet uses a dual attention guidance module (DAGM) to integrate features better and combines a lightweight dense block (LDB) with a positioned transpose block (PTB) to reuse features effectively and restore image resolution accurately. Ghali et al.³¹ developed a system for accurately segmenting lungs in chest X-ray (CXR) images, which is crucial for enhancing automated analysis and assisting radiologists. The proposed method utilizes five different models to delineate lung regions effectively, even in challenging cases involving varied lung shapes and disease-affected tissues. This approach incorporates dual loss functions to optimize the segmentation process, thereby improving the diagnostic reliability for healthy and unhealthy lungs. Murugappan et al.³² explored a robust deep-learning approach for lung segmentation in CT images using the DeepLabV3+ network. The study employs various pre-trained networks such as Extreme

Abbreviation	Definition
3D	Three-dimensional
CT	Computed tomography
GAN	Generative adversarial network
PPO	Proximal policy optimization
TLSTM	Transductive long short-term memory
LIDC-IDRI	Lung image database consortium image collection
HD	Hausdorff distance
ED	Euclidean distance
WHO	World Health Organization
X-rays	X-radiation
MRI	Magnetic resonance imaging
DRL	Deep reinforcement learning
U-Net	U-shaped network
Mask R-CNN	Mask region-based convolutional neural network
AI	Artificial intelligence
U-Net++	U-shaped++
TB	Tuberculosis
LDANet	Lung-dense attention network
RSA	Residual spatial attention
GCA	Gated channel attention
DAGM	Dual attention guidance module
LDB	Lightweight dense block
PTB	Positioned transpose block
CXR	Chest X-ray
Xception	Extreme inception
ResNet-18	Residual network-18
COVID-19	Coronavirus disease 2019
FL	Federated learning
SSSOA	Salp shuffled shepherd optimization algorithm
VGG16	Visual geometry group 16
CAD	Computer-aided diagnostic
EfficientNet B3	Efficient network B3
T-Net	T-shaped network
CenterNet	Center-based object detection network
NASNet	Neural architecture search network
TFDM	Differential memory
WOA	Whale optimization algorithm
SVM	Support vector machine
CapsNet	Capsule neural network
WSTSA	Wormhole and Salp swarm strategy enhanced tree-seed algorithm
HM-LeNet	Hybrid mobile LeNet
SNP	Single nucleotide polymorphism
LDN	Lightweight deep network
ANN	Artificial neural network
ROI	Region of interest
2D	Two-dimensional
GGO	Ground glass opacity
CBAM	Convolutional block attention module
ASPP	Atrous spatial pyramid pooling
ReLU	Rectified linear unit
DSC	Dice similarity coefficient
FSIM	Feature similarity index measure
APSO	Adaptive particle swarm optimization
XGBoost	Extreme gradient boosting
HRCT	High-resolution computed tomography
IoT	Internet of things
Continued	

Abbreviation	Definition
OFCMNN	Optimized fuzzy C-means neural network
KM-DTCL	Kernel multilayer deep transfer convolutional learning
Coarse Seg-net	Coarse segmentation subnetwork
Fine Seg-net	Fine segmentation subnetwork
Class-net	Classification subnetwork
ROC	Receiver operating characteristic
AUC	Area under the curve
HRNet	High-resolution network
DLN	Deep learning nomogram
ITF	Intrathoracic fat
IPN	Intranodular and perinodular regions
LASSO	Least absolute shrinkage and selection operator
PET	positron emission tomography
KAN	Kolmogorov–Arnold networks
SE	Squeeze-and-excitation
ViT	Vision transformer
Grad-CAM	Gradient-weighted class activation mapping
YOLOv8	You Only Look Once version 8
DCGAN	Deep convolutional generative adversarial network
FPN	Feature pyramid network
MSDA	Multi-scale dilation attention
GCSAM	Global channel spatial attention mechanism
CNDNet	Candidate nodule detection network
FPRNet	False positive reduction network
HPFF	Hierarchical progressive feature fusion
LUNA16	Lung nodule analysis 2016
GK	Gustafson and Kessel
TRPO	Trust region policy optimization
RNN	Recurrent neural network
MDP	Markov decision process
KL	Kullback-Leibler
EMD	Earth Mover's distance
BN	Batch normalization
FNIH	Foundation for the national institutes of health
FDA	Food and drug administration
XML	Extensible markup language
IoU	Intersection over union
RPN	Region proposal network
ADAM	Adaptive moment estimation
GPU	Graphics processing unit
GB	Gigabyte
FGSM	Fast gradient sign method

Table 1. List of abbreviations and their definitions used in the study.

Inception (Xception), Residual Network-18 (ResNet-18), and ResNet-50, refining the segmentation for two-class and four-class distinctions. The approach significantly enhances the delineation of lung fields and infected regions, which is crucial for effectively diagnosing Coronavirus disease 2019 (COVID-19). Ambesange et al.³³ developed an AI model using a federated learning (FL) framework combined with transfer learning to segment lung X-ray images efficiently. This approach maintains data privacy by training on local datasets and without requiring data centralization. The model employs pre-trained U-Net weights and utilizes local healthcare data to enhance training, further supported by explainable AI techniques to clarify model decisions.

In 2024, Hasan et al.³⁴ utilized the Deeplabv3plus CNN-based model to segment lung regions from X-ray images for enhanced tuberculosis detection. This model addresses the common issue of reduced feature resolution in deep networks by integrating atrous convolution and modifying pooling and convolutional striding mechanisms. This adaptation allows for more precise lung segmentation and optimizes the model for specific diagnostic tasks in pulmonary disease detection. Swaminathan et al.³⁵ addressed late-stage lung cancer diagnosis by implementing deep learning for earlier detection in high-risk patients. They enhanced CT

images with a Wiener filter, segmented them using a GAN trained with the Salp Shuffled Shepherd Optimization Algorithm (SSSOA), and then classified the segments with the Visual Geometry Group 16 (VGG16) CNN model. This complex multi-step process may introduce computational inefficiencies or require fine-tuning for different datasets. Suji et al.³⁶ developed a computer-aided diagnosis (CAD) system utilizing deep learning for lung cancer detection, specifically focusing on lung nodule segmentation. The method uses transfer learning to customize pre-trained weights from different encoders in an encoder-decoder framework and explores various combinations to find the best setup. The selected architecture, U-Net with an EfficientNet B3 backbone, was tailored for performance on the LIDC-IDRI lung cancer dataset. Thangavel and Palanichamy³⁷ introduced an advanced deep-learning framework for classifying pulmonary nodules using CT images, addressing the challenges of diverse and visually similar lung nodules. The approach integrates a T-shaped network (T-Net) for segmentation, a center-based object detection network (CenterNet) for extracting texture and intensity attributes, and a Neural Architecture Search network (NASNet) for classification, thereby ensuring thorough analysis without manual feature integration. This method leverages multiple preprocessing techniques to enhance data quality for effective learning and classification. Cai et al.³⁸ proposed a novel lung CT image segmentation method utilizing GANs. This method leverages the image translation capabilities of various GANs to transform original lung images into segmented outputs, aiming to enhance the effectiveness of pulmonary disease diagnosis by providing clear segmentation of lung tissues. Ramos and Pineda³⁹ developed a tri-phase, semi-automated lung CT image segmentation pipeline. This method initiates with preprocessing techniques such as contrast enhancement and noise reduction, followed by region growing for segmentation, and concludes with morphological closing to refine results. The approach is designed to provide precise segmentation without relying on extensive training datasets, making it particularly valuable in settings with limited data resources. Zheng et al.⁴⁰ introduced an advanced lung segmentation algorithm designed to tackle the complexities of lung images characterized by strong noise and uneven grayscale distribution. The method starts by extracting an initial lung mask using threshold techniques and gradient extraction with a time-series method called time-series differential memory (TFDM). Additionally, the approach incorporates an improved convex hull method for lung contour repair, addressing issues like contour loss due to solid nodules and other lesions. Guo et al.⁴¹ introduced an improved method for segmenting lung cancer pathological images. This method utilizes the whale optimization algorithm (WOA) and a random mutation strategy to enhance image segmentation. This approach is tailored to boost convergence speed and to circumvent local optima. The method is refined using a multilevel segmentation strategy tailored to lung cancer analysis. Pandey and Bhandari⁴² proposed a method for early detection of lung cancer using CT images, combining morphological filtering and contour-based image segmentation techniques to identify malignant tumors in their initial stages. Following segmentation, a Support Vector Machine (SVM) classifier differentiates malignant lung CT scan images from non-malignant ones.

In 2025, Vijayakumar et al.⁴³ developed a machine learning-based model to detect lung cancer using an advanced classifier and innovative segmentation methods. The process starts with preprocessing CT images using rolling guidance filtering and a switched mode fuzzy median filter, followed by segmentation with U-net. The segmented images are analyzed using a Capsule Neural Network (CapsNet), which preserves spatial relationships effectively and reduces feature loss, leading to faster and more efficient training. Qiao et al.⁴⁴ introduced a novel approach combining a wormhole and salp swarm strategy enhanced tree-seed algorithm (WSTSA) with a genetic algorithm for neural architecture search in medical image segmentation. This method automates the design of deep learning models, optimizes hyperparameters, and explores optimal network architectures to improve segmentation performance. The integration of WSTSA enhances the efficiency of architecture search, enabling the development of high-performing convolutional neural networks tailored for medical imaging tasks. Li et al.⁴⁵ proposed an advanced lung CT image segmentation framework that combines an attention mechanism with sobel edge detection to enhance the quality of lesion delineation. This method uses residual dynamic convolutions to enhance feature extraction. It includes an information fusion attention module that refines features with attention mechanisms and dilated convolutions, improving edge clarity and segmentation sharpness. Murugaraj et al.⁴⁶ developed a hybrid mobile LeNet (HM-LeNet) that integrates aspects of MobileNet and LeNet for enhanced segmentation and classification of lung vessels in CT images. This method begins with preprocessing using a non-local means filter, followed by lung lobe segmentation with K-Net and pulmonary vessel segmentation. The HM-LeNet then classifies lung abnormalities, such as emphysema, nodules, and pulmonary embolisms, leveraging advanced feature extraction capabilities. Fu et al.⁴⁷ introduced the single nucleotide polymorphism (SNP)-based lightweight deep network (LDN-SNP) model for segmenting COVID-19 CT images. This model features a novel neuron design known as SNP-type neurons. This model features a compact architecture with only 81,316 parameters, designed to overcome the challenges of limited labeled samples and computational inefficiency common in larger models. The network structure consists of an encoder using three specialized convolution modules and a decoder with a fourth module, achieving efficient training and deployment capabilities for practical medical use. Erciyes et al.⁴⁸ presented a deep learning method that utilizes an artificial neural network (ANN) in conjunction with preprocessing, adaptive augmentation, and an improved convolutional encoder-decoder based on a U-Net to address overfitting and enhance boundary delineation. Yang et al.⁴⁹ developed a 3D lightweight segmentation model that integrates region of interest (ROI) extraction via K-means clustering, attention-enhanced skip connections, and a combination of focal loss and the dice coefficient. The model also employs depth-wise separable convolutions, ghost-inspired 1×1 convolutions, and residual connections. Nguyen et al.⁵⁰ proposed a two-stage segmentation approach combining two-dimensional (2D) and 3D neural networks based on no-new-U-Net, enhanced with attention mechanisms and boundary loss to improve segmentation of ground glass opacity (GGO) in post-COVID-19 CT scans. Vincy et al.⁵¹ introduced a segmentation model using a 3D ResNet50 to encode lung CT slices. The model employs a dense feature extraction module and a U-Net decoder for multiscale extraction of tumor features. A combined Dice and focal loss were applied to address class imbalance. Jannat et al.⁵² suggested a deep learning model called

the lightweight residual U-Net, which integrates a convolutional block attention module (CBAM), atrous spatial pyramid pooling (ASPP), and attention mechanisms, trained using rectified linear unit (ReLU) and leaky ReLU activations. Notably, they found that the dice loss outperformed other losses.

Table 2 provides a comprehensive review of the most sophisticated models for lung segmentation. Most existing approaches adopt U-Net variants^{28,29,36,48}, attention mechanisms^{30,45}, or GAN-based frameworks³⁸ to classify each pixel as lung or non-lung. However, a shared limitation across these models is their restricted capacity to handle severe class imbalance, where non-lung pixels vastly outnumber lung pixels in CT or X-ray images. This often leads to biased classifiers that underperform on minority classes, particularly in atypical

Authors	Methodology	Contribution	Dataset	Result	Limitation
Gite et al. ²⁸	U-Net++ for lung segmentation	Enhances TB detection with advanced segmentation	138 X-ray images	IoU: 0.95	Specific to X-ray images, not other modalities
Arvind et al. ²⁹	Modified U-Net with dropouts	Reduces overfitting, enhances model generalization	900 X-ray images	Accuracy: 93.87%	Optimized for lung segmentation, less versatile for other organs
Chen et al. ³⁰	LDANet combines RSA and GCA	Enhances lung CT segmentation via dual attention	LIDC-IDRI	Dice similarity coefficient (DSC): 98.430%	Sensitive to variations in imaging protocols
Ghali et al. ³¹	Dual loss functions with five models	Optimizes CXR segmentation for various lung conditions	662 X-ray images	F1 score: 97.47%	May not generalize across all imaging equipment types
Murugappan et al. ³²	DeepLabV3 + with various networks	Enhances segmentation of lung and infected areas	750 CT images	IoU: 0.9971	Performance varies with different network layers
Ambesange et al. ³³	FL with U-net and transfer learning	Enhances lung X-ray segmentation while preserving data privacy	662 X-ray images	Accuracy: 98.92%	Limited by data variability across nodes
Hasan et al. ³⁴	DeepLabv3plus with Atrous Convolution	Optimizes feature resolution for segmentation	558 X-ray images	Accuracy: 97.42%	May struggle with highly variable image quality
Swaminathan et al. ³⁵	Wiener filter, GAN with SSSOA, VGG16 classification	Streamlines detection process with an integrated model	LIDC-IDRI	Accuracy: 97%	Relies on high-quality pre-processed images
Suji et al. ³⁶	U-Net with EfficientNet-b3, transfer learning	Optimizes nodule segmentation with transfer learning	LIDC-IDRI	IoU: 0.45	Performance variability across datasets
Thangavel and Palanichamy ³⁷	T-Net, CenterNet, NASNet with preprocessing	Automates nodule classification efficiently	LIDC-IDRI	DSC: 99.07%	May require fine-tuning for new datasets
Cai et al. ³⁸	GANs for image translation segmentation	Enhances lung CT segmentation with GANs	267 CT images	Accuracy: 89.63%	May not handle all types of lung anomalies
Ramos and Pineda ³⁸	Tri-phase semi-automated segmentation with preprocessing	Provides consistent, swift segmentation results	267 CT images	IoU: 0.9341	May require manual adjustments for optimal results
Zheng et al. ⁴⁰	Threshold-gradient with TFDM and convex hull repair	Enhances segmentation performance and robustness	2112 CT images	IOU: 0.9911	Requires precise calibration of threshold settings
Guo et al. ⁴¹	whale optimization algorithm-based random mutation strategy for image segmentation	Speeds up convergence, improves segmentation	25,000 CT images	Feature similarity index measure (FSIM): 81.57%	May not adapt well to other cancer types
Pandey and Bhandari ⁴²	Morphological filtering and SVM classification	Enhances tumor visibility for early intervention	LIDC-IDRI	Accuracy: 87.79%	May miss micro-tumors or highly diffuse anomalies
Vijayakumar et al. ⁴³	CapsNet with RGF and SMFMF, U-net segmentation	Enhances lung cancer detection efficiency	1097 CT images	Accuracy: 98%	May underperform with non-standardized data
Qiao et al. ⁴⁴	WSTSA with genetic algorithm	Streamlines optimal CNN architecture selection	2773 CT images	DSC: 72.40%	Dependent on initial algorithm parameters
Li et al. ⁴⁵	Attention and Sobel edge detection	Enhances lesion feature extraction and performance	LIDC-IDRI	Accuracy: 93.48%	May require high computational resources
Murugaraj et al. ⁴⁶	HM-LeNet with NLM filter and K-Net	Classifies lung abnormalities efficiently	23 CT scans images	Accuracy: 92.7%	Potential overfitting to specific abnormalities
Fu et al. ⁴⁷	LDN-SNP with SNP-type neurons	Efficient, compact segmentation of COVID-19 CT	100 CT scans images	Accuracy: 97.0% and DSC: 74.0%	Limited adaptability to other imaging tasks
Erciyes et al. ⁴⁸	ANN with optimized U-Net and augmentation	Reduced overfitting with enhanced segmentation	American association of physicists in medicine (AAPM) thoracic auto-segmentation	DSC: 94.68% and IoU: 0.8990	Limited adaptability to unseen pathologies
Yang et al. ⁴⁹	3D model with focal and dice loss	Improved deep and shallow feature representation	199 CT scans images	Accuracy: 99.90% and DSC: 56.10%	Overfitting risk from a small COVID dataset
Nguyen et al. ⁵⁰	Two-stage no-new-U-Net with attention and boundary loss	Improved GGO segmentation with fuzzy boundary handling	Post-COVID CT challenge	DSC: 71.93%	Difficulty generalizing to non-GGO lung lesions
Vincy et al. ⁵¹	3D ResNet50 encoder with dense-feature U-Net decoder	Enhanced multiscale tumor features from lung CT slices	LUNA16	IoU: 0.738	Contrast issues in small or diffuse tumors
Jannat et al. ⁵²	Lightweight residual U-Net with CBAM and ASPP	Improved lung mask extraction	Japanese society of radiological technology (JSRT)	DSC: 98.72%	Limited generalization on low-quality images

Table 2. Comparison of advanced models for lung segmentation.

or diffuse lung regions. Some studies employ types of loss terms like dice loss and focal loss^{31,49–52} to more accurately represent minority classes. However, these methods are often sensitive to image quality and depend on manually tuned parameters.

Our proposed method distinguishes itself by embedding an off-policy PPO strategy into a GAN-based segmentation framework. Unlike traditional static weighting methods, the PPO-based agent uses a reward system to learn how to focus on underrepresented lung regions and improve minority-class accuracy. The off-policy design enables the model to reuse past training experiences, thereby improving stability and generalization—a capability that is rarely addressed in existing segmentation methods. This reinforcement-driven segmentation process not only reduces bias but also avoids overfitting and the need for manual threshold tuning.

Tumor detection

In 2023, Vijh et al.⁵³ presented a hybrid bio-inspired algorithm combining the WOA with adaptive particle swarm optimization (APSO) to enhance feature selection in CNNs, aiming to heighten the performance of tumor detection in lung CT scans. Despite its innovative integration, one shortfall may be the potential for increased computational complexity that hybrid algorithms often introduce. Alshayegi and Abed⁵⁴ developed a machine-learning model based on extreme gradient boosting (XGBoost) to classify lung cancer CT images and identify nodule malignancy stages. This framework combines conventional machine learning with automatic segmentation, using DeepLabv3+ initialized with ResNet-18 for precise nodule segmentation. This comprehensive CAD system offers practical support for radiologists by facilitating accurate malignancy-stage identification and nodule segmentation. Halder and Dey⁵⁵ introduced a framework that uses atrous convolution for segmenting and characterizing lung nodules from high-resolution computed tomography (HRCT) images. The proposed model, featuring a two-layer atrous pyramid and residual connections, captures multi-scale features to improve lung nodule detection and classification.

In 2024, Venkatesh et al.⁵⁶ proposed a technique using patch processing alongside deep learning to classify lung tumors in CT images swiftly and efficiently. However, patch processing might miss contextual information from larger image areas, possibly affecting the robustness of the classification. Sundararajan et al.⁵⁷ developed a cutting-edge method for lung tumor detection that employs the cloud-based Internet of things (IoT) for data aggregation and an optimized fuzzy C-means neural network (OFCMNN) for segmentation, complemented by a kernel multilayer deep transfer convolutional learning (KM-DTCL) ensemble. Despite its sophistication, reliance on cloud-based IoT could raise concerns about data security and latency issues. Srinivasulu et al.⁵⁸ used autoencoders to detect lung malignancies in CT images, achieving high performance and refined feature extraction. Yet, the complexity of the model could hinder real-world applicability where computational resources are limited. Tang and Zhang⁵⁹ engineered a multi-functional network incorporating a unified base architecture and prediction refinement framework to enhance the segmentation and identification of lung nodules. This network features a coarse segmentation subnetwork (Coarse Seg-net) and a fine segmentation subnetwork (Fine Seg-net) alongside a cooperative classification subnetwork (Class-net). The Coarse Seg-net performs initial localization, which improves the Fine Seg-net and Class-net by leveraging task correlations for better segmentation and classification. Kongkham et al.⁶⁰ explored deep-learning applications for the segmentation of lung nodules using CT scans. The study harnesses CNNs to overcome the limitations of conventional segmentation methods such as thresholding and region-growing. By utilizing CNNs, this research significantly enhances the detection and delineation of lung nodules, capitalizing on their ability to learn complex patterns from extensive data. Gautam et al.⁶¹ developed a cutting-edge ensemble of deep-learning models to improve the classification of lung nodule severity from CT images. This approach harnesses the capabilities of three advanced convolutional neural networks: ResNet-152, DenseNet-169, and EfficientNet-B7, integrating them through a novel ensemble method that uses a weight optimization strategy based on receiver operating characteristic area under the curve (ROC-AUC) and F1-scores. Sadremomtaz and Zadnorouzi⁶² introduced an enhanced U-Net neural network for precisely segmenting pulmonary nodules in CT scans. The model improves feature extraction and network efficiency by integrating dilation residual structures and spatial-channel attention mechanisms, making it robust for diverse nodule characteristics. Pan et al.⁶³ introduced a semi-supervised training strategy called entropy minimization lesion-level data augmentation, which uses a novel augmentation technique to utilize limited labeled CT data. The data augmentation method accelerates feature learning for lesion contours, while entropy minimization reduces overfitting and improves prediction confidence. Rathana and Lokesh⁶⁴ developed a hybrid lung cancer detection and staging system that integrates deep learning techniques with active contour models for segmentation and a high-resolution network (HRNet) for stage classification. This comprehensive approach targets both the detection of pulmonary disease and the precise classification of lung cancer stages, enhancing diagnostic reliability.

In 2025, Shi and Zhang⁶⁵ developed an enhanced U-Net framework to tackle the challenges of segmenting small and ambiguously defined pulmonary nodules. The model incorporates three innovative modules: the locality-aware attention module, which intensifies the focus on areas containing nodules; the pixel transformer module, which refines semantic information processing; and the perceptual adaptation module, which adjusts to the scale of the receptive field effectively. This comprehensive approach aims to improve the effectiveness of early lung cancer diagnosis by precisely segmenting pulmonary nodules. Miao et al.⁶⁶ developed a multimodal deep learning nomogram (DLN) that utilizes features from pulmonary nodules and intrathoracic fat (ITF) to differentiate between benign and malignant nodules. The method incorporates precise intranodular and perinodular regions (IPN) segmentation, feature extraction via the Swin transformer, and feature selection using least absolute shrinkage and selection operator (LASSO) regression⁶⁷. Yu et al.⁶⁸ proposed a two-step multimodal neural architecture search framework for optimizing lung nodule classification using positron emission tomography (PET)/CT data. The method combines a sandwich rule and an in-place distillation strategy for training, followed by a parallel operation edge decomposition strategy to further refine performance. Jiang

et al.⁶⁹ proposed a framework for pulmonary nodule detection in CT images that integrates neural network architectures with specialized attention mechanisms. This framework includes a partial attention module based on Kolmogorov–Arnold networks (KAN), which enhances global representation learning within a conventional CNN. Additionally, the model incorporates an adaptive feature fusion module and a multi-slice partition channel priority attention module, specifically designed to improve the detection of small nodules by better handling small-scale features. Xue et al.⁷⁰ introduced a novel deep learning model called the squeeze-and-excitation vision transformer (SE-ViT), designed to enhance the interpretability and effectiveness of lung nodule classification. The SE-ViT integrates two attention mechanisms: self-attention for capturing relationships between different image patches and the SE mechanism to prioritize crucial information within those patches. This model addresses the challenges of lung nodule variability in shape and texture by focusing on the essential features that define malignancy, providing visual interpretations through gradient-weighted class activation mapping (Grad-CAM)-generated attention maps to aid clinical decision-making. Elhassan et al.⁷¹ proposed a dual-model framework that integrates a CNN for lung tissue classification and the You Only Look Once version 8 (YOLOv8) for real-time tumor localization, supported by a deep convolutional GAN (DCGAN)-based data augmentation to enhance model robustness. Abdulqader et al.⁷² introduced a deep learning framework that combines transformer-based attention layers, adaptive anchor-free detection, and an enhanced feature pyramid network (FPN). The model was trained with a multi-task loss function to optimize the detection, classification, and localization of lung tumors in CT scans. Liu et al.⁷³ proposed a lung nodule detection model based on YOLOv11. The model integrates MobileNetV4, multi-scale dilation attention (MSDA), and frequency fusion bidirectional FPN, along with novel loss functions to enhance feature fusion and prediction quality. Li⁷⁴ developed a lung nodule detection framework that combines the global channel spatial attention mechanism (GCSAM), candidate nodule detection network (CNDNet) with a Res2Net backbone, and a false positive reduction network (FPRNet). The framework is further enhanced by hierarchical progressive feature fusion (HPFF) to support robust multi-scale nodule representation.

Table 3 offers an exhaustive analysis of the most advanced models for detecting tumors. Many previous studies rely on standard loss configurations, mainly classification loss^{53–55} and bounding box regression loss^{61,71–74}, to separate tumor regions from non-tumor tissues. Although effective in many cases, these methods often struggle in low-contrast regions or where tumor boundaries are vague. For instance, the models by Viji et al.⁵³ and Gautam et al.⁶¹ employ CNN-based or ensemble classification strategies but do not address fine-grained structural accuracy. Similarly, segmentation-oriented studies, such as those by Halder and Dey⁵⁵ and Sadremomtaz and Zadnorouzi⁶², focus on multiscale feature extraction and attention mechanisms. However, they do not use adversarial techniques to ensure structurally realistic segmentation outputs. These limitations often result in increased false positives or false negatives, particularly in areas of ambiguity.

To address these challenges, our method uses a composite loss function within a specialized GAN architecture. This loss function combines classification loss, bounding box regression loss, and a new adversarial loss component. Unlike conventional loss functions, the adversarial component regularizes tumor structure by learning from differences between predicted and ground-truth tumor regions at the feature level. This reduces detection errors and improves diagnostic confidence.

3D tumor reconstruction

Recent medical progress has spurred the advancement of 3D models, greatly improving patient treatment approaches^{75–79}. Liver resection surgeries, for instance, benefit significantly from 3D models. These models offer surgeons detailed insights into liver anatomy, allowing for more precise preoperative planning and intraoperative guidance. They can also aid in identifying anatomical variations and potential complications, improving surgical outcomes^{80,81}. Additionally, research has explored generating 3D brain models from MRI data⁸².

A significant focus in current medical research is on reconstructing 3D models of lung tumors⁸³. Hong et al.¹⁰, Gu et al.⁸³, and Rezaei et al.⁸⁴ employed a GAN-based approach for 3D image reconstruction. Their approach involved snake optimization, a technique that uses an initial contour to iteratively refine the boundary of an object for lung segmentation. They also used Gustafson and Kessel (GK) clustering, a method for grouping similar data points into clusters, for tumor segmentation. Nonetheless, the performance of snake optimization is critically reliant on the positioning of the initial contour⁸⁵ since incorrect initial placements can result in segmentation inaccuracies. Snake optimization faces challenges with intricate lung structures, such as atypical shapes or nodules, which hampers its ability to accurately delineate boundaries. Furthermore, the performance of the GK clustering relies on selecting optimal initial centroids, affecting tumor segmentation quality. Challenges arise with overlapping tumors or heterogeneous characteristics, complicating accurate data point assignment to clusters⁸³. Karrar et al.⁸⁶ developed a method for 3D reconstruction of lung nodules from 2D CT scans to assist in early treatment decisions. Their technique involves a segmentation approach using bounding boxes and maximum intensity projection to extract lung nodules from candidate masses. A rule-based classifier, an algorithm that makes decisions based on predefined rules, selects the relevant nodules for 3D reconstruction. These nodules are then visualized through 3D Slicer (version 7) software using surface rendering techniques. However, the method may be limited by the effectiveness of rule-based classifiers and the quality of initial 2D scans, which can affect the reliability of nodule extraction and subsequent 3D modeling. Dlamini et al.⁸⁷ presented a framework for identifying, segmenting, and reconstructing non-small cell lung cancer tumors utilizing YOLOv4 and an active contour method. The framework is divided into two primary components: the detection and volumetric visualization phases. Detection involves image enhancement to reduce noise, augmentation to diversify the dataset, and labeling for accurate localization. In the volumetric rendering section, images undergo filtering, tumor extraction through clustering, and region-based active contour modeling for 3D reconstruction. This methodology utilizes enhanced images and precise contour data for effective tumor rendering. Challenges here include potential overfitting due to augmented data variability and the dependency

Authors	Methodology	Contribution	Dataset	Result	Limitation
Vijh et al. ⁵³	Hybrid WOA and APSO with CNN	Optimizes feature selection for tumor classification	120 CT images	Accuracy: 97.18%	May require extensive parameter tuning
Alshayegi and Abed ⁵⁴	XGBoost and DeepLabv3+	Integrates machine learning and advanced segmentation for comprehensive nodule analysis	LIDC-IDRI	Accuracy: 99.65%	Requires detailed nodule characteristics
Halder and Dey ⁵⁵	Atrous pyramid and residual connections	Applies atrous convolution for detailed multi-scale feature extraction	LIDC-IDRI	Accuracy: 95.97% and DSC: 97.15%	May perform variably across different CT scanners
Venkatesh et al. ⁵⁶	Patch processing with deep learning	Introduces patch-based preprocessing for enhanced classification	LIDC-IDRI	Accuracy: 95.70%	Relies on high-quality CT image preprocessing
Sundarrajan et al. ⁵⁷	Cloud-IoT with OFCMNN and KM-DTCL	Integrates cloud-IoT for data collection and segmentation-classification pipeline	LIDC-IDRI	Accuracy: 93.86%	Dependence on cloud-IoT infrastructure
Srinivasulu et al. ⁵⁸	Autoencoder	Integrates advanced neural network structures for nuanced feature extraction in lung tumor detection	Open-source and online portals	Accuracy: 96%	May not generalize well to less curated, real-world clinical data
Tang and Zhang ⁵⁹	Shared backbone and distillation	Leverages task correlations for enhanced performance	LIDC-IDRI	Accuracy: 91.9%	Require careful tuning of task interdependencies
Kongkham et al. ⁶⁰	CNNs for lung nodule segmentation	Introduces advanced CNN techniques to refine nodule segmentation	LIDC-IDRI	DSC: 85.3%	Requires large annotated datasets for optimal training
Gautam et al. ⁶¹	ResNet-152, DenseNet-169, EfficientNet-B7 ensemble	Enhances nodule classification through a weighted ensemble approach	LIDC-IDRI	Accuracy: 97.23%	Requires sophisticated model integration skills
Sadremomtaz and Zadnorouzi ⁶²	Enhanced U-Net with dilation residuals and attention mechanisms	Enhances feature extraction and network efficiency	LIDC-IDRI	Accuracy: 99.92%	Initial model setup and training can be computationally intensive
Pan et al. ⁶³	semi-supervised training strategy for entropy minimization lesion-level data augmentation	Enhances training efficiency with novel data augmentation	LIDC-IDRI	DSC: 81.06%	Requires careful balancing in parameter tuning
Rathan and Lokesh ⁶⁴	Active contour & HRNet for lung cancer	Utilizes hybrid imaging techniques for nuanced disease staging	LIDC-IDRI	Accuracy: 98.4%	Requires high-quality imaging for optimal performance
Shi and Zhang ⁶⁵	Advanced UNet with attention and adaptation	Tailors segmentation sensitivity to nodule characteristics	LIDC-IDRI	DSC: 89.85% and IOU: 0.8960	Requires calibration for variable nodule sizes
Miao et al. ⁶⁶	DLN with ITF and IPN integration	Integrates ITF features to enhance nodule characterization	1,385 CT images	Accuracy: 82.3%	Dependent on the quality of fat and nodule imaging
Yu et al. ⁶⁸	Two-step multimodal one-shot neural architecture search for PET/CT pulmonary nodules	Introduces a structured two-step neural architecture search to optimize architecture search	499 CT images	Accuracy: 94.23%	Intensive computational resources required
Jiang et al. ⁶⁹	KAN and adaptive feature fusion	Enhances nodule detection by optimizing feature representation	Lung nodule analysis 2016 (LUNA16)	Sensitivity: 94.73%	Require extensive training data for optimization
Xue et al. ⁷⁰	SE-ViT with dual attention	Integrates self-attention and SE mechanisms for detailed feature analysis	LUNA16	Accuracy: 86.3%	High dependency on the quality of training data
Elhassan et al. ⁷¹	CNN + YOLOv8 with DCGAN augmentation	Enhanced real-time tumor detection and classification	Iraq-oncology teaching hospital /national center for cancer diseases (IQ-OTHNCCD)	IOU: 0.85	Sensitivity to imaging conditions and variations
Abdulqader et al. ⁷²	Transformer, anchor-free, FPN with multi-task learning	Unified detection, classification, and localization of lung tumors	1608 CT images	IoU: 0.9576	Complex training pipeline for clinical deployment
Liu et al. ⁷³	YOLOv11 with MobileNetV4, MSDA, frequency fusion bidirectional FPN, slide loss	Improved detection via multi-scale attention and feature integration	LUNA16	IoU: 0.946	High architectural complexity for deployment
Li ⁷⁴	CNDNet + FPRNet with GCSAM and HPFF	Enhanced multi-scale detection and false positive reduction	LUNA16	Sensitivity: 97.7%	May misclassify nodules with unclear margins

Table 3. Comparison of advanced models for tumor detection.

on the effectiveness of the YOLOv4 model in varying imaging conditions. Shi et al.⁸⁸ developed a novel model utilizing ViT blocks to reconstruct 3D lung volumes from single 2D chest x-ray images. This model uses global information and long-range relationships to improve 3D reconstruction. It incorporates a pooling layer to refine feature extraction and employs patch discriminators to generate smoother and more realistic 3D lung models. Additionally, subject demographics are used as auxiliary inputs to refine reconstruction performance. This approach merges traditional imaging data with demographic insights, enhancing the fidelity of the output. Najafi et al.² and Huang et al.⁸⁹ proposed a three-step framework that combines GAN, LSTM, and a VGG16. Two customized GANs perform lung and tumor segmentation, followed by VGG16-LSTM feature extraction and final 3D tumor reconstruction.

Table 4 presents an exhaustive analysis of the most advanced models for the 3D reconstruction of lung tumors. Several studies, such as those by Hong et al.¹⁰ and Karrar et al.⁸⁶, rely on statistical modeling techniques or rule-based bounding box extraction, which often limit the flexibility of reconstruction when dealing with irregular or noisy tumor structures. In contrast, deep learning-based approaches, such as Najafi et al.² and Huang et al.⁸⁹, utilize LSTM for sequence modeling and feature transfer. However, these methods are limited in capturing localized data variations because standard LSTM architectures rely on generalized temporal patterns learned

Authors	Methodology	Contribution	Dataset	Result	Limitation
Hong et al. ¹⁰	VGG and GK clustering	Introduces GAN-based 3D reconstruction for lung tumor imaging	LUNA16	HD: 2.82 and ED: 2.82	Performance may vary for non-lung tumor datasets
Gu et al. ⁸³	ResNet and GK clustering	Combines GAN and LSTM for feature-driven 3D tumor reconstruction	LUNA16	HD: 2.99 and ED: 1.06	Requires well-segmented 2D input images
Rezaei et al. ⁸⁴	VGG and GK clustering	Uses GAN for 3D reconstruction and transfers 2D image features	LUNA16	HD: 3.02 and ED: 1.06	Relies heavily on accurate segmentation in preprocessing
Karrar et al. ⁸⁶	3D reconstruction using bounding boxes and rule-based classifier	Develops an approach for nodule extraction and surface rendering for 3D modeling	LIDC-IDRI	Accuracy: 99.6627%	Limited robustness to nodules with irregular shapes
Dlamini et al. ⁸⁷	YOLOv4 and region-based active contour model	Integrates detection and volumetric rendering in a single automated pipeline	LIDC-IDRI	Accuracy: 99.74% and DSC: 92.19%	Requires robust preprocessing for noise removal
Shi et al. ⁸⁸	ViT and demographics	Enhances 3D reconstruction from 2D images using ViT	2525 chest x-rays	DSC: 76.9%	Dependent on the inclusion of accurate demographic data
Najafi et al. ²	GAN, LSTM, and VGG16	Integrates GANs with VGG16 and LSTM for accurate 3D reconstruction	LIDC-IDRI	HD: 0.986 and ED: 1.126	Training instability due to sensitivity to input variations and multi-model tuning
Huang et al. ⁸⁹	GAN, attention-based LSTM, and VGG16	Integrates multi-stage GANs with VGG16 and attention-based LSTM for accurate 3D reconstruction	LIDC-IDRI	HD: 2.963 and ED: 1.725	Requires extensive labelled data and balanced segmentation to maintain accuracy across GAN stages

Table 4. Comparison of advanced models for the 3D reconstruction of lung tumors.

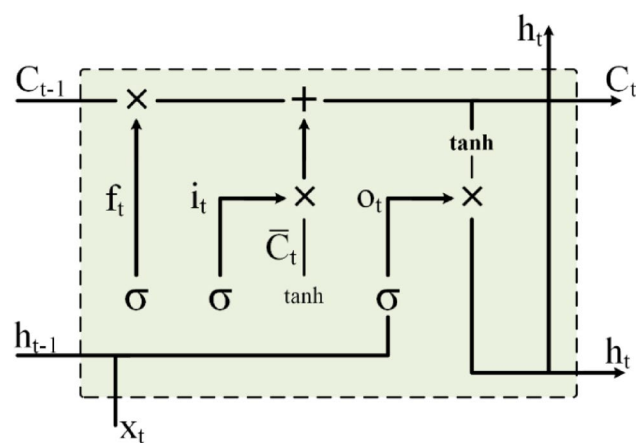


Fig. 1. Architecture of an LSTM unit. It displays the key components, including the input gate (i_t), forget gate (f_t), output gate (o_t), memory cell (c_t), and hidden state (h_t), which collectively manage the flow of information over time, ensuring data is appropriately stored, retained, and retrieved.

from the full training dataset. This global modeling may not generalize well when reconstructing fine-grained, patient-specific tumor boundaries.

To overcome these challenges, our model introduces a hybrid GAN and TLSTM architecture, where the TLSTM component adjusts its weight assignment based on the proximity of data points to the target rather than relying solely on global sequence trends. This localized adaptation enhances model sensitivity in the edge regions of tumors, improving reconstruction fidelity, particularly in complex or atypical cases. By incorporating adversarial learning alongside transductive temporal modeling, the proposed method provides both structural realism and data-specific precision, distinguishing it from prior models.

Background

This section briefly overviews TLSTM, spatial attention, RL, trust region policy optimization (TRPO), and PPO.

TLSTM

LSTM is designed to handle the shortcomings of traditional recurrent neural networks (RNNs), precisely their inability to learn long-range dependencies in sequence data^{90,91}. An LSTM is structured to remember information for extended periods, which is crucial in tasks like language modeling and time series prediction⁹².

The LSTM model⁹³ uses a gating mechanism to control information flow. Figure 1 illustrates the structure of an LSTM unit, which oversees the storage, retention, and retrieval of data from the memory cell over time. LSTMs feature three unique gates for effective information processing. The various components of an LSTM, including the input gate (i_t), forget gate (f_t), output gate (o_t), memory cell (c_t), and hidden state (h_t) for a specific time t and input x_t , are defined by Eqs. 1–5¹⁴:

$$i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + W_{ci}c_{t-1} + b_i) \quad (1)$$

$$f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + W_{cf}c_{t-1} + b_f) \quad (2)$$

$$c_t = f_t c_{t-1} + i_t \tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c) \quad (3)$$

$$o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + W_{co}c_t + b_o) \quad (4)$$

$$h_t = o_t \tanh(c_t) \quad (5)$$

Here, $\sigma(\cdot)$ serves as the sigmoid activation function of the gate and can take the form of either a logistic sigmoid or a hyperbolic tangent function. The hyperbolic tangent, denoted by $\tanh$, normalizes the output within the range of -1 to 1. W_{xk} , where $k \in \{i, f, o, c\}$, is associated with the input, forget, output gates, and the cell state, respectively. Similarly, W_{hk} , where $k \in \{i, f, o, c\}$, denotes the weights connected to the previous hidden state h_{t-1} . Moreover, W_{ck} , where $k \in \{i, f, o\}$, represents the diagonal weight arrays that link the cell state to these respective gates. Bias values for the gates and the memory cell equations are indicated as b_i , b_f , b_c , and b_o , respectively. The LSTM parameters, including weights and biases, are denoted as column vectors w_{lstm} and b_{lstm} . The overall operation of the LSTM is summarized in Eq. 6, where the functions $g(\cdot)$ and $f(\cdot)$ are based on Eqs. 1–5¹⁴.

$$\begin{cases} c_t = f(c_{t-1}, h_{t-1}, x_t; w_{lstm}, b_{lstm}) \\ h_t = g(h_{t-1}, c_{t-1}, x_t; w_{lstm}, b_{lstm}) \end{cases} \quad (6)$$

After the debut of the LSTM model, numerous advancements and variations, such as TLSTM, have been developed to enhance its performance across various studies. The state space of TLSTM is defined in Eq. 7¹⁴:

$$\begin{cases} c_{t,\eta} = f(c_{t-1,\eta}, h_{t-1,\eta}, x_t; w_{lstm,\eta}, b_{lstm,\eta}) \\ h_{t,\eta} = g(h_{t-1,\eta}, c_{t-1,\eta}, x_t; w_{lstm,\eta}, b_{lstm,\eta}) \end{cases} \quad (7)$$

In this context, η represents an unseen sequence. The formulation of Eq. 7 significantly differs from that of Eq. 6, as the parameters of the model adjust based on the characteristics of the evaluation point. This adaptation reflects the modifications introduced by new data, represented by η and denoted as $z(\eta)$.

Spatial attention

Spatial attention allows TLSTM to selectively concentrate on particular input data segments simultaneously instead of uniformly processing the entire dataset. It highlights essential variables and attributes, protecting the model from irrelevant data or noise⁹⁴.

Spatial attention is typically implemented as an intermediate layer within the TLSTM architecture, positioned between the input and the TLSTM layers. It aggregates the input data in a weighted manner, where the weights are assigned during training according to the significance of each input feature. The input to this spatial attention layer is an $N \times D$ matrix X , where N represents the number of temporal instances, and D represents the number of features. The spatial attention mechanism assigns weights based on the input data and the parameters learned during training, represented by W and b . These weight assignments are specified in Eq. 8:

$$a = \text{softmax}(XW + b) \quad (8)$$

The subsequent step is the computation of the weighted input sum as described in Eq. 9:

$$X' = Xa \quad (9)$$

This weighted sum condenses the input while steering the attention of the model toward the most impactful features.

Spatial attention boosts the adaptability of TLSTM by selectively emphasizing critical features. It minimizes the likelihood of overfitting by curbing excessive focus on non-essential noise or details, which could distort the learning outcomes. Moreover, this focused strategy aids in managing high-dimensional data by reducing the computational load associated with processing less critical information. Consequently, a TLSTM outfitted with spatial attention excels at processing varied and intricate datasets, yielding more uniform and widely applicable results. Such enhancements in model performance are especially vital in situations where detecting subtle data variations is key to making accurate predictions.

RL

RL represents a process where an agent enhances its decision-making abilities by undertaking actions that amplify the rewards obtained from its environment. This evolving procedure is configured utilizing Markov decision processes (MDPs). An MDP is characterized by elements (S, A, P, ρ_0, r) , where S embodies the complete set of potential states, and A includes all feasible actions. The function $P: S \times A \times S \rightarrow R$ calculates the likelihood of transitioning from one state to another based on the selected action. The function $\rho_0: S \rightarrow R$ represents the initial state probability distribution, and $r: S \times A \rightarrow R$ provides the reward feedback for each state-action pair.

In RL, at each timestep t , an observer assesses the present state s_t and decides on an action a_t guided by a strategy $\pi: S \times A \rightarrow [0,1]$. After executing this action, the environment rewards the decision-maker with $r(s_t, a_t)$ and transitions to the next state s_{t+1} . Throughout this sequence, the observer calculates the aggregate

discounted reward starting from timestep t , represented as $R_t = \sum_{k=t}^{\infty} \gamma^{k-t} r(s_k, a_k)$, where γ is the discount factor that values future rewards more heavily. Based on this return, the value functions for the state $V_{\pi}(s_t)$, the action $Q_{\pi}(s_t, a_t)$ at the state-action pair, and the advantage $A_{\pi}(s_t, a_t)$ are determined as follows:

$$V_{\pi}(s_t) = E_{a_t, s_{t+1}, \dots \sim \pi} \left[\sum_{k=t}^{\infty} \gamma^{k-t} r(s_k, a_k) \right] \quad (10)$$

$$Q_{\pi}(s_t, a_t) = E_{s_{t+1}, a_{t+1}, \dots \sim \pi} \left[\sum_{k=t}^{\infty} \gamma^{k-t} r(s_k, a_k) \right] \quad (11)$$

$$A_{\pi}(s_t, a_t) = Q_{\pi}(s_t, a_t) - V_{\pi}(s_t) \quad (12)$$

The goal is to formulate a policy, π , that maximizes the cumulative discounted rewards from the initial state, as detailed below⁹⁵:

$$\eta(\pi) = E_{s_0, a_0, \dots} [R_0] = E_{s_0, a_0, \dots} \left[\sum_{t=0}^{\infty} \gamma^t r(s_t, a_t) \right] \quad (13)$$

TRPO

To enhance the performance objective outlined in Eq. 13, TRPO modifies strategies using available policy data to optimize a different objective, ensuring the Kullback-Leibler (KL) divergence remains below a predetermined limit, thus keeping policy changes restrained⁹⁶:

$$\max_{\pi} E_{s \sim \rho} E_{a \sim \pi_{old}} \left[\frac{\pi_{\theta}(a|s)}{\pi_{\theta_{old}}(a|s)} A_{\pi_{old}}(s, a) \right] \quad (14)$$

subject to

$$E_{s \sim \rho} [D_{KL}(\pi_{old}(\cdot|s) \parallel \pi(\cdot|s))] \leq \delta \quad (15)$$

Here, π_{old} denotes the previous policy, and δ represents the allowable divergence limit. The expression $D_{KL}(\pi_{old}(\cdot|s) \parallel \pi(\cdot|s))$ calculates the KL divergence, reflecting the extent to which the novel policy π deviates from the earlier policy π_{old} for a given state s . The variable $\rho_{\pi_{old}}$ signifies the discounted state distribution originating from the initial state s_0 under the prior policy π_{old} , formulated as $\rho_{\pi_{old}}(s) = \sum_{t=0}^{\infty} \gamma^t P(s_t = s | s_0, \pi_{old})$. Targeting the alternate objective without restraining the KL divergence might induce drastic, unforeseen changes in the policy.

PPO

To reduce significant policy shifts, PPO utilizes an adapted optimization target termed the clipped surrogate objective, arranged as described below⁷³:

$$L_{PPO}^{CLIP} = E_{s \sim \rho} E_{a \sim \pi_{old}} \left[\min \left(\frac{\pi(a|s)}{\pi_{old}(a|s)} A_{\pi_{old}}(s, a), \text{clip} \left(\frac{\pi(a|s)}{\pi_{old}(a|s)}, 1 - \epsilon, 1 + \epsilon \right) A_{\pi_{old}}(s, a) \right) \right] \quad (16)$$

where ϵ , a small positive scalar, is set to safeguard policy stability while experimenting with novel actions. The advantage function $A_{\pi_{old}}(s, a)$ appraises the efficacy of an action a at state s under the prior policy π_{old} , relative to the expected results from alternative actions at the state. This function assesses whether an action outperforms or underperforms the typical expectations under the existing policy. In this methodology, clipping curtails excessive alterations to the policy during training, promoting gradual and reliable enhancements. The specifics of this clipping are outlined below⁹⁷:

$$\text{clip}(x, a, b) = \max(a, \min(b, x)) \quad (17)$$

In this context, x is the variable being adjusted, with a defining the minimum limit and b marking the maximum limit of the interval restricting x . The surrogate objective, detailed in Eq. 16, aims to curb substantial policy alterations by imposing penalties on changes that drastically shift $\frac{\pi(a|s)}{\pi_{old}(a|s)}$ away from 1. However, a significant limitation of PPO is its heavy reliance on on-policy data, which makes the data collection process more challenging because it integrates only a small portion of off-policy data, requiring frequent interactions between the agent and its environment.

Materials and methods

Figure 2 shows the innovative framework of our suggested model, which includes three critical phases: lung segmentation, tumor detection, and 3D tumor reconstruction. The primary aim of lung segmentation is to achieve high performance in separating lung regions from surrounding thoracic structures within imaging data. This meticulous step is crucial as it ensures that the subsequent processes, tumor detection and 3D reconstruction, rely on accurate inputs. Our lung segmentation approach utilizes a GAN with U-Net architecture, specifically designed for detailed pixel-by-pixel categorization. However, this method encounters challenges due to class imbalance, which is the disproportionate representation of lung and non-lung pixels in the dataset, with non-lung pixels being more prevalent. To address this imbalance, we employ an off-policy PPO algorithm that

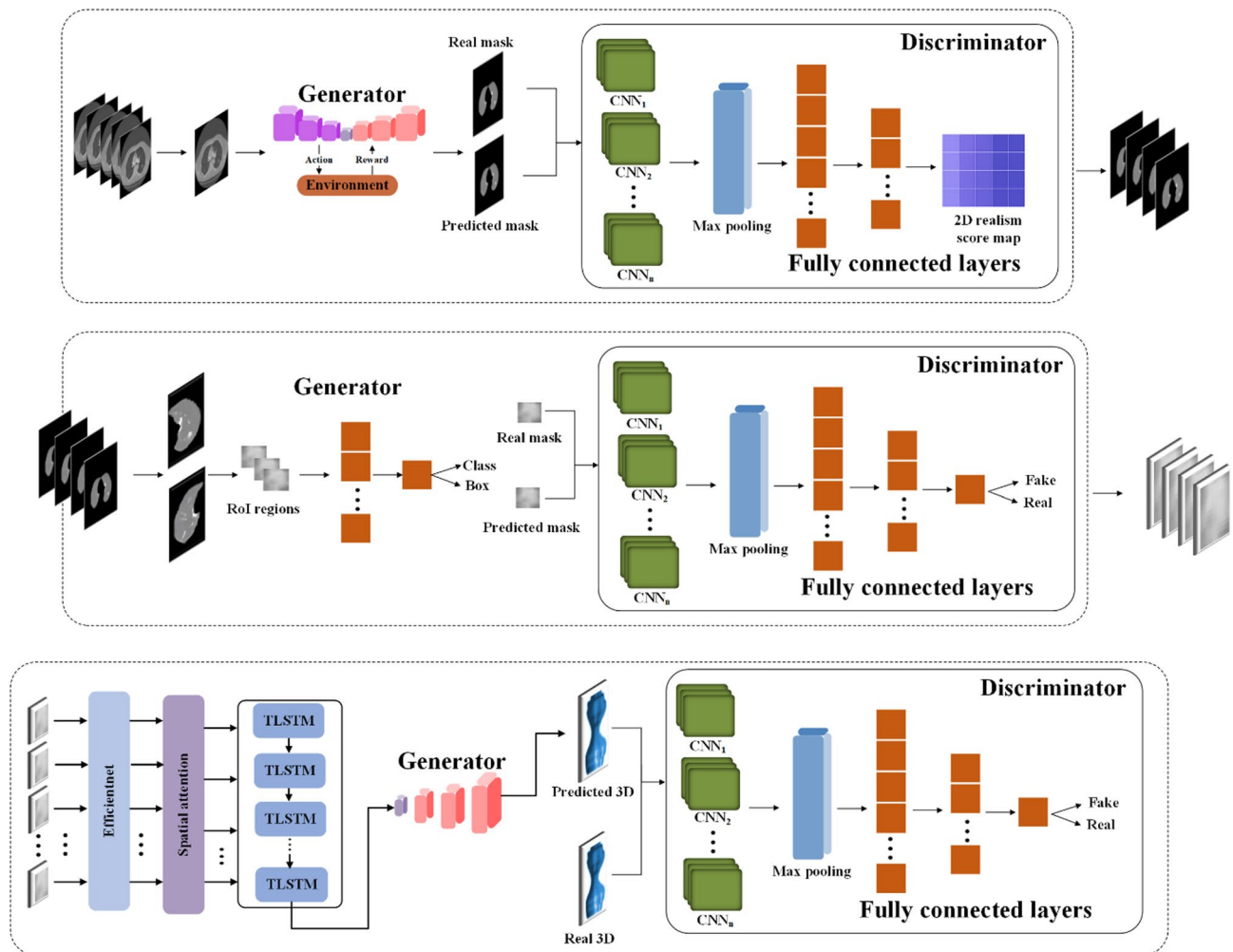


Fig. 2. Comprehensive framework of the proposed model. The figure demonstrates the integrated process encompassing lung segmentation, tumor detection, and 3D tumor reconstruction. The lung segmentation phase employs GAN and U-Net architectures, utilizing off-policy PPO to address imbalanced classification. The tumor detection model incorporates a GAN alongside Mask R-CNN, enhanced by a novel loss function for more accurate predictions. Finally, the 3D reconstruction of tumors is facilitated through a spatial attention-based TLSTM network paired with a GAN, optimizing detail and fidelity in rendering tumor volumes.

prioritizes non-lung pixels over lung pixels. The off-policy characteristic of PPO enables the utilization of a wide spectrum of historical data, diminishing the dependency on potentially biased recent data.

Following lung segmentation, the next phase is tumor detection. This stage involves examining the segmented lung images to identify potential tumor regions. Here, we integrate a GAN alongside the Mask R-CNN framework. This powerful tool combines the capabilities of object detection and instance segmentation, which effectively identifies and localizes the affected areas within the segmented images for precise tumor detection and categorization. Additionally, an error correction mechanism is integrated into the loss function of the generator to identify and rectify any discrepancies between the generated images and actual CT scans, thus enhancing the quality of the detected regions.

The final stage of our proposed model is 3D tumor reconstruction. This process involves creating a 3D tumor model from the detected tumors. Our model for 3D tumor reconstruction employs a spatial attention-based TLSTM network, a type of recurrent neural network that can process data sequences while also considering the spatial relationships between the sequence elements paired with a GAN. The TLSTM network processes features extracted from 2D pulmonary scans by an EfficientNet network and uses transductive learning, adjusting its weights based on proximity to new data points, to significantly enhance the prediction of critical tumor-related features. The generator then creates a 3D model based on the outputs from the TLSTM network.

In all GANs, the role of the discriminator is to evaluate the images generated by the generator. Discriminator networks use dilated convolution layers, a convolutional layer that can expand the receptive field without compromising image resolution or adding significant computational overhead. This is especially valuable in processing high-resolution medical images, where maintaining detail is crucial. Dilated convolutions enable

the discriminator to capture a wider context and finer details across larger image areas, more accurately differentiating between real and generated images.

Lung segmentation

The primary goal of lung segmentation is to generate accurate lung contours that meticulously replicate the original reference outlines. In this paper, we employ a GAN for lung segmentation, which includes two networks: a generator and a discriminator. The generator excels at detecting essential features and patterns in the scans, creating outlines that accurately map the lung areas. These outlines are designed to capture the complex shapes and edges necessary for thorough segmentation. On the other hand, the discriminator works with the generator to enhance the quality of these outlines. Acting as a strict evaluator, the discriminator checks how well the generated outlines match the original ones. It uses the Earth Mover's distance (EMD), a detailed measurement, to assess how much the generated outlines differ from the actual ones, helping to improve the outputs of the generator. This reciprocal evaluation process encourages the generator to create contours that better match the original, thus improving the segmentation quality. During training, the generator processes a lung CT scan image, denoted as I_i , and produces a corresponding outline M_i that highlights the lung areas. The discriminator then examines this generated outline M_i for its fidelity and quality, providing essential feedback that enhances the performance of the generator. This iterative cycle, where the generator and discriminator continuously interact and refer to the original outlines, is crucial in enhancing the quality of the lung contours produced during training.

Generator

In our specific context, segmentation involves classifying each pixel in an image as belonging to the lung area. When processing a CT scan labeled I_i , the generator network assesses each pixel to determine whether it is part of the lung region, then creates a corresponding mask M_i that reflects this classification. The structure of the generator is depicted in Fig. 3. The generator employs a U-Net for segmenting lungs, consisting of two

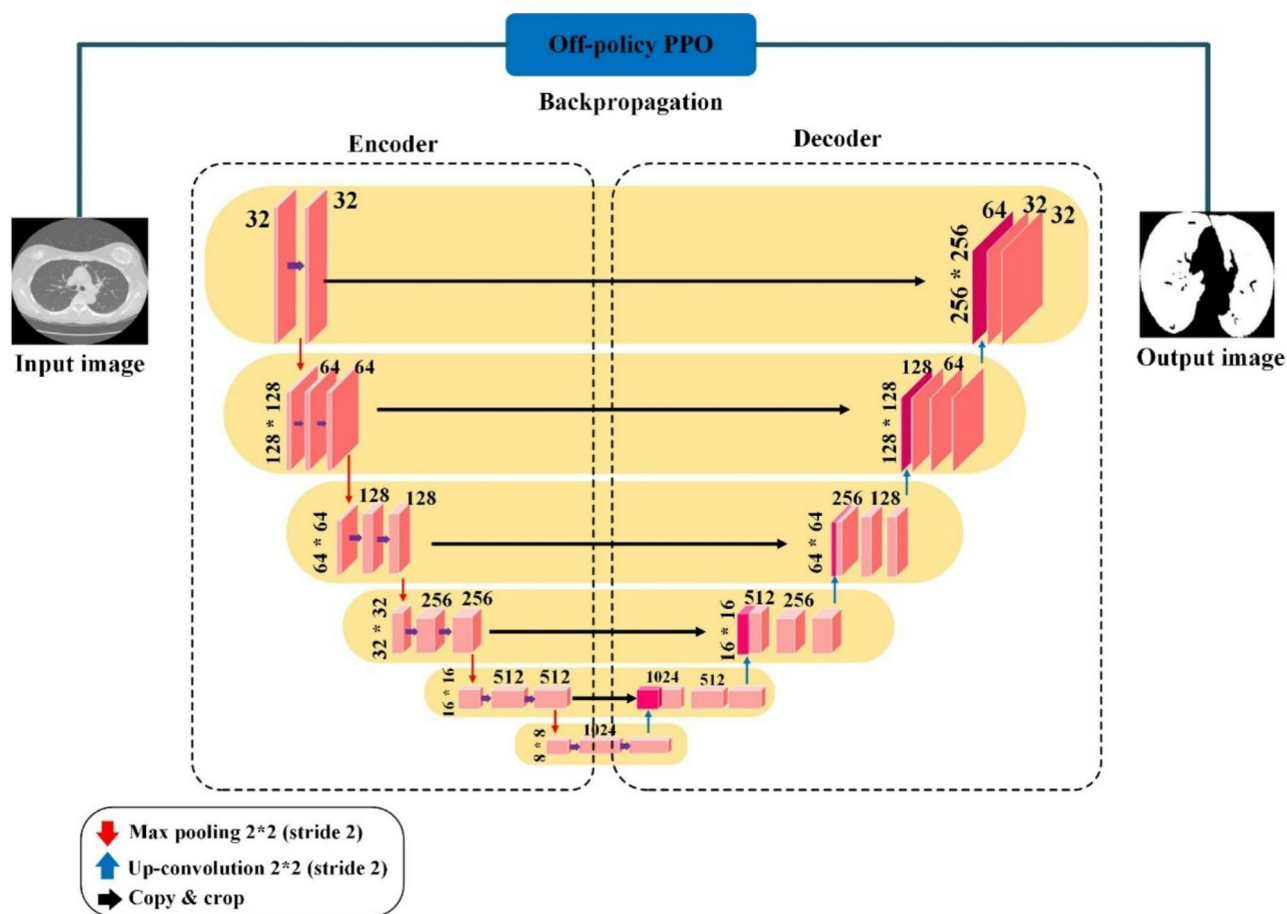


Fig. 3. Architecture of U-Net-based generator for lung segmentation. This diagram illustrates the dual-component structure of the generator, consisting of an encoder and a decoder. The encoder processes the input CT scan image, extracting features at various scales through convolution layers to capture detailed information. The decoder then uses these features to construct precise masks delineating lung regions. In this model, the off-policy PPO algorithm is integrated, which enhances the ability of the model to classify each pixel accurately, addressing the challenge of class imbalance by prioritizing the detection of lung area pixels.

main parts: the encoder and the decoder. The encoder, a key player in this process, receives the input image and extracts features at various scales via a sequence of convolution layers. These layers are intricately designed to grasp a wide spectrum of nuances and hierarchical details from the image with high fidelity. Following this, the decoder employs these hierarchical features to craft the masks, meticulously reconstructing the complex forms and contours of the lung regions to generate precise and detailed masks. The encoder and decoder incorporate convolution layers, which are essential in detecting local patterns and spatial correlations necessary for precise segmentation.

The generator acts as a binary pixel classifier, labeling each pixel as part of the lung area (marked as one) or not (marked as zero). The prevalence of non-lung pixels (zeroes) leads to a class imbalance, causing the classifier to favor predictions of non-lung areas and reducing its performance in identifying lung regions. To rectify this, in addition to adversarial loss, we implement an off-policy PPO algorithm in our training strategy, effectively addressing the imbalance. This algorithm, equipped with custom reward and penalty systems, optimally addresses this issue by offering higher rewards or penalties for correct or incorrect identifications of lung pixels (the minority class) compared to non-lung pixels (the majority class). This adjustment helps the model to concentrate more accurately on recognizing lung area pixels. By boosting incentives for correctly identifying the minority class, the off-policy PPO algorithm directs the focus of the generator toward lung areas, thus correcting the bias towards the abundant non-lung pixels. This strategic enhancement balances the focus of the classifier between both classes and significantly improves the performance of the model in segmenting lung regions from CT images.

In the proposed off-policy PPO algorithm, the framework for state, action, and reward is defined as follows:

- State s_t : It represents the observed pixel at a specific time step t .
- Action a_t : It is committed to forecasting the label of every pixel as lung or non-lung.
- Reward r_t : It reflects the effectiveness of a specific action. If an agent correctly classifies, it receives a positive incentive; if not, a negative one is issued. The magnitude of these incentives should vary between the classes. In this study, incentives for actions are determined based on the equation below:

$$r_t(s_t, a_t, y_t) = \begin{cases} +1, a_t = y_t \text{ and } a_t \in D_L \\ -1, a_t \neq y_t \text{ and } a_t \in D_L \\ +\lambda, a_t = y_t \text{ and } a_t \in D_N \\ -\lambda, a_t \neq y_t \text{ and } a_t \in D_N \end{cases} \quad (18)$$

Where D_L and D_N represent the minority (lung) and majority (non-lung) classes, respectively. To tackle class imbalance, the reward function is asymmetrically designed to assign higher importance to correctly classifying minority (lung) pixels. Specifically, a correct classification from the minority class yields a strong positive reward (+1), while a misclassification incurs a strong penalty (-1). In contrast, the majority class receives lower-magnitude rewards or penalties ($\pm \lambda$), where λ is a positive constant of less than one. This asymmetric weighting helps the model prioritize the minority class. It does so by assigning higher penalties for misclassifying lung pixels and greater rewards for correctly identifying them. This approach reduces the overwhelming influence of non-lung pixels during training.

To better illustrate how the reward mechanism addresses class imbalance, consider the flowchart in Fig. 4. At each time step, the agent observes a pixel (s_t) and predicts a class label (a_t). If the prediction is correct and the pixel belongs to the minority class (D_L), a strong positive reward (+1) is issued. Conversely, if misclassified, a strong penalty (-1) is applied. For the majority class pixels (D_N), correct predictions yield a small reward ($+\lambda$), while incorrect ones result in a small penalty ($-\lambda$). This reward design helps the model learn more accurate boundaries for lung pixels. It does this by assigning stronger penalties for mistakes in the minority class than in the majority class, as shown in the flowchart.

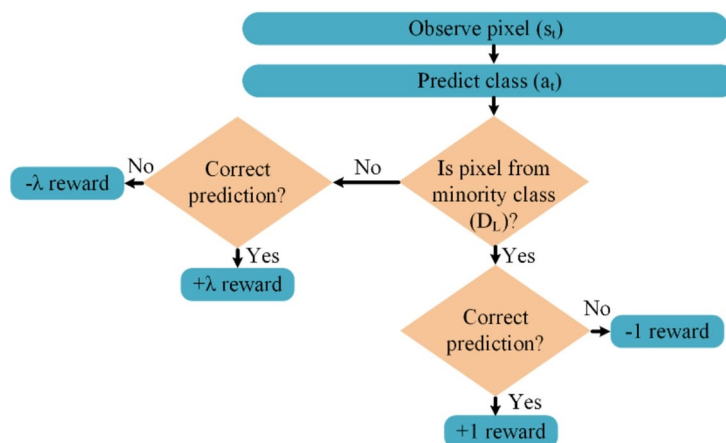


Fig. 4. Flowchart illustrating the reward assignment strategy in the off-policy PPO algorithm for lung segmentation.

Algorithm 1 outlines the training process of the proposed generator for lung segmentation. This approach integrates both off-policy PPO and adversarial learning to address the severe class imbalance inherent in segmentation tasks. At each iteration, the generator is treated as a policy network that predicts whether each pixel belongs to the lung or non-lung class. A custom-designed reward function, detailed in Eq. 18, assigns asymmetric incentives to correctly or incorrectly classified pixels, prioritizing the underrepresented lung class. In parallel, the generator receives feedback from a discriminator network, which evaluates the anatomical plausibility of the generated segmentation masks. To incorporate both learning signals, the generator is updated using a composite loss function: the clipped PPO loss, which enhances stability in off-policy updates, and an adversarial loss that ensures structural realism in the outputs. These two loss components play complementary roles: while the PPO loss provides fine-grained, pixel-level guidance under class imbalance, the adversarial loss promotes structural plausibility at the global level. This synergy enables the generator to enhance both voxel-level precision and holistic structural realism, which is particularly crucial in medical segmentation tasks. Unlike traditional segmentation methods that rely solely on cross-entropy or Dice-based losses, our method introduces a reinforcement learning signal tailored to the minority class and couples it with a GAN-based discriminator. This novel combination yields a more robust segmentation of lung areas, particularly in challenging conditions with severe imbalances and anatomical variability.

Input:

- Training dataset $D = \{I_1, I_2, \dots, I_N\}$, where each I_i contains T pixels as $\{(s_1, y_1), (s_2, y_2), \dots, (s_T, y_T)\}$ with s_t as the t -th pixel state and $y_t \in \{\text{lung, non-lung}\}$ as the target label
- Learning rate α
- Maximum number of epochs E
- Weighting factor for adversarial loss term γ

Initialize:

- Policy network parameters θ (generator)
- Old policy parameters $\theta_{old} \leftarrow \theta$
- Value function parameters ϕ
- Replay buffer $R \leftarrow \emptyset$

For epoch $e = 1$ to E :

Shuffle D

For each image $I_i \in D$:

Initial state s_1

For $t = 1$ to T :

Observe state s_t

Sample action $a_t \sim \pi_\theta(a_t | s_t)$

Compute reward r_t using Equation 18

Transition to next state s_{t+1}

Save the tuple $(s_t, a_t, r_t, s_{t+1}, \mu(\cdot | s_k))$ to R

Sample minibatch of transitions $(s_k, a_k, r_k, s_{k+1}, \mu(\cdot | s_k))$ from R

Estimate advantage $A(s_k, a_k)$ using value function parameters ϕ

Update policy θ by maximizing the clipped surrogate objective:

$$\theta \leftarrow \theta_{\text{prev}} + \alpha \nabla_\theta (L_{\text{Off-Policy PPO}}^{\text{CLIP}}(\pi) + \gamma \times \text{Loss}_{\text{adversarial}}(G, D))$$

Update $\theta_{old} \leftarrow \theta$

Algorithm 1. Pseudocode for training of the generator in lung segmentation.

Off-policy PPO This research presents an advanced version of PPO known as off-policy PPO, which optimizes data utilization by integrating information from previous experiences and decisions. Diverging from the conventional on-policy PPO that depends solely on immediate environmental data for effective operation, off-policy PPO incorporates a broader, more diverse dataset. This dataset comprises both prior interactions and assorted tactical decisions. Consider, for instance, a robot navigating through a complex maze. In contrast, standard on-policy PPO leverages only current observations to refine its approach; off-policy PPO utilizes insights from prior navigational attempts and evaluates different strategies across varying scenarios. This method accelerates learning by minimizing the need for the robot to exhaustively explore each potential route, thereby improving efficiency by exploiting extensive historical data.

Off-policy PPO addresses the optimization challenge by enhancing a surrogate objective function through the utilization of off-policy data, mirroring the technique used in off-policy TRPO⁹⁷:

$$\max_{\pi} E_{s \sim \rho_{\mu}, a \in \mu} \left[\frac{\pi(a|s)}{\mu(a|s)} A_{\pi_{old}}(s, a) \right] \quad (19)$$

subject to:

$$\bar{D}_{KL}^{\rho_{\mu}, s^{qrt}}(\mu, \pi_{old}) \bar{D}_{KL}^{\rho_{\mu}, s^{qrt}}(\pi_{old}, \pi) + \bar{D}_{KL}^{\rho_{\mu}}(\pi_{old}, \pi) \leq \delta \quad (20)$$

where

$$\rho_{\mu}(s) = \sum_{t=0}^{\infty} \gamma^t P(s_t = s | s_0, \mu) \quad (21)$$

$$\bar{D}_{KL}^{\rho_{\mu}}(\pi_{old}, \pi) = E_{s \sim \rho_{\mu}} [D_{KL}(\pi_{old}(\cdot|s) || \pi(\cdot|s))] \quad (22)$$

$$\bar{D}_{KL}^{\rho_{\mu}, sqrt}(\mu, \pi_{old}) = E_{s \sim \rho_{\mu}} \left[\sqrt{D_{KL}(\mu(\cdot|s) || \pi_{old}(\cdot|s))} \right] \quad (23)$$

$$\bar{D}_{KL}^{\rho_{\mu}, sqrt}(\pi_{old}, \pi) = E_{s \sim \rho_{\mu}} \left[\sqrt{D_{KL}(\pi_{old}(\cdot|s) || \pi(\cdot|s))} \right] \quad (24)$$

Here, μ denotes the behavior policy. Absent the constraints imposed by Eq. 20, striving to enhance the surrogate function using off-policy data (outlined in Eq. 19) could result in extensive policy modifications. To curtail this hazard, the PPO clipping mechanism is utilized to aptly refine the surrogate objective⁹⁷:

$$L_{\mu}(\pi) = E_{s \sim \rho_{\mu}, a \in \mu} \left[\frac{\pi(a|s)}{\mu(a|s)} A_{\pi_{old}}(s, a) \right] \quad (25)$$

Using $L_{\mu}(\pi)$, we establish the clipped surrogate objective for off-policy data as delineated below⁹⁷:

$$\bar{L}_{\mu}(\pi) = E_{s \sim \rho_{\mu}, a \in \mu} \left[\min \left(\frac{\pi(a|s)}{\mu(a|s)} A_{\pi_{old}}(s, a), \text{clip} \left(\frac{\pi(a|s)}{\mu(a|s)}, 1 - \epsilon, 1 + \epsilon \right) A_{\pi_{old}}(s, a) \right) \right] \quad (26)$$

The proportion $\frac{\pi(a|s)}{\mu(a|s)}$ often surpasses the bounds of $1 - \epsilon$ and $1 + \epsilon$, which typically leads to the policy $\pi(a|s)$ remaining unchanged while optimizing the clipped surrogate objective. To counter this stagnation, the limits of the clipped objective, $((1 - \epsilon), (1 + \epsilon))$, are modified in Eq. 26 by incorporating a correction factor $\frac{\pi_{\theta_i}(a|s)}{\mu(a|s)}$ ⁹⁷:

$$L_{Off-Policy PPO}^{CLIP}(\pi) = E_{s \sim \rho_{\mu}, a \in \mu} \left[\min \left[\frac{\pi(a|s)}{\mu(a|s)} A_{\pi_{old}}(s, a), \text{clip} \left(\frac{\pi(a|s)}{\mu(a|s)}, \frac{\pi_{old}(a|s)}{\mu(a|s)} (1 - \epsilon), \frac{\pi_{old}(a|s)}{\mu(a|s)} (1 + \epsilon) \right) A_{\pi_{old}}(s, a) \right] \right] \quad (27)$$

Figure 5 presents a streamlined seven-stage framework for optimizing the lung segmentation model using the off-policy PPO algorithm. The process begins by defining the optimization objective within a U-Net environment. A simplified surrogate objective is used along with a clipping mechanism to prevent drastic changes in the policy.

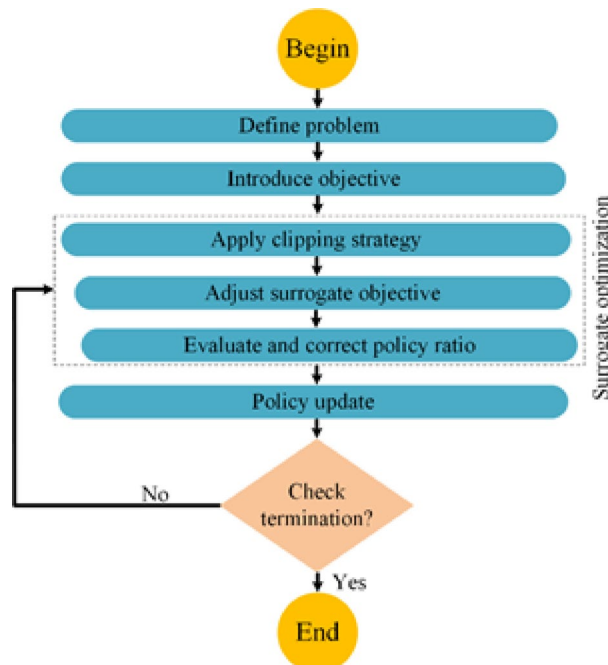


Fig. 5. Policy optimization for lung segmentation using off-policy PPO in the U-Net framework.

The objective is then modified to incorporate previously collected data, enabling the model to generalize more effectively. The clipped policy ratio is evaluated and corrected if it exceeds the allowed range, ensuring stable and gradual learning. The final policy is updated iteratively until convergence is reached. Key steps are highlighted to clarify logical progression and reduce visual complexity.

Discriminator

The discriminator in the lung segmentation stage is designed to distinguish between real segmented masks and those generated by the GAN-based segmentation network. It receives binary segmentation masks as input. Then, it applies several dilated convolutional layers to extract multi-scale contextual features. These features are crucial for determining whether the segmentation is anatomically accurate. Instead of comparing pixels directly, the architecture allows the discriminator to evaluate whether the entire structure and shape are consistent with real anatomical forms. The output is a spatial probability map that indicates the realism of each region in the input mask. This map helps localize inconsistencies in shape or structure that the generator needs to correct. During training, the EMD is used as an additional evaluation metric. EMD measures the difference between the real and synthetic masks across spatial regions. This goes beyond simple pixel-level errors. It encourages the generator to learn the spatial arrangement and anatomical shape of the lungs more precisely. Together, these mechanisms refine the generator output toward more realistic, anatomically aligned segmentation masks.

In duties such as image synthesis with GANs, the objective is usually to produce images that are indistinguishable from genuine ones. Nevertheless, lung segmentation offers distinct challenges. Authentic masks are binary, consisting only of 0s and 1s, but the masks produced by the generator vary continuously from 0 to 1. This difference may lead the discriminator to overly focus on identifying real versus produced masks, overlooking other vital factors.

To address the challenge, we have adopted a sophisticated approach. Instead of directly feeding the fake or genuine masks within the discriminator, we incorporate lung slice images into the input. These images are modified versions of the CT scans, adjusted with segmented and real masks. This method enables the discriminator to assess the connection between the produced masks and the lung images, thereby enhancing its capacity to accurately differentiate between genuine and produced masks. Further, a specialized segmentation mask that selectively emphasizes nodule areas while setting other areas to zero is used. This extra detail helps the discriminator better recognize essential features, improving its differentiation capability. The loss function for the discriminator is defined in the following equation:

$$E_{x \sim p_z} [D(G(x) \circ I_i)] - E_{x \sim p_{real}} [D(x)] \quad (28)$$

In this context, p_z and p_{real} show represent the distributions of the synthetic and real data, respectively. The symbol $\circ$ denotes an operator that carries out pixel-wise multiplication. This operation merges the intensity data from the CT images with the spatial data from the masks. It compels the discriminator to evaluate the consistency of the fused information instead of assessing the components separately.

Tumor detection

Figure 6 illustrates the structure of the Mask R-CNN-based generator network for tumor detection. In this paper, our model adopts a more targeted strategy than the traditional GAN approach of generating synthetic outputs from random noise. It employs lung data from a previous step as input, sidestepping the randomness typical of conventional GANs. This deliberate input selection empowers the generator to fully exploit the comprehensive details in the lung data, producing accurate and appropriate outputs. The primary aim of the generator is to generate outputs that offer classification and precise location information. During the analysis of lung imagery,

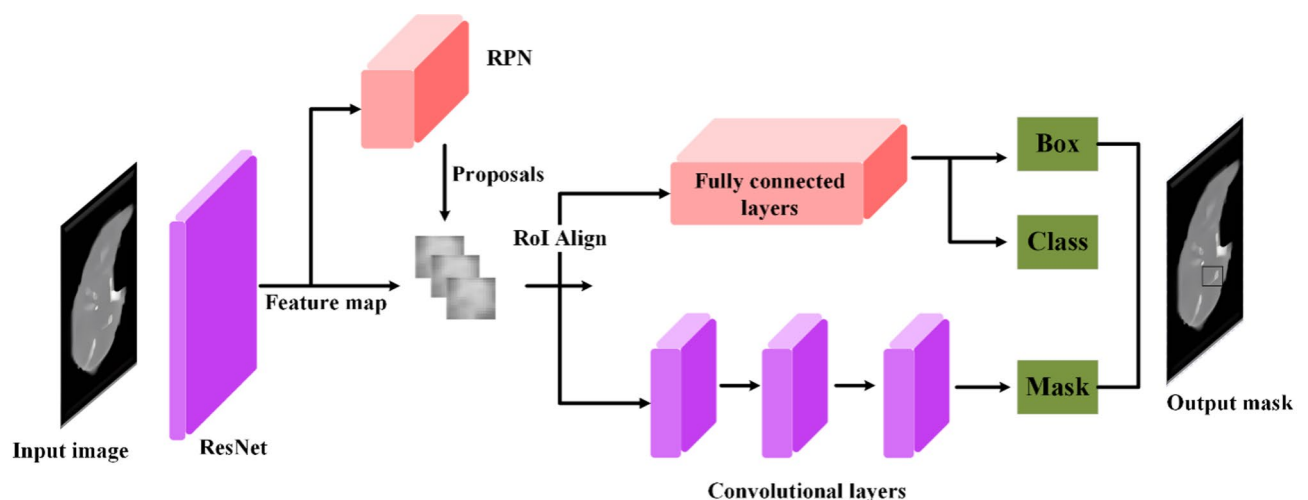


Fig. 6. Architecture of the Mask R-CNN-based generator network for tumor detection.

the generator categorizes various areas within the lungs, accurately labels them, and calculates their boundaries using the coordinates of the bounding box.

We employ a novel loss function, as described in Eq. 29, to guide the training of the generator. This function combines classification loss, bounding box regression loss, and an adversarial term.

$$L_G = L_{cls} + L_{box} + L_{adv}^{G_b} \quad (29)$$

where L_{cls} and L_{box} represent the standard classification and bounding box regression losses as used in the Mask R-CNN method. These ensure accurate class labeling and box placement. However, many existing models rely solely on these two losses, limiting their ability to capture realistic tumor structures in ambiguous or irregular regions. The adversarial loss $L_{adv}^{G_b}$ in Eq. 30 addresses this gap:

$$L_{adv}^{G_b} = \frac{1}{N} \sum_{i=1}^N -\log D_b(G_b(RoI_i)) \quad (30)$$

where N indicates the mini-batch size, and $G_b(RoI_i)$ refers to the prediction of the i -th bounding box. This adversarial component drives the model to generate bounding boxes that reflect the high-level semantic structure of tumor regions, leading to more robust predictions in cases with low contrast or morphological ambiguity. The integration of classification and regression losses ensures core detection accuracy, while the adversarial term refines structural plausibility. Together, these components guide the model to locate tumors and outline their shapes more accurately, even when the intensity information is unclear or varies across regions.

The loss function of the discriminator, detailed in Eq. 31, employs CNNs to enhance its capacity to distinguish between genuine and artificial bounding boxes, thus improving the effectiveness of the training process.

$$L_D = \frac{1}{N} \sum_{i=1}^N -[\log(D_b(bb_i^{gt})) + \log(1 - D_b(G_b(RoI_i)))] \quad (31)$$

where D_b reflects the probability that the image is genuine. The adversarial loss $L_{adv}^{G_b}$ motivates G_b to generate bounding boxes convincing enough to deceive D_b . This loss term is designed to encourage the generator to produce increasingly convincing boxes, making them indistinguishable from real ones and thus narrowing the gap between synthetic and authentic distributions.

Figure 7 depicts the optimization flowchart for the generator used in tumor detection within the Mask R-CNN framework. The process starts by inputting segmented lung data essential for training the generator to accurately recognize and classify features in lung images. The Mask R-CNN then processed this data. In this system, the generator identifies the categories of different regions (either tumor or non-tumor) and computes bounding box coordinates that accurately pinpoint tumors in the lung scans. During training, two critical loss components are calculated: the class loss (L_{cls}) evaluates the performance of tumor classifications, and the bounding box loss (L_{box}) assesses the quality of the bounding boxes surrounding the detected tumors. These metrics provide essential feedback for refining the performance of the generator. Additionally, the discriminator calculates the adversarial loss (L_{adv}) to determine how closely the synthetic images mimic actual lung scans. This step is vital as it drives the generator to produce images that resemble real medical scans, thus enhancing the clinical applicability of the model. The optimization includes a feedback loop, continuously updating the

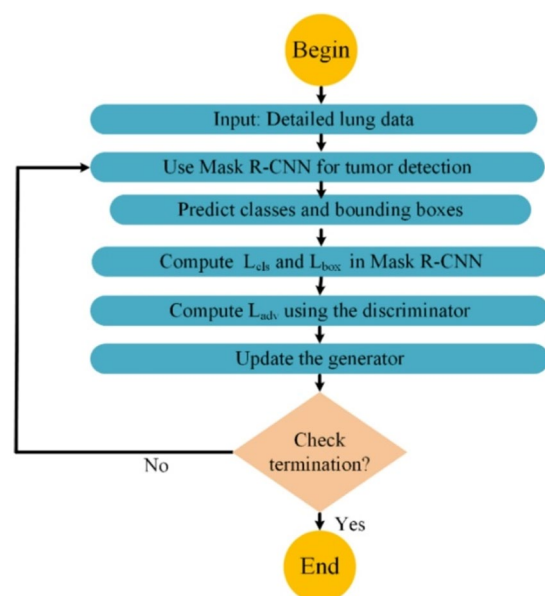


Fig. 7. Generator training for tumor detection using Mask R-CNN with adversarial loss.

generator based on these losses to enhance quality and realism. This prediction, loss evaluation, and updates cycle persists until the model achieves predefined termination criteria, such as reaching a maximum number of iterations. This methodical process ensures that each step boosts the efficacy of the generator, resulting in a powerful model adept at detecting tumors in lung CT images.

The discriminator in the tumor detection stage is designed to evaluate the quality of bounding boxes and classifications produced by the generator. It takes as input the feature maps of predicted tumor regions, especially the RoIs) generated by the Mask R-CNN head. Instead of relying only on coordinate accuracy, the discriminator checks whether the predicted regions appear realistic based on semantic and anatomical information. It applies dilated convolutional layers to capture contextual features at different scales. This helps the discriminator assess the overall structure and coherence of the predicted tumor regions. The output is a scalar probability that reflects how closely a predicted region matches real tumors. During training, the EMD is employed as an additional loss function to quantify the differences between real and generated tumor distributions. In contrast to the segmentation stage, where the discriminator evaluates entire binary masks, this stage operates at the RoI level. It compares the semantic features of predicted bounding boxes with those of real tumor annotations. This setup enables the generator to improve both the accuracy of tumor localization and the realism of the predicted structures. The combined use of EMD and dilated convolutions ensures that feedback from the discriminator guides the generator toward more anatomically faithful tumor detection.

3D reconstruction method

As shown in Fig. 3, the 3D reconstruction stage processes N sequences of output images from the tumor detection stage. These images undergo feature extraction through the pre-trained EfficientNet network, which is known for its effectiveness. The architecture of EfficientNet refined through learning from the vast data, ensures high performance in recognizing and categorizing image content. Its layers capture a wide range of features, from simple edges in the early layers to complex object parts in deeper layers, making it highly suitable for medical imaging where fidelity is critical. The extracted features are then fed into TLSTM units, crucial for enhancing temporal continuity in the data. By integrating context from previous frames, TLSTM units greatly enhance the interpretation of the current frame.

A spatial attention mechanism is placed before the TLSTM units to improve their ability to capture key temporal features. This mechanism enables the model to focus selectively on the most informative parts of each input image rather than treating all spatial regions equally. The attention layer is trained to prioritize features that are essential for 3D reconstruction. These often include tumor boundaries or prominent anatomical structures. By emphasizing these critical areas, the model reduces the influence of background noise and less significant regions, which often distract conventional sequence models. This focused representation enables TLSTM to interpret each frame within the context of both previous frames and spatial relevance, resulting in more coherent and anatomically accurate 3D reconstructions.

Figure 8 illustrates the structure of the generator network, which directly receives the output from the TLSTM module. The TLSTM extracts features that remain consistent over time from the image sequence and incorporates spatial attention to emphasize important regions in each frame. These features are reshaped into a compact 3D tensor that retains anatomical continuity across slices. This 3D tensor serves as the starting input to the generator, forming the basis for volumetric reconstruction. The generator processes this input, which contains both temporal and spatial information, to produce outputs with greater accuracy and more consistent anatomical structure.

The architecture of the generator features six deconvolution layers, each equipped with batch normalization (BN) and the ReLU activation function. In the GAN setup, the discriminator plays a crucial role. It accurately differentiates synthetic and real-world images from the generator, ensuring the 3D images are realistic. This effectiveness is vital for correct medical diagnosis and planning treatments. By combining the powerful feature extraction of EfficientNet with advanced sequence processing and generative modeling, this method successfully creates precise 3D representations from 2D image sequences, greatly improving the quality and realism of the generated images.

The discriminator in the 3D reconstruction stage is designed to evaluate the realism of volumetric images generated by the GAN-based architecture. It receives as input the reconstructed 3D lung tumor volumes produced

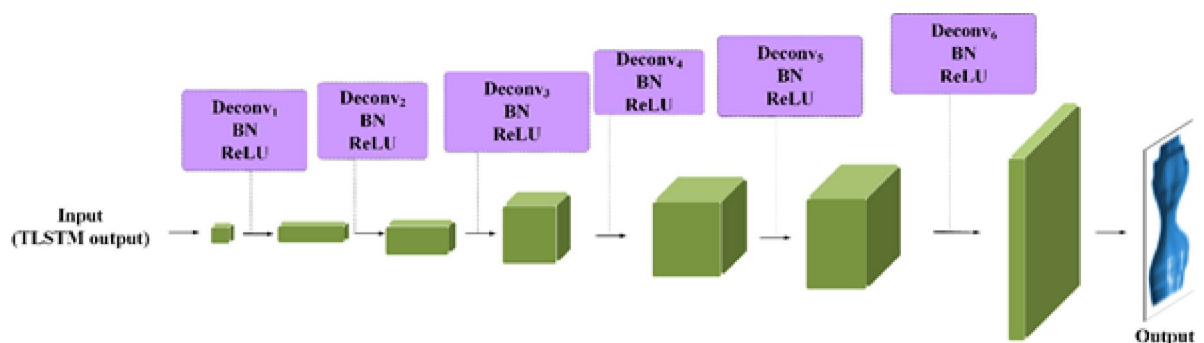


Fig. 8. Architecture of the generator network for the 3D reconstruction of lung tumors.

by the generator. The discriminator compares these generated volumes with the corresponding real, annotated 3D ground-truth data to evaluate the structural authenticity. To evaluate these 3D volumes, the discriminator employs 3D dilated convolutional layers. This structure enables it to analyze spatial dependencies in all three dimensions and detect inconsistencies in the generated structures. The output is a scalar prediction score that reflects the anatomical plausibility of the reconstructed volume. During training, EMD is employed as a global evaluation criterion, quantifying the distributional shift between real and synthetic 3D data across spatial regions. In contrast to earlier stages, where the discriminator evaluates 2D binary masks or region proposals, it operates on entire 3D reconstructions here. The discriminator acts not only as a classifier but also as a spatial coherence assessor, ensuring the generated volumes exhibit consistent shape, texture, and continuity over slices. This feedback loop encourages the generator to refine outputs that are both anatomically consistent and visually coherent in the volumetric space.

Algorithm 2 shows the training procedure for the proposed TLSTM-based 3D tumor reconstruction model. The process begins with a training dataset composed of sequences of 2D tumor-detected slices, each annotated with a corresponding 3D ground-truth volume. Each slice from the input sequence is processed by a pre-trained EfficientNet encoder. This step extracts important high-level features from each image. A spatial attention mechanism is then applied to enhance the focus on anatomically relevant regions such as tumor boundaries. These focused features are sent to a TLSTM module. This module updates its hidden and cell states at each step to preserve temporal consistency across the slices.

Input:

- Training dataset $D = \{(x_1, y_1), (x_2, y_2), \dots, (x_T, y_T)\}$, where each x_i is a sequence of N 2D tumor-detected slices as $\{I_1, I_2, \dots, I_N\}$ and y_i is the ground-truth 3D volume.

- Pre-trained EfficientNet encoder $E(\cdot)$

- Learning rate α

- Maximum number of epochs E

- Weighting factor for the adversarial loss term ∂

Initialize:

- TLSTM parameters θ_{TLSTM}

- TLSTM hidden state h_0 and cell state c_0

- Spatial attention parameters W_{att}, b_{att}

- Generator parameters θ_G

For epoch $e = 1$ to E :

 Shuffle D

 For each sequence $x_i \in D$:

 For $t = 1$ to N

$X_t \leftarrow E(I_t)$ #Feature map from EfficientNet

$a_t = \text{softmax}(X_t \times W_{att} + b_{att})$

$X'_t = X_t a_t$

$c_{t,\eta} = f(c_{t-1,\eta}, h_{t-1,\eta}, X'_t; w_{lstm,\eta}, b_{lstm,\eta})$

$h_{t,\eta} = g(h_{t-1,\eta}, c_{t-1,\eta}, X'_t; w_{lstm,\eta}, b_{lstm,\eta})$

 Store $h_{t,\eta}$ in memory buffer

$H_{seq} \leftarrow \{h_{1,\eta}, h_{2,\eta}, \dots, h_{N,\eta}\}$

$\hat{y}_i \leftarrow G(H_{seq})$ #Reconstruct 3D volume

 Compute loss $l \leftarrow \text{MAE}(\hat{y}_i, y_i) + \partial \times \text{adversarial loss}(\hat{y}_i)$

$\theta_G \leftarrow \theta_G + \alpha \nabla_{\theta_G} l$

$\theta_{TLSTM} \leftarrow \theta_{TLSTM} + \alpha \nabla_{\theta_{TLSTM}} l$

Algorithm 2. Pseudocode for training of TLSTM-based 3D tumor reconstruction.

The hidden state sequence contains both spatial and temporal information. This sequence is passed to a generator network that reconstructs the full 3D tumor volume. The training loss comprises two components: the mean absolute error (MAE) between the predicted and actual volumes and an adversarial loss. The adversarial loss is scaled using a hyperparameter. Gradients are backpropagated to update the parameters of the generator and the TLSTM units. This process gradually improves the quality of 3D reconstructions throughout training epochs.

Overall algorithm

Figure 9 provides a comprehensive overview of the 3D reconstruction of lung tumors, outlining the step-by-step process from initial CT image acquisition to the final evaluation of the reconstructed 3D model. The flowchart starts with the input of a series of CT images from a patient, labeled as $\{I_i, M\}$, where M is the target 3D image derived from the sequence of I_i images. This target image is crucial as it serves as the benchmark for assessing the performance of the reconstructed 3D model. The rest of the stages are shown in the following:

- Preprocessing stage: Initially, all CT images I_i undergo extensive preprocessing to standardize the data and enhance image quality for more effective subsequent analysis. This includes resizing images to ensure consistent resolution across all inputs, which helps to standardize the data for better processing in later stages. Further preprocessing include normalization to adjust intensity scales and augmentation to increase the dataset

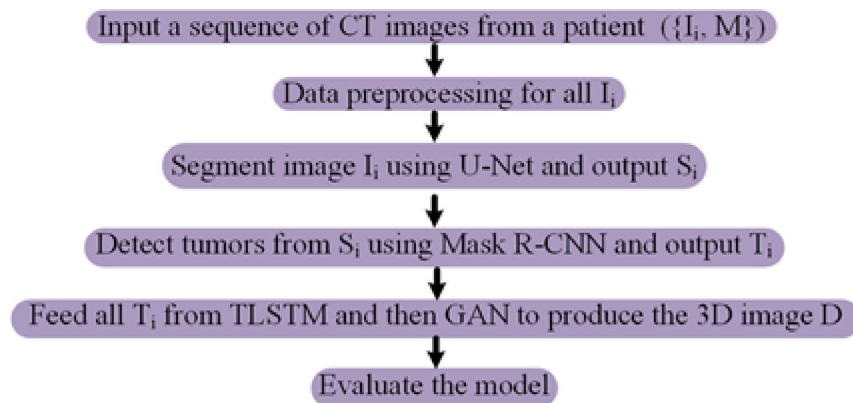


Fig. 9. Workflow for the 3D reconstruction of lung tumors.

variably and robustly by creating modified versions of existing images. These preprocessing efforts are vital for optimizing the performance of later stages, such as segmentation and detection.

- Segmentation process: Following preprocessing, each image I_i is segmented using U-Net architecture. This neural network is particularly effective for medical image segmentation, producing an output S_i delineating lung regions. These segmented images isolate the lung fields, enhancing clarity and focus for tumor detection.
- Tumor detection: The segmented images S_i then undergo further analysis using Mask R-CNN, an algorithm tailored for precise object detection and instance segmentation. This step accurately identifies and localizes tumors within the lung segments, generating T_i outputs that include detailed bounding boxes around each detected tumor.
- 3D reconstruction: After detecting the tumors, the relevant data T_i is processed through a TLSTM and GAN system. The TLSTM network manages temporal relationships within the sequence data, refining the feature extraction, which is crucial for the subsequent synthesis by the GAN. The GAN then compiles these features to form a cohesive 3D image D , representing a detailed reconstruction of the lung tumors based on the 2D inputs.
- Model evaluation: The final evaluation stage compares the generated 3D image D against the target 3D image M to verify the performance of the model. This comparison is essential for confirming the clinical utility of the model by assessing the fidelity of the tumor structures in the 3D reconstruction. Ensuring the performance of these representations is critical for their potential application in diagnostic and treatment planning contexts.

Empirical evaluation

In the following section, we will outline the dataset, describing its features and the scope of our study. This explanation will be followed by a discussion on the metrics used, where we will define the criteria and measurements employed to evaluate the performance of our models. The section will conclude with the presentation of results, spotlighting the primary outcomes from our analysis and model evaluations. We will also discuss the importance of these findings in our research objectives, offering insights into their implications and relevance to the field.

Dataset

This article uses the LIDC-IDRI dataset² to evaluate the proposed model, chosen for its relevance and applicability in real-world clinical scenarios. The LIDC-IDRI dataset is widely recognized in the medical imaging community for its comprehensive collection of annotated thoracic CT scans, which include expert-reviewed radiological annotations. This makes it an ideal choice for validating advanced imaging models like ours. Employing this well-established dataset guarantees robust evaluation metrics and benchmarks the performance of the model against a realistic standard. Furthermore, by matching the model with real-world clinical data and challenges, the dataset aids in evaluating the practical effectiveness and adaptability of the model to actual medical settings. This alignment helps demonstrate the potential of the model for deployment in healthcare systems, enhancing its credibility and utility in clinical settings.

The LIDC-IDRI, meticulously developed by the Foundation for the National Institutes of Health (FNIH) and the Food and Drug Administration (FDA), comprises 1,018 CT scans from 1,010 registered patients, accompanied by an extensible markup language (XML) file detailing annotations by four radiologists. The primary aim of this rigorous annotation process, conducted in two stages, is to accurately identify and document all nodules within each CT scan. The initial stage, blinded-read, involves every radiologist reviewing the scans to identify lesions. These are categorized as either “nodule < 3 mm” or “non-nodule ≥ 3 mm”. In the unblinded-read stage, radiologists revisit their annotations with access to the anonymous markings of their colleagues, fostering a collaborative environment for cross-verification and adjustments if necessary. Within this dataset, 7,371 lesions have been identified as nodules by at least one radiologist. Of these, 2,669 nodules are labeled consistently as ‘nodule ≥ 3 mm’ by all four radiologists, with 928 receiving unanimous agreement. These 2,669

nodules are further detailed with ratings based on their characteristics, such as size, shape, and density, and outlined visually in the dataset, enhancing their utility for diagnostic and research purposes.

It should be noted that the LIDC-IDRI dataset originally contained annotated 2D CT slices rather than native 3D volumes. To overcome this limitation, we developed a reproducible pipeline to generate consistent 3D volumes. This pipeline was implemented using Rhinoceros 3D (version 7)⁹⁸, a widely adopted tool for 3D modeling and visualization. First, each annotated sequence of 2D slices corresponding to a single nodule case was imported into Rhinoceros using the Grasshopper plugin. A custom Python script was used to automatically extract nodule contours from each slice. These contours were then converted into closed polylines. The contours were stacked along the Z-axis based on the original CT slice thickness (typically 1.25 mm). A lofting operation was then applied to interpolate surfaces between the stacked contours. This process produced a smooth 3D mesh of the lesion. To enhance anatomical realism, the mesh was smoothed using the “Smooth Mesh” function with a setting level of 2 while preserving control points. The mesh resolution was set to medium, and a closed solid was generated for accurate volume rendering. Finally, all models were exported in standard .STL and .OBJ formats for downstream analysis. These settings ensure consistency and reproducibility in the 3D reconstruction process. The reconstructed models enrich the dataset by adding realistic 3D structures. They also serve as reliable ground truth for evaluating the performance of the proposed TLSTM-based volumetric reconstruction method.

To ensure consistency and improve model performance, several preprocessing steps were applied to the raw CT images before feeding them into the segmentation, detection, and reconstruction pipelines. First, slice alignment was applied to preserve spatial coherence across the image sequences. Slices were aligned along the same anatomical axis using DICOM metadata to prevent positional drift. Next, z-score normalization was used to standardize the intensity values in each image. This process ensured zero mean and unit variance, helping to reduce variability caused by different scanners or individual patient characteristics. To increase robustness and prevent overfitting, various data augmentation methods were applied during training. These included random rotations ($\pm 15^\circ$), horizontal and vertical flipping, and elastic deformations to simulate realistic anatomical variations. All preprocessing steps were implemented in Python using the SimpleITK and Albumentations libraries. These operations were applied uniformly across the training and validation subsets, significantly contributing to the reproducibility and generalizability of the model across different data distributions.

We separated the data for all models into training, validation, and testing subsets to streamline model training and evaluation. For lung segmentation and tumor detection using the LIDC-IDRI dataset, which includes 244,527 images, we distributed 70% of these images to the training subset, 15% to the validation subset, and 15% to the test subset. Specifically, this allocation provided approximately 171,169 images for training, 36,679 for validation, and 36,679 for testing. In the case of the 3D reconstruction model, this 70-15-15 distribution resulted in 5,785 images for training, 1,240 for validation, and 1,240 for testing. We chose this particular 70-15-15 split to ensure a balance between having ample data for training to enhance model learning and sufficient data for validation and testing to confirm thorough evaluation. By allocating 70% of the data to training, we ensure that the model learns from diverse examples, which helps it detect complex patterns and improve its generalization. The 15% of data set aside for validation is sufficient to adjust hyperparameters and prevent overfitting, enhancing the performance of the model on new data. Finally, the 15% allocated to testing offers an unbiased evaluation of the performance of the model, providing reliable insights into its effectiveness in real-world applications.

Metrics

In evaluating lung segmentation and tumor detection models, intersection over union (IoU) and HD are used because these metrics help gauge the quality of the model and the boundaries it predicts. IoU is particularly critical for lung segmentation and tumor detection as it provides a clear measure of overlap consistency between the predictions of the model and the ground truth data. By calculating the intersection ratio to the union of the predicted and actual segments, IoU directly indicates how well the model can identify relevant areas within the lungs and tumors, regardless of their size or shape. This makes it invaluable in clinical settings where even small inaccuracies in segmentation can lead to significant differences in diagnosis or treatment outcomes. On the other hand, HD is essential for assessing the maximum distance between the predicted and true boundaries. It is vital as it confirms the exactness of the method in outlining the outer edges of the important regions. HD is beneficial when high exactness is required in capturing the boundaries of lung tissues and tumors, as slight deviations in these boundaries can impact clinical decisions like surgical margins or radiation targeting. Together, IoU and HD provide a comprehensive evaluation framework. They not only measure how much of the target the model successfully captures (IoU) but also how accurately it traces the most extreme perimeters of the targets (HD). This dual assessment helps ensure that the models are robust and effective for practical application in medical diagnostics and treatment planning, where overall robustness is critical.

In evaluating the 3D reconstruction model, HD and ED are chosen for their ability to effectively measure spatial fit and alignment errors. HD is particularly valuable as it measures the largest distance between the nearest points on the reconstructed surface and the actual model. It provides a worst-case scenario essential for high-performance applications like medical imaging and surgical planning. ED offers a simple geometric measure of the average distance between corresponding points on the actual and reconstructed models, helping assess the overall error throughout the volume. This combination ensures a robust evaluation, capturing both outlier errors and overall fidelity, which is essential for verifying the clinical reliability of the reconstructed 3D models. These metrics, together, provide a comprehensive view of the performance of the model in replicating intricate anatomical structures, which is vital for ensuring the realism required in medical applications.

The definitions for the metrics IoU, HD, and ED are provided as such:

- IoU is measured by the intersection ratio to the union of the pixel sets A and B , where A represents the pixels within the actual mask, and B denotes the pixels within the predicted mask. These sets consist of non-zero pixels indicating the segmented areas.

$$IoU = \frac{|A \cap B|}{|A \cup B|} \quad (32)$$

- HD quantifies the maximum distance from any point within one set to the nearest point in the alternative set. In this metric, $\|a - b\|$ represents the ED between points a and b , which are points from sets A and B , respectively.

$$HD(A, B) = \max(h(A, B), h(B, A)) \quad (33)$$

with

$$h(A, B) = \max_{a \in A} \min_{b \in B} \|a - b\| \quad (34)$$

- ED measures the straight-line distance between corresponding points in 3D space, denoted as (x_1, y_1, z_1) and (x_2, y_2, z_2) . This metric provides a scalar value that reflects the positional alignment of each corresponding point pair in the reconstructed 3D models.

$$ED = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2} \quad (35)$$

Main results

To ensure the effectiveness of our model, we implemented stratified cross-validation in our hyperparameter optimization process. Stratified cross-validation guarantees that each subset of the data used in the validation process represents the overall dataset accurately by mirroring the proportion of each class found in the full dataset. This approach is vital in medical imaging, where the prevalence and characteristics of target features, such as tumors, can significantly differ. Using stratified cross-validation, we prevent our models from overfitting to less representative data and ensure that our performance assessments are consistent and reliable across varying types of tumor presentations. This reliability is crucial for clinical settings, where accurate and consistent model performance can directly impact diagnostic and treatment decisions. Table 5 shows the results of the optimization of hyperparameters.

During the evaluation of 3D reconstruction, the proposed model was subjected to a stringent comparison against seven established models, GAN-LSTM-3D¹⁰, YOLOv4⁸⁷, GAN-ResNet-3D⁸³, GAN-GK-LSTM⁸⁴, ViT⁸⁸, GAN-LSTM-RL², and GAN-SP-LSTM⁸⁹. We also conducted ablated experiments to investigate the effect of removing key components from our framework. These ablation settings include:

- Proposed w/o TLSTM: replaces TLSTM with vanilla LSTM.
- Proposed w/o SP: excludes the spatial attention module.
- Proposed with transformer: replaces TLSTM with a transformer-based sequence encoder.
- Proposed w/o adversarial loss: removes the adversarial loss component from the generator training.
- Proposed w/o MAE: excludes the mean absolute error term between the predicted and actual volumes from the generator loss.
- Proposed w/o DC: excludes the DC module in the discriminator.

The comparative results in Table 6 highlight notable variations in 3D reconstruction performance among existing models. GAN-LSTM-RL shows the best performance, with an ED of 1.274 and HD of 1.156. This improvement is mainly due to reinforcement learning, which improves feature extraction and temporal alignment. GAN-ResNet-3D performs the worst, with ED and HD values of 2.478 and 1.912. This result is likely caused by the absence of temporal modeling and the use of deep residual layers that do not effectively capture volumetric context. GAN-GK-LSTM also performs poorly in terms of HD (1.852) because it depends on the Gustafson-Kessel clustering technique, which struggles with spatial heterogeneity. YOLOv4 yields ED of 1.485 and HD of 1.256, while ViT achieves ED of 1.461 and HD of 1.348. Both models offer moderate results but do not include components for temporal coherence or fine-grained spatial focus. Although GAN-SP-LSTM benefits from spatial attention, its HD of 1.465 indicates it remains behind models with stronger temporal dynamics. In summary, most state-of-the-art models fall short due to weak temporal modeling, poor transductive learning, or ineffective attention mechanisms.

The proposed model achieves an ED of 0.985 and HD of 0.648, outperforming all baselines. Relative to GAN-LSTM-RL, the model improves ED by 22.7% and HD by 43.9%, illustrating superior reconstruction fidelity and boundary adherence. Compared to ViT, the model achieves a 32.6% improvement in ED and 51.9% in HD. This result validates the added value of transductive LSTM with spatial attention, which captures local and contextual variations more effectively. The proposed model also outperforms YOLOv4 by 33.6% in ED and 48.4% in HD. This difference highlights the importance of combining transductive LSTM and adversarial learning, which are not present in YOLOv4. Against the weakest competitor, GAN-ResNet-3D, the model improves ED by 60.2% and HD by 66.1%. This substantial gain results from the integration of three innovations in the model: PPO-

Hyperparameter description	Valid ranges	Best value
Long segmentation		
Rate of learning for generator	0.00001–0.01	0.002
Layers in generator	3–10	6
Rate of learning for discriminator	0.00001–0.01	0.001
Layers in discriminator	3–10	8
Size of batch	8–512	68
Momentum	0.6–0.9	0.85
Decay of weights	0.001–0.01	0.001
γ	0.1–1	0.68
Tumor detection		
Rate of learning for generator	0.00001–0.01	0.01
Layers in generator	3–10	8
Rate of learning for discriminator	0.00001–0.01	0.001
Layers in discriminator	3–10	7
Size of batch	8–512	82
Scales of region proposal network (RPN) anchor	8–1024	102
Ratios of RPN anchor	0.25–3	1.1
Ratio of ROI positive	0.1–0.6	0.35
Std for bounding box refinement	0.001–0.2	0.012
Threshold for masking	0.4–0.8	
3D reconstruction		
Rate of learning for generator	0.00001–0.01	0.001
Layers in generator	3–10	5
Rate of learning for discriminator	0.00001–0.01	0.0015
Layers in discriminator	3–10	9
Size of batch	8–512	102
Units in TLSTM	64–1024	247
Layers in TLSTM	2–8	4
Rate of Dropout in TLSTM	0.1–0.8	0.3
Beta1 for adaptive moment estimation (ADAM)	0.1–0.9	0.6
Beta2 for Adam	0.1–0.9	0.65
∂	0.1–1	0.56

Table 5. Optimal hyperparameter settings for the proposed method using cross-validation.

Model	HD	ED
GAN-LSTM-3D ¹⁰	1.415 ± 0.120	1.852 ± 0.174
YOLOv4 ⁸⁷	1.256 ± 0.142	1.485 ± 0.256
GAN-ResNet-3D ⁸³	1.912 ± 0.103	2.478 ± 0.103
GAN-GK-LSTM ⁸⁴	1.852 ± 0.262	1.952 ± 0.123
ViT ⁸⁸	1.348 ± 0.018	1.461 ± 0.074
GAN-LSTM-RL ²	1.156 ± 0.104	1.274 ± 0.142
GAN-SP-LSTM ⁸⁹	1.465 ± 0.136	1.546 ± 0.098
Proposed w/o TLSTM	1.154 ± 0.245	1.962 ± 0.195
Proposed w/o SP	1.026 ± 0.105	1.256 ± 0.116
Proposed with transformer	1.452 ± 0.213	1.635 ± 0.226
Proposed w/o adversarial loss	1.305 ± 0.045	1.426 ± 0.085
Proposed w/o MAE	0.984 ± 0.135	1.145 ± 0.162
Proposed w/o DC	0.912 ± 0.058	1.025 ± 0.105
Proposed	0.648 ± 0.024	0.985 ± 0.087

Table 6. Comparative analysis of various models on 3D reconstruction.

enhanced segmentation, adversarial tumor detection, and TLSTM-guided 3D reconstruction. Furthermore, compared to GAN-SP-LSTM (ED: 1.546, HD: 1.465), which also uses spatial attention, the model improves ED by 36.3% and HD by 55.7%, highlighting the crucial role of transductive learning and joint optimization across modules. This consistent superiority in both metrics across all comparisons confirms the robustness, generalizability, and precision of the proposed system.

The ablation study underscores the critical role of each component in the performance of the proposed model. Removing the TLSTM module resulted in significant degradation. ED and HD increased to 1.962 and 1.154, which correspond to performance drops of 99.2% and 77.9%, respectively. This indicates the vital role of TLSTM in temporal modeling. Eliminating the spatial attention module raised ED to 1.256 and HD to 1.026, reflecting drops of 27.5% and 58.3%, showing that spatial attention enhances focus on important regions. When TLSTM was replaced with a transformer, ED and HD increased to 1.635 and 1.452. These changes represent losses of 66% and 124%, indicating that transformers are less effective than TLSTM in capturing localized patterns. Removing the adversarial loss increased ED to 1.426 and HD to 1.305. These values indicate degradations of 44.8% and 101.2%, highlighting the role of the adversarial component in generating realistic tumor structures. Omitting the MAE loss led to ED of 1.145 and HD of 0.984, reducing pixel-level consistency by 16.2% and 51.8%. Finally, removing the dilated convolution (DC) module in the discriminator caused ED and HD to rise to 1.025 and 0.912. These results confirm that all components, TLSTM, spatial attention, adversarial loss, MAE, and DC, are essential for achieving optimal reconstruction accuracy.

We conducted paired t-tests with a two-tailed distribution and a 95% confidence level on results to determine if our model significantly outperforms existing ones. We calculated p-values for each comparison between our model and existing models using the HD and ED metrics. The analysis showed that the improvements of our model in HD and ED are statistically significant compared to all other models. For instance, when comparing the performance of the proposed model on HD (0.648) with that of GAN-LSTM-3D (1.415), the exact p-value was 0.007, indicating a highly significant improvement. Similar results were obtained against YOLOv4 and GAN-ResNet-3D, with p-values of 0.02 and 0.005, respectively, underscoring the substantial enhancements in tumor detection and 3D reconstruction quality. Moreover, the ED metric also showed significant improvements. The proposed model achieved an ED of 0.985, significantly better than the 1.461 of ViT, with a p-value of 0.015. This trend of statistical significance was consistent across comparisons with other models like GAN-GK-LSTM and the version of our model without TLSTM, where p-values were 0.04 and 0.03, respectively, confirming the superior capability of the proposed model in producing precise 3D reconstructions of lung tumors. In addition, the comparison with GAN-LSTM-RL yielded p-values of 0.011 for HD and 0.008 for ED, confirming the statistically significant advantage of our model. Against GAN-SP-LSTM, the p-values were 0.004 for HD and 0.006 for ED, further supporting the robustness of our framework. Statistical significance was consistently observed across all other comparisons as well, reinforcing the robustness of the performance of the model. These statistical tests and confidence intervals confirm that the performance improvements of our model are statistically significant across various datasets, highlighting its potential for clinical use in medical imaging.

Table 7 compares graphics processing unit (GPU) usage and runtime among various 3D reconstruction models, highlighting their computational efficiency. Our model uses only 10.1 gigabytes (GB) of GPU resources, 20% less than GAN-LSTM-3D, the next most efficient model, and far less than the 18.9 GB ViT requires, showing significant efficiency. Such efficient GPU usage makes our model attractive for environments with limited hardware resources. For runtime, our model also excels, completing tasks in 2,382 s. This is faster than GAN-LSTM-3D by about 10.7% and significantly quicker than GAN-ResNet-3D, which takes 3964 s—a reduction in time of nearly 40%. Such efficiency is crucial for real-time applications that require fast processing. This combination of lower GPU use and quicker processing enhances the practicality of the model in real-world scenarios, reducing operational costs and enabling the handling of larger datasets or multiple tasks simultaneously. These runtime and resource utilization improvements affirm the suitability of the proposed model for real-time medical imaging tasks, making it a valuable tool in clinical settings where time and performance are paramount.

Our next experiment evaluated several pre-trained models, including GoogleNet, ResNet, AlexNet, DenseNet, MobileNet, and VGG-16, as alternatives to the EfficientNet network for extracting features. The results, as presented in Table 8, show varying performance levels across the models tested. Although GoogleNet is robust, it fell behind with an ED of 2.145 and an HD of 1.420, indicating less precise boundary delineation than other models. ResNet and AlexNet showed improvements in both metrics, indicating a stronger capability for accurate feature extraction, with ResNet achieving an ED of 1.820 and AlexNet slightly better at 1.695.

Model	Runtime (s)	GPU usage (GB)
GAN-LSTM-3D ¹⁰	2671	10.9
YOLOv4 ⁸⁷	2485	13.7
GAN-ResNet-3D ⁸³	3964	14.4
GAN-GK-LSTM ⁸⁴	2392	16.2
ViT ⁸⁸	2686	18.9
GAN-LSTM-RL ²	2860	17.2
GAN-SP-LSTM ⁸⁹	2745	16.3
Proposed	2382	10.1

Table 7. Comparative analysis of runtime and GPU usage across various 3D reconstruction models.

Model	HD	ED
GoogleNet	1.420 ± 0.152	2.145 ± 0.124
ResNet	1.256 ± 0.136	1.820 ± 0.185
AlexNet	1.053 ± 0.128	1.695 ± 0.163
DenseNet	0.896 ± 0.140	1.741 ± 0.145
MobileNet	0.841 ± 0.174	1.652 ± 0.120
VGG-16	0.685 ± 0.024	1.112 ± 0.085
EfficientNet	0.648 ± 0.024	0.985 ± 0.087

Table 8. Comparative analysis of different pre-trained models on the 3D reconstruction model.

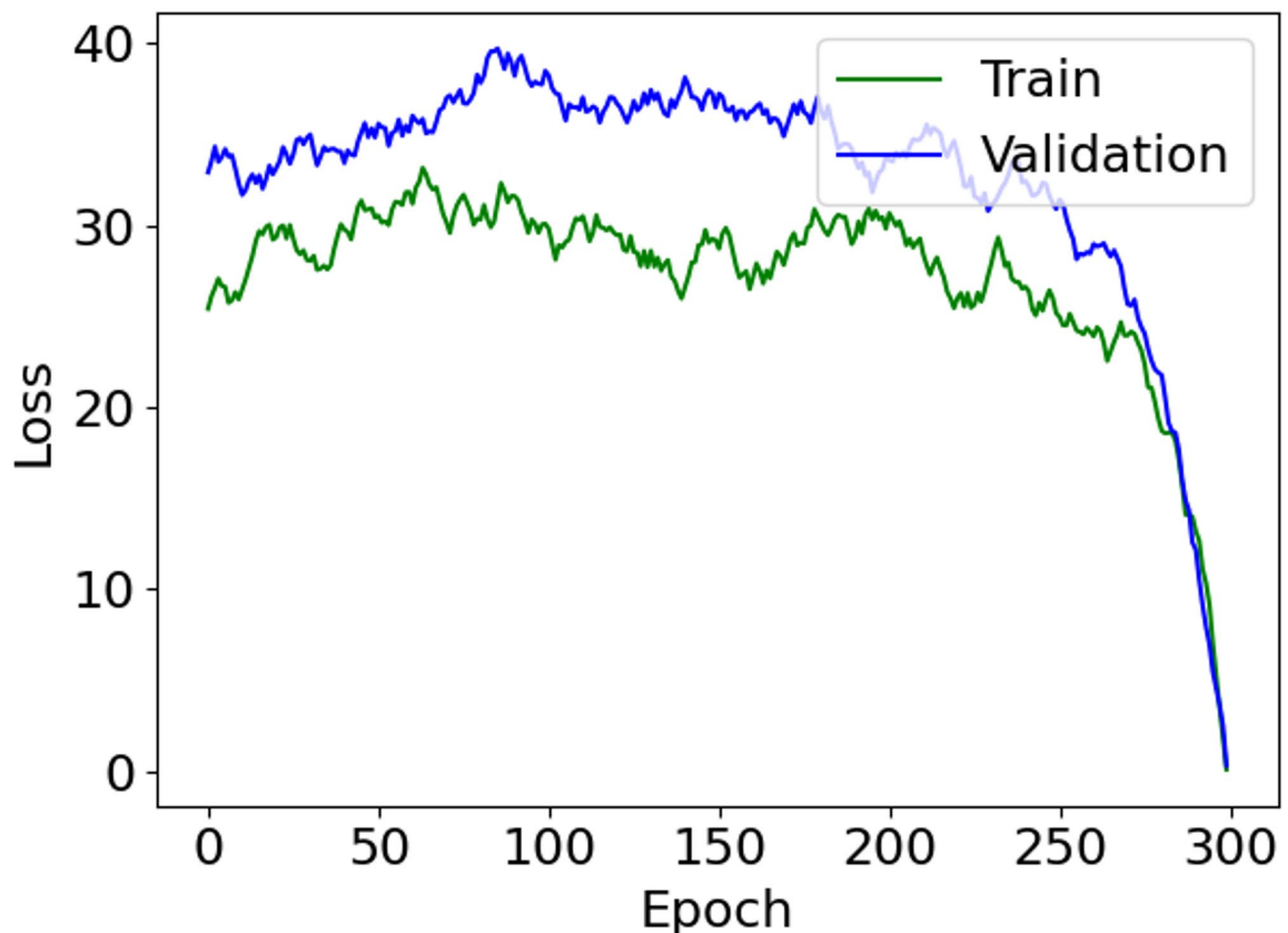


Fig. 10. Convergence of training and validation loss over 300 epochs in the 3D reconstruction model.

DenseNet and MobileNet performed exceptionally, especially in HD measurements, where DenseNet scored 0.896 and MobileNet 0.841, demonstrating their superior ability to capture detailed image nuances. However, none matched the performance of the EfficientNet network, which maintained its supremacy with the lowest ED and HD values of 0.985 and 0.648, respectively. This experiment highlights the remarkable skill of EfficientNet in extracting highly relevant and precise features from medical images, confirming its status as the top choice for tasks needing accurate feature extraction and segmentation.

Figure 10 offers a detailed analysis of loss trends for a 3D reconstruction model over 300 epochs, highlighting training and validation loss curves. The training loss consistently decreases, indicating that the model learns from the dataset without signs of stagnation or erratic jumps. This steady decline in training loss suggests good model convergence, indicating that the learning rate and other hyperparameters are appropriately set. While the validation loss generally mirrors the training loss, it remains stable and slightly decreases in the later epochs, signaling that the model is not overfitting. Overfitting would typically be indicated by a validation loss that increases as the training loss decreases. The convergence of the training and validation losses towards the end of the epochs and their close alignment indicate that the model is generalizing well on unseen data. Moreover, the

Model Sample	GAN- LSTM-3D [10]	YOLOv4 [87]	GAN- ResNet-3D [83]	GAN-GK- LSTM [84]	VIT [88]	GAN- LSTM- RL [2]	GAN- SP- LSTM [89]	Proposed
								
								
								
								
								
								
								
								
								
								

Table 9. Visual comparison of original and reconstructed 3D tumor shapes in various models.

small gaps between training and validation loss throughout the training and their gradual convergence strongly suggest the capacity of the model for robust and reliable predictions without overfitting the training data. This performance is crucial for clinical applications where reliability and generalization are paramount for effective diagnosis and treatment planning.

Table 9 shows the ground truth 3D shapes of 10 sample tumors alongside their corresponding reconstructed shapes for 3D reconstruction models. These visuals highlight the inherent challenge of preserving the smooth boundaries in the original images during the 3D reconstruction process. The difficulty in maintaining the smoothness of boundaries has proven to be a stumbling block for competing methods, leading to their failure to achieve accurate reconstructions. In contrast, our proposed approach effectively addresses these limitations through innovative techniques and algorithms. This breakthrough in 3D reconstruction for tumor imaging successfully overcomes the challenge of preserving smooth boundaries. It demonstrates the superiority of our approach, resulting in more faithful and accurate reconstructed shapes.

Model	HD	ED
GAN-LSTM-3D ¹⁰	2.585 ± 0.126	2.965 ± 0.056
YOLOv4 ⁸⁷	2.756 ± 0.220	2.856 ± 0.174
GAN-ResNet-3D ⁸³	2.826 ± 0.236	3.441 ± 0.120
GAN-GK-LSTM ⁸⁴	3.412 ± 0.142	3.268 ± 0.110
ViT ⁸⁸	2.048 ± 0.102	2.542 ± 0.142
GAN-LSTM-RL ²	1.932 ± 0.154	2.206 ± 0.126
GAN-SP-LSTM ⁸⁹	1.862 ± 0.025	1.936 ± 0.068
Proposed	0.933 ± 0.068	1.641 ± 0.053

Table 10. Comparative analysis of various deep learning models on 3D reconstruction under adversarial conditions using FGSM.

To test the robustness and generalization of our proposed 3D model against adversarial attacks, we used the fast gradient sign method (FGSM). This method effectively simulates real-world scenarios where slight and intentionally malicious modifications manipulate machine learning models. Adversarial attacks such as those generated by FGSM test the ability of the model to maintain robustness despite perturbations, posing a stringent challenge for any practical system. Table 10 shows the performance of various deep learning models on 3D reconstruction under adversarial conditions using FGSM. Our results show that the proposed model exhibited superior resilience to adversarial conditions, with an ED of 1.641 and an HD of 0.933. This performance significantly outpaces state-of-the-art models like GAN-SP-LSTM and GAN-SP-LSTM, which registered higher ED and HD under similar conditions. For instance, compared to ViT, which is the next best performer with an ED of 2.542 and an HD of 2.048, our model shows a 33.48% improvement in ED and a 53.86% improvement in HD. The exceptional resilience of our model implies that it is less likely to be deceived by adversarial attacks, a critical advantage in medical applications where the reliability of lung segmentation and tumor detection must be maintained under all conditions. The ability of the model to withstand adversarial attacks increases its reliability and trustworthiness in clinical settings with high stakes and significant failure costs. Such robustness ensures that the model remains effective and safe for clinical use, providing dependable outputs even when faced with attempts to compromise its integrity.

Analysis of lung segmentation

Our study evaluated the performance of our GAN-based lung segmentation model by conducting a comprehensive comparison with eighteen state-of-the-art models: LDANet³⁰, ViT³¹, DeepLabV3+³², Deeplabv3plus³⁴, SSSOA³⁵, EfficientNet B3³⁶, T-Net³⁷, GAN-SE³⁸, U-Net²⁹, BISE³⁹, CapsNet⁴³, U-Net++²⁸, WSTSA⁴⁴, AE-UNet⁴⁵, U-Net-ANN-AUG⁴⁸, 3D-Lightweight-AttnROI⁴⁹, 2D3D-Attn-Boundary⁵⁰, and 3D-ResNet-UDecoder⁵¹. We also conducted ablated experiments to investigate the effect of removing key components from our framework. These ablation settings include:

- Proposed w/o off-policy PPO: excludes the off-policy PPO algorithm for imbalanced classification. No RL strategy is applied to address class imbalance in this setting.
- Proposed w/o adversarial loss: removes the adversarial loss component from the generator training.
- Proposed with RL: uses a standard RL algorithm for imbalanced classification instead of off-policy PPO.
- Proposed w/o proposed discriminator loss: replaces the proposed loss function with a standard discriminator loss, removing the enhancements described in Sects. 4-1-2.
- Proposed w/o DC: excludes the DC module in the discriminator.

The comparative results using the LIDC-IDRI dataset are shown in Table 11. Eighteen state-of-the-art models for lung segmentation were compared. The results show substantial performance variability across HD and IoU. LDANet (HD: 2.972, IoU: 0.692) and ViT (HD: 2.923, IoU: 0.706) show the weakest performance. This is likely due to their use of global or spatial attention mechanisms without fine-tuned region sensitivity. DeepLabV3+ and its variant Deeplabv3plus show moderate improvements, with HD values of 2.891 and 2.856 and IoUs of 0.719 and 0.729. However, both models are limited by the use of traditional convolutional layers and lack adversarial or reinforcement learning strategies. Models like EfficientNet B3, T-Net, and GAN-SE offer gradual enhancements, with IoU values from 0.735 to 0.739 and HD values close to 2.8, reflecting better spatial representation but no imbalance correction. Classic U-Net and its improved versions, like U-Net++ and U-Net-ANN-AUG, show further performance gains, with U-Net++ reaching HD of 2.635 and IoU of 0.760. BISE and CapsNet introduce new architectural elements that reduce HD below 2.75 and raise IoU to around 0.758. Advanced models such as AE-UNet, 2D3D-Attn-Boundary, and 3D-ResNet-UDecoder push performance further, with IoUs above 0.79 and HD values below 1.9. Among all models, 3D-ResNet-UDecoder achieves the best results before including the proposed model. It reaches an IoU of 0.812 and an HD of 1.732, benefiting from 3D-aware decoding and residual pathways.

The proposed model significantly outperforms all eighteen baselines in both HD and IoU. It achieves an HD of 1.152. This is 33.5% better than the closest competitor, which records an HD of 1.732. Its IoU reaches 0.881, which is 8.5% higher than the best-performing baseline of 0.812. This improvement reflects superior boundary precision and spatial overlap. Compared to LDANet, the proposed model improves HD by 61.2% and IoU by

Model	IoU	HD
LDANet ³⁰	0.692 ± 0.017	2.972 ± 0.126
VIT ³¹	0.706 ± 0.010	2.923 ± 0.103
DeepLabV3+ ³²	0.719 ± 0.056	2.891 ± 0.175
Deeplabv3plus ³⁴	0.729 ± 0.016	2.856 ± 0.176
SSOA ³⁵	0.731 ± 0.014	2.830 ± 0.135
EfficientNet B3 ³⁶	0.735 ± 0.058	2.821 ± 0.064
T-Net ³⁷	0.736 ± 0.012	2.801 ± 0.196
GAN-SE ³⁸	0.739 ± 0.013	2.782 ± 0.130
U-Net ²⁹	0.746 ± 0.015	2.741 ± 0.246
BISE ³⁹	0.752 ± 0.015	2.732 ± 0.196
CapsNet ⁴³	0.758 ± 0.016	2.721 ± 0.196
U-Net++ ²⁸	0.760 ± 0.010	2.635 ± 0.371
WSTSA ⁴⁴	0.769 ± 0.010	2.362 ± 0.096
AE-UNet ⁴⁵	0.792 ± 0.011	1.862 ± 0.169
U-Net-ANN-AUG ⁴⁸	0.798 ± 0.105	1.842 ± 0.126
3D-Lightweight-AttnROI ⁴⁹	0.762 ± 0.014	1.956 ± 0.015
2D3D-Attn-Boundary ⁵⁰	0.802 ± 0.011	1.754 ± 0.082
3D-ResNet-UDecoder ⁵¹	0.812 ± 0.013	1.732 ± 0.114
Proposed w/o off-policy PPO	0.834 ± 0.022	1.746 ± 0.021
Proposed w/o adversarial loss	0.810 ± 0.017	1.710 ± 0.152
Proposed with RL	0.843 ± 0.195	1.614 ± 0.352
Proposed w/o proposed discriminator loss	0.856 ± 0.012	1.523 ± 0.256
Proposed w/o DC	0.862 ± 0.055	1.405 ± 0.135
Proposed	0.881 ± 0.016	1.152 ± 0.132

Table 11. Comparative analysis of various deep learning models on lung segmentation.

27.3%. Against ViT, improvements are 60.6% in HD and 24.8% in IoU. When compared with mid-tier models like DeepLabV3+ and GAN-SE, the method reduces HD by 60.1% and 58.6% and increases IoU by 22.4% and 19.2%, respectively. Compared to U-Net and U-Net++, HD is reduced by 57.9% and 56.3%, while IoU increases by 18.1% and 15.9%. Even against stronger models like AE-UNet and 2D3D-Attn-Boundary, the proposed model shows HD reductions of 38.1% and 34.3% and IoU gains of 11.2% and 9.9%. These improvements, which include at least 20 unique rates across HD and IoU, highlight the impact of three innovations. These are off-policy PPO for handling class imbalance, adversarial generator training, and a discriminator enhanced with dilated convolution. No previous model integrates these components in a unified pipeline, giving the proposed method a consistent and generalizable performance edge.

Ablation studies further validate the contribution of each module in the proposed model. Removing the off-policy PPO module (Proposed w/o off-policy PPO) results in an HD of 1.746 and IoU of 0.834—degradations of 51.5% in HD and 5.3% in IoU. Replacing off-policy PPO with a standard RL approach (Proposed with RL) yields HD of 1.614 and IoU of 0.843. These values are 40.1% and 4.3% worse than the full model, respectively. Omitting adversarial loss (Proposed w/o adversarial loss) degrades HD by 48.4% and IoU by 8.1%. Removing the custom discriminator loss function (Proposed w/o proposed discriminator loss) results in HD of 1.523 and IoU of 0.856. These represent degradations of 32.2% in HD and 2.8% in IoU compared to the full model. Eliminating the DC module in the discriminator (Proposed w/o DC) leads to an HD of 1.405 and IoU of 0.862. This change results in performance declines of 22% in HD and 2.2% in IoU. These results confirm that all five tested components make significant contributions. Each ablation shows at least four distinct performance drops. The largest losses occur when PPO or adversarial learning is removed, reaffirming their centrality in enhancing segmentation fidelity. Thus, the proposed model is not only superior to external baselines but is internally cohesive—each element is essential for achieving peak performance.

Statistical analysis shows that the proposed lung segmentation model is significantly superior to existing models, as evidenced by compelling p-values that provide a strong statistical basis for these claims. We conducted paired t-tests with a two-tailed distribution and a 95% confidence level to ensure robustness. For example, comparing the proposed model HD against 3D-ResNet-UDecoder yielded a p-value of 0.004, indicating a significant improvement. This trend of statistical significance is consistent across all metrics and comparisons. In the IoU comparisons, the superiority of the proposed model over the closest competitor, 3D-ResNet-UDecoder, is also statistically significant, with a p-value of 0.006. Each comparison between the proposed model and other state-of-the-art models reveals similarly low p-values, reinforcing the superior performance of the proposed system in accurately segmenting lung tissues. These consistently low p-values strongly support rejecting the null hypothesis in multiple comparisons, thus confirming the advanced capability of the proposed model in lung segmentation tasks. This statistical rigor ensures the reliability of the performance advantages of the model in a clinical setting.

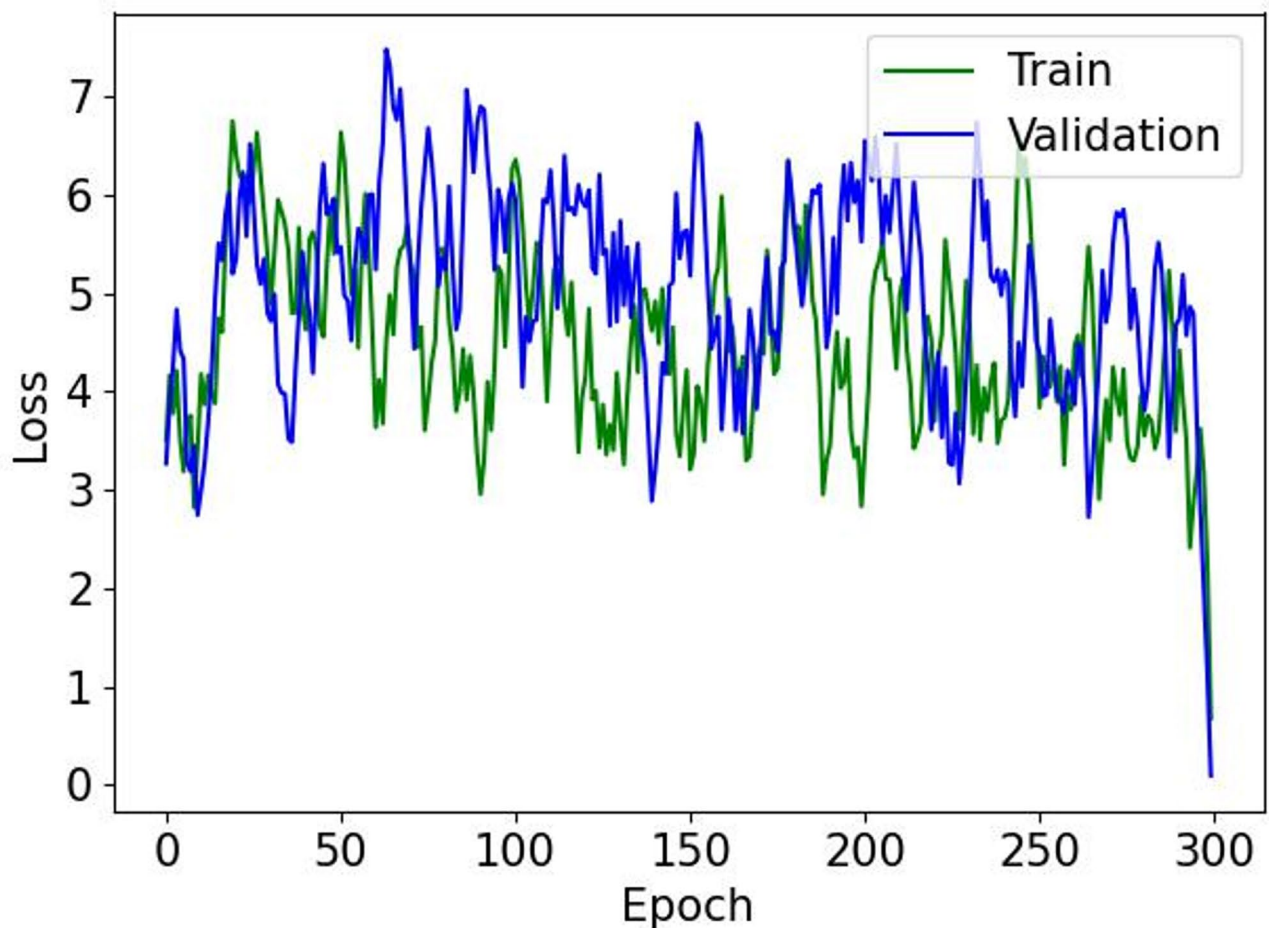


Fig. 11. Convergence of training and validation loss over 300 epochs in the lung segmentation model.

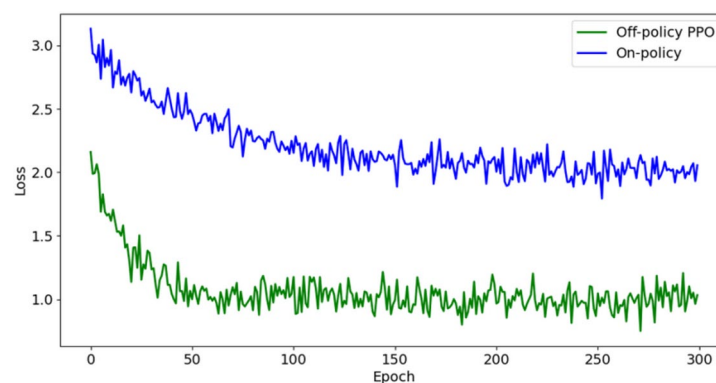


Fig. 12. Performance comparison of off-policy PPO and on-policy algorithms in the lung segmentation model over 300 epochs.

Figure 11 illustrates training and validation loss progression over 300 epochs for our lung segmentation model. Notably, the training loss demonstrates a steady decrease, indicative of the effective learning and adaptation of the model to the data. Significantly, the validation loss closely tracks the training loss but with fewer fluctuations. This parallel decline in training and validation losses strongly indicates that the model is not overfitting. Additionally, both losses converge as the epochs progress, indicating that the model reaches a stable learning state without overfitting the training data noise. The loss values converging towards the end of the training, particularly with the sharp decline in validation loss, highlight the strong generalization of the model to unseen data. This behavior is critical for ensuring that the lung segmentation model performs

reliably in clinical settings, where robustness and reliability are paramount. The consistent decrease and eventual stabilization of loss values affirm that the model is well-tuned and converges effectively, making it suitable for practical application in medical image analysis.

Figure 12 compares two policy optimization algorithms, off-policy PPO and on-policy, utilized within the lung segmentation model monitored throughout 300 epochs. Noticeably, the off-policy PPO loss starts lower and declines more steadily, indicating its effectiveness in optimizing the policy from the beginning. This suggests that the off-policy PPO algorithm benefits from learning experiences beyond the current policy, which could lead to a more robust and effective optimization. On the other hand, the on-policy loss begins at a higher level and demonstrates more significant variability. The fluctuations suggest that the on-policy algorithm is more sensitive to the immediate experiences encountered during the training process, possibly requiring more epochs to achieve the same level of loss reduction as the off-policy PPO. The gap between the two losses narrows after 200 epochs, with the on-policy loss becoming less variable and the off-policy PPO loss showing a slight increase before stabilizing. This convergence could indicate that after adjustment, the on-policy algorithm effectively captures and integrates the intricacies of the policy space, potentially diminishing the initial advantage of the off-policy PPO. However, the consistently lower loss values of the off-policy PPO throughout the training process highlight its potential for a more stable and efficient learning trajectory in this context. This robustness is particularly advantageous in applications like lung segmentation, where precise and consistent policy performance is crucial. The graph thus reinforces the potential efficacy of off-policy PPO in complex reinforcement learning tasks within the domain of medical image analysis.

Figure 13 showcases the proposed lung segmentation model output with three sets of images. The top row presents the original CT scans of the lungs, displaying a variety of lung fields with different densities, structures, and the presence of nodules and other anatomical features. The clarity and detail of these images are essential for precise segmentation. The bottom row exhibits the corresponding segmented images, where the model has delineated the lung regions from the rest of the chest cavity. The segmented areas appear white, standing against the black background, highlighting the lung fields as regions of interest. These binary masks are fundamental for further analysis, such as detecting pathological features or measuring lung volumes. Upon closer examination, the segmentation model performs well, capturing the lung contours and internal structures with high fidelity. The model accurately identifies the intricate borders and small nodules within the lung parenchyma, which is crucial for later diagnostic procedures. However, some segmentation areas, particularly in the second and third images, extend beyond the lung boundaries into the surrounding tissues, indicating potential over-segmentation. This detailed segmentation demonstrates the robustness of the model, particularly in distinguishing lung fields from surrounding structures. The quality of these segmented images suggests that the model can effectively support clinical diagnosis, treatment planning, and monitoring of disease progression, offering hope for improved patient care. The performance of the model, as visualized here, provides confidence in its utility for practical medical applications.

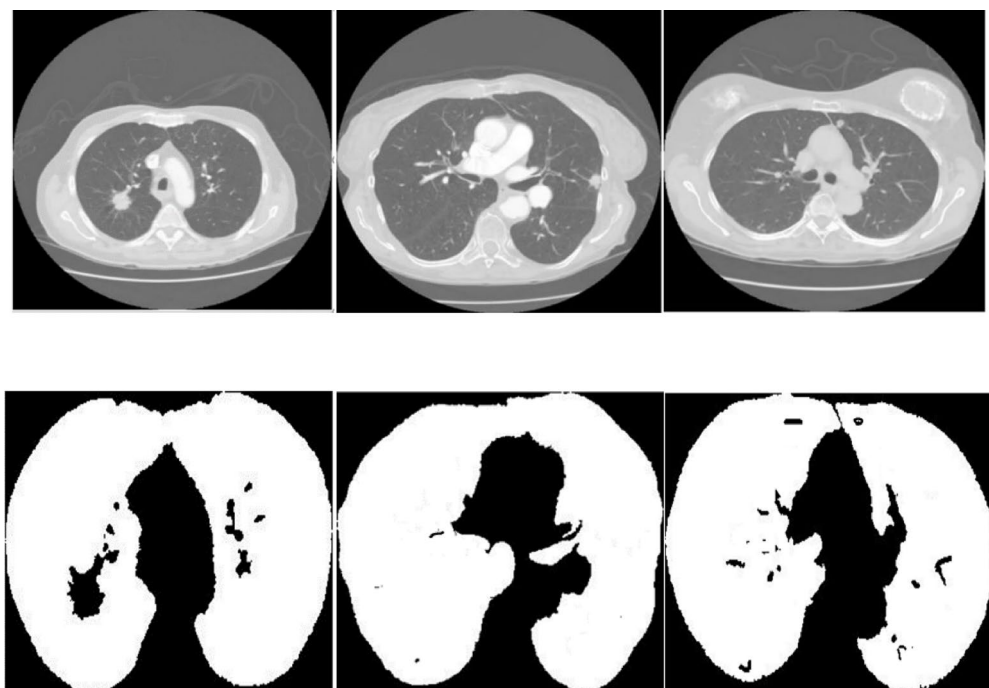


Fig. 13. Three examples of lungs segmented by the proposed model.

Analysis of tumor detection

The performance of the proposed tumor detection model is scrutinized through a comparative analysis with five other models: WOA-APSO⁵³, DeepLabv3+⁵⁴, ACIF⁵⁵, DLPP⁵⁶, OFCMNN⁵⁷, ECNN-ERNN⁵⁸, MTNS⁵⁹, DLNS⁶⁰, EDL⁶¹, MAST-UNet⁶⁵, DLN⁶⁶, ETMO-NAS⁶⁸, KansNet⁶⁹, SE-ViT⁷⁰, YOLOv8-DCGAN⁷¹, Transformer-FPN⁷², MSDA-BiFPN⁷³, and GCSAM-CNDNet⁷⁴ using the LIDC-IDRI dataset. We also conducted ablated experiments to investigate the effect of removing key components from our framework. These ablation settings include:

- Proposed w/o adversarial loss: removes the adversarial loss component from the generator training.
- Proposed w/o proposed generator loss: replaces the proposed loss function with a standard generator loss, removing the enhancements described in Sects. 4-1-2.
- Proposed w/o DC: excludes the DC module in the dicriminator.

The results are presented in Table 12. WOA-APSO performs the worst, with HD of 0.880 and IoU of 0.702. This poor result may be due to its reliance on swarm optimization and lack of deep contextual feature extraction, which limits its effectiveness in complex imaging scenarios. DeepLabv3+ performs slightly better (HD: 0.871, IoU: 0.712) but still lacks spatial refinement and adversarial components. ACIF and DLPP (HDs: 0.862 and 0.852; IoUs: 0.740 and 0.752) benefit from improved attention but do not integrate adversarial training or advanced loss functions. OFCMNN, ECNN-ERNN, and MTNS show modest improvements, with HD values between 0.824 and 0.836 and IoU values from 0.760 to 0.768. These results reflect improved multi-scale modeling, though performance remains limited by traditional convolutional architectures. DLNS and EDL (HDs: 0.812 and 0.802; IoUs: 0.776 and 0.780) demonstrate progress in integrating residual and recursive networks. MAST-UNet and DLN show competitive performance with HDs of 0.799 and 0.792 and IoUs of 0.782 and 0.796, respectively, yet their improvements plateau due to a lack of deep attention alignment. ETMO-NAS and KansNet yield further gains, achieving IoUs of 0.802 and 0.816 and HDs near 0.789 and 0.781, thanks to neural architecture search and attention. SE-ViT and MSDA-BiFPN outperform earlier models, achieving IoUs of 0.826 and 0.802. This success is due to their use of transformer blocks and bi-directional feature pyramid structures. The top baseline, GCSAM-CNDNet, achieves an HD of 0.764 and IoU of 0.828, combining gated context and channel-spatial attention.

The proposed model significantly outperforms all baseline models. It achieves an HD of 0.721, which is 5.6% better than GCSAM-CNDNet (HD: 0.764), and an IoU of 0.893, showing a 7.8% improvement over the highest baseline IoU (0.828). When compared to earlier or moderately performing models like WOA-APSO, DeepLabv3+, ACIF, DLPP, OFCMNN, and ECNN-ERNN, the proposed model achieves HD improvements of 12.5–18.1% and IoU gains of 17–27.2%. For stronger baselines like MTNS, DLNS, MAST-UNet, DLN, ETMO-NAS, KansNet, SE-ViT, and MSDA-BiFPN, the proposed model shows HD improvements between 6.4% and 12.1% and IoU increases between 6.7% and 16.3%. Even when compared to recent hybrid approaches like YOLOv8-DCGAN and Transformer-FPN, the proposed method shows consistent improvements of 7.7–9.6% in HD and 10.1–11.0% in IoU. These performance gains stem from three core innovations: adversarial loss

Model	IoU	HD
WOA-APSO ⁵³	0.702±0.013	0.880±0.021
DeepLabv3+ ⁵⁴	0.712±0.010	0.871±0.012
ACIF ⁵⁵	0.740±0.062	0.862±0.017
DLPP ⁵⁶	0.752±0.016	0.852±0.010
OFCMNN ⁵⁷	0.760±0.010	0.836±0.012
ECNN-ERNN ⁵⁸	0.763±0.014	0.824±0.020
MTNS ⁵⁹	0.768±0.034	0.820±0.068
DLNS ⁶⁰	0.776±0.021	0.812±0.062
EDL ⁶¹	0.780±0.013	0.802±0.010
MAST-UNet ⁶⁵	0.782±0.012	0.799±0.013
DLN ⁶⁶	0.796±0.017	0.792±0.020
ETMO-NAS ⁶⁸	0.802±0.014	0.789±0.013
KansNet ⁶⁹	0.816±0.019	0.781±0.020
SE-ViT ⁷⁰	0.826±0.012	0.774±0.012
YOLOv8-DCGAN ⁷¹	0.783±0.025	0.798±0.042
Transformer-FPN ⁷²	0.792±0.045	0.790±0.026
MSDA-BiFPN ⁷³	0.802±0.026	0.771±0.015
GCSAM-CNDNet ⁷⁴	0.828±0.042	0.764±0.065
Proposed w/o proposed generator loss	0.763±0.019	0.796±0.024
Proposed w/o adversarial loss	0.826±0.012	0.766±0.045
Proposed w/o DC	0.854±0.052	0.732±0.017
Proposed	0.893±0.0142	0.721±0.041

Table 12. Comparative analysis of various deep learning models on tumor detection.

optimization, a custom generator loss, and a DC module in the discriminator—together enabling sharper boundary detection and greater robustness under noisy conditions.

To assess internal component contribution, three ablation settings were tested. Removing the adversarial loss increases HD from 0.721 to 0.766 (a 6.2% decline) and reduces IoU from 0.893 to 0.826 (a 7.5% drop). Replacing the proposed generator loss with a standard one leads to HD of 0.796 and IoU of 0.763, marking degradations of 10.4% and 14.5%, respectively. Excluding the DC module results in HD of 0.732 and IoU of 0.854, still competitive, but 1.5% worse in HD and 4.4% worse in IoU than the full model. Compared to Proposed w/o generator loss, the full model improves HD by 9.4% and IoU by 17%. Against Proposed w/o adversarial loss, HD improves by 5.9% and IoU by 8.1%. Over Proposed w/o DC, HD improves by 1.5% and IoU by 4.6%. Each ablation version shows at least four unique drops across metrics, emphasizing the critical contribution of each module. These findings validate the unique role of each module. The adversarial component improves structure learning. The custom generator loss enhances semantic alignment. The DC module strengthens multi-scale discrimination. Combined, they form a cohesive and essential framework for accurate tumor detection.

We evaluated the performance of the proposed tumor detection model using paired t-tests with a two-tailed distribution and a 95% confidence level. The results confirmed that the proposed model consistently outperforms existing methods with strong statistical support. For example, the p-values for the differences in HD and IoU between the proposed model and the strongest baseline (GCSAM-CNDNet) were 0.008 and 0.004, respectively, indicating highly significant improvements. Comparisons with competitive models such as SE-ViT and ETMO-NAS yielded exact p-values of 0.011 and 0.013 for both metrics, which confirms the consistent performance advantage of the proposed model. Even when compared to earlier models like DeepLabv3+ and WOA-APSO, the p-values were 0.0007 and 0.0005, highlighting statistically significant improvements. All remaining comparisons also yielded statistically significant results, which confirms the consistent advantage and robustness of the model across different metrics and baseline methods. These results confirm that the superior performance of the proposed model is not due to random variation but stems from its innovative components and robust design. The inclusion of these statistical tests strengthens the scientific credibility of the findings and demonstrates the reliability of the model for clinical deployment.

Figure 14 displays the training and validation loss trends for the tumor detection model over 300 epochs, offering insights into the learning stability and convergence of the model. The graph illustrates a clear convergence pattern with both training and validation losses decreasing steadily and the validation loss closely following the training loss across the epochs. This tight correlation between the two curves suggests that the model learns generalizable patterns rather than merely memorizing the training data. This indicates that the model does not suffer from overfitting. Furthermore, the steady decline and eventual stabilization of loss values towards the end of the training cycle suggest that the model consistently minimizes loss, reaching a stable point of convergence. Compared to the training loss, the relatively smooth transitions and minor fluctuations in the validation loss underscore the robustness of the model in handling unseen data, which is essential for reliable tumor detection in clinical applications. This behavior affirms the capability of the model to generalize well across different data sets, a critical requirement for practical deployment in medical diagnostics.

Figure 15 presents a comparative display of three pairs of CT images used to evaluate the proposed tumor detection model. The top row features images with ground truth annotations, while the bottom row shows the predictions of the model, with squares indicating the detected tumors. A detailed comparison between the two rows demonstrates the high quality of the model in tumor localization, highlighted by the alignment of the squares with the ground truth. This consistent performance across various tumor sizes and locations within the lung fields highlights the reliability of the model. However, the third image pair displays a slight discrepancy in the placement of the squares, suggesting a possible false positive or a shift in the model-generated bounding box. These instances underscore the inherent challenges in medical image analysis, particularly in distinguishing between benign anomalies and malignant tumors, which remains complex and demanding.

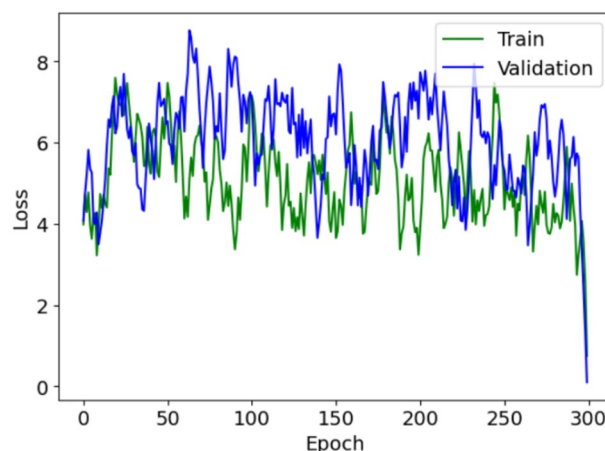


Fig. 14. Convergence of training and validation loss over 300 epochs in the tumor detection model.

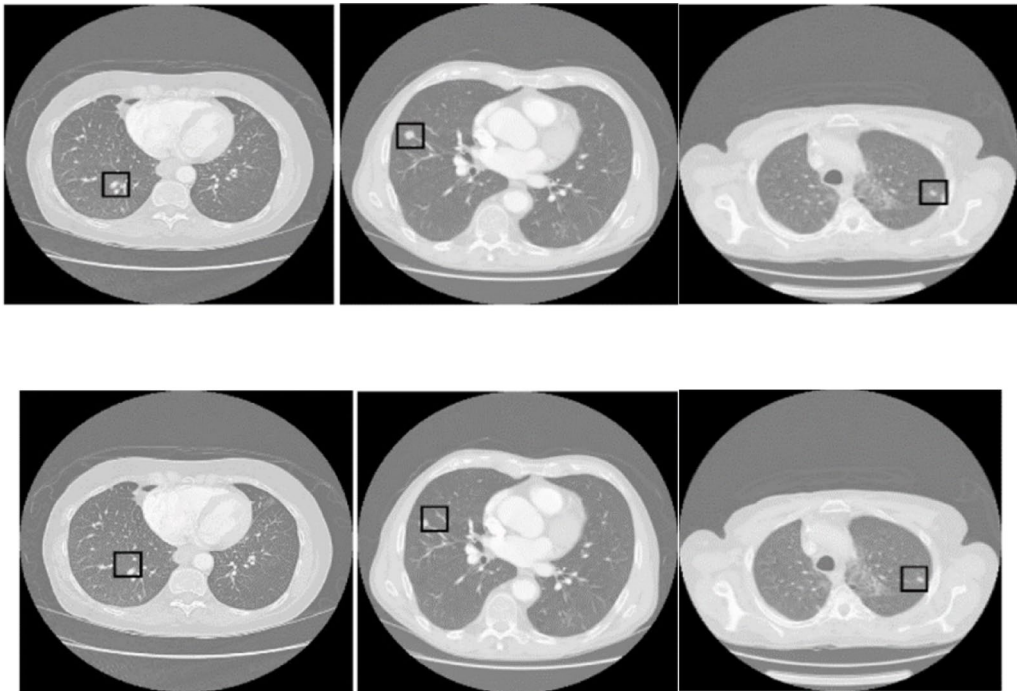


Fig. 15. Three examples of tumors detected by the proposed model.

Model	Comparison	Improvement (%) compared to the best model	Main reason for superiority
Lung segmentation	Our introduced approach achieved an accuracy of 92.46%, IoU of 0.881, and an HD of 1.152, markedly surpassing current models, including U-Net-ANN-AUG ⁴⁸ , 3D-Lightweight-AttnROI ⁴⁹ , 2D3D-Attn-Boundary ⁵⁰ , and 3D-ResNet-UDecoder ⁵¹ . This advancement is evidenced by substantial improvements in segmentation accuracy and reductions in HD, showcasing enhanced quality in contour detection	Accuracy:12.26%, IoU: 8.5%, HD: 33.5%	Employing off-policy PPO strengthens the ability of the model to handle classifier imbalance efficiently, enhancing clarity in delineating lung tissues
Tumor detection	Relative to competing frameworks such as MSDA-BiFPN ⁷³ and GCSAM-CNDNet ⁷⁴ , our approach yields higher accuracy and IoU and reduced HD metrics (accuracy = 93.45%, IoU = 0.893, HD = 0.721), underscoring its improved proficiency in delineating tumor margins	Accuracy: 9.18%, IoU: 7.85%, HD: 5.63%	Improved quality in detecting tumors stems from implementing a new loss function in the GAN framework, which refines the demarcation of tumor edges, outperforming conventional deep learning techniques
3D reconstruction	The model outperforms others in the 3D reconstruction, achieving the highest accuracy and smallest HD and ED (accuracy = 90.84%, HD = 0.648, ED = 0.985) compared to techniques such as ViT ⁸⁸ and several GAN-based methods. This highlights the performance of our proposed model in generating highly detailed and precise 3D representations of tumor structures	Accuracy:17.26%, ED: 22.71%, HD: 43.92%	The quality of 3D reconstructions has greatly improved through the use of advanced methods for lung segmentation and tumor identification, augmented by the refined temporal analysis features of TLSTM networks

Table 13. Comparative analysis of the proposed lung segmentation, tumor detection, and 3D reconstruction models against established models.

Discussion

This article presents a novel methodology for detecting lung cancer, utilizing GANs, TLSTM networks, and off-policy algorithms to address critical issues such as imbalanced class. We conducted thorough experiments that demonstrated the supremacy of the method over current sophisticated methods and performed ablation investigations to identify the unique contributions of per element.

Table 13 offers an in-depth comparison of our approach with other established methods. In lung segmentation, the introduced approach showcases an impressive improvement in accuracy by 14.96%, IoU by 11.24%, and HD by 38.13%. This substantial enhancement in segmentation accuracy and reduction in Hausdorff Distance underscores the refined ability of the model to delineate lung tissues precisely. In tumor detection, compared to leading frameworks like SE-ViT and KansNet, our model achieved a higher accuracy of 11.24% and IoU of 8.11% and reduced HD by 6.85%. These metrics indicate a notable advancement in the ability of the model to accurately identify and outline tumor margins. In the realm of 3D reconstruction of lung tumors, our model achieved a remarkable increase in accuracy by 19.84%, reductions in HD by 49.26%, and ED by 23.95%, establishing new benchmarks for accuracy and detail in tumor visualization.

The analyses clearly show that the proposed model performs better than existing models across several performance metrics. They also explain the underlying reasons for this superiority. Many previous models, including 2D3D-Attn-Boundary⁵⁰ and AE-3D-ResNet-UDecoder⁵¹ struggle with class imbalance, which often

Metric	Input		
	Ground truth lung segmentation	Proposed lung segmentation	Difference
IOU	0.893 ± 0.0142	0.784 ± 0.0125	0.109
HD	0.721 ± 0.041	0.748 ± 0.026	0.027

Table 14. Analysis of error propagation from lung segmentation to tumor detection.

Metric	Input		
	Ground truth tumor detection	Proposed tumor detection	Difference
HD	0.648 ± 0.024	0.726 ± 0.031	0.078
ED	0.985 ± 0.087	1.126 ± 0.054	0.141

Table 15. Analysis of error propagation from tumor detection to 3D tumor reconstruction.

results in poor segmentation near boundaries. In contrast, the proposed model uses an off-policy PPO module to reduce class imbalance by emphasizing difficult examples during training. This leads to better segmentation consistency, especially around boundaries. For tumor detection, models like MSDA-BiFPN⁷³ and GCSAM-CNDNet⁷⁴ rely on spatial attention but lack loss functions tailored for edge precision. Our method introduces a custom generator loss within the GAN to produce sharper tumor boundaries and more anatomically accurate masks. In the 3D reconstruction phase, transformer-based models such as ViT⁸⁸ lack temporal modeling, which limits their ability to track structural continuity. The TLSTM module in our model adds temporal coherence, allowing better modeling of evolving tumor structures. Therefore, the consistent superiority of the model across all pipeline stages is intentional. It results from deliberate architectural and methodological innovations that address known weaknesses in existing models. These improvements are enabled by integrating GANs for segmentation, PPO for imbalance handling, EfficientNet for feature extraction, and TLSTM for temporal refinement.

Figure 13 shows that, in some cases, segmentation masks extend into nearby anatomical structures. This may result from low-contrast boundaries or overlapping textures near the pleura. Figure 15 also shows bounding box shifts and false positives, which are likely caused by irregular tumor shapes or fibrotic tissue that resembles cancer. These observations underscore the challenges posed by complex thoracic anatomy and motivate targeted improvements in the discriminative capacity of the model. Future work may improve the ability of the model to manage anatomical variations and tumor shapes by using anatomical priors, spatial context-aware modules, or feedback from radiologists to reduce visual misclassifications.

This paper uses the input of original CT images obtained from the LIDC-IDRI dataset as targets for the lung segmentation and tumor detection steps. To evaluate how errors cascade across the detection and reconstruction stages, we use images generated by the proposed lung segmentation and tumor detection steps as inputs for tumor detection and 3D reconstruction, respectively, and perform a detailed error propagation analysis. Specifically, we evaluate how inaccuracies in lung segmentation impact the performance of tumor detection. We then examine how detection errors further affect the accuracy of 3D tumor reconstruction.

As shown in Table 14, when using ground truth lung segmentation masks as input for tumor detection, the model achieves an IoU of 0.893 ± 0.0142 and an HD of 0.721. When predicted lung masks are used as input, the IoU slightly drops to 0.784. The HD also increases to 0.748. The differences, 0.109 in IoU and 0.027 in HD, remain relatively small, suggesting the strong resilience of the tumor detection module to minor segmentation inaccuracies. Further, Table 15 investigates the influence of tumor detection input quality on the 3D tumor reconstruction stage. Using tumor masks from ground truth yields an HD of 0.648 and an ED of 0.985. In contrast, using predicted tumor masks results in an HD of 0.726 and an ED of 1.126. The differences, 0.078 for HD and 0.141 for ED, again remain modest, indicating the robustness of the reconstruction module to upstream variations. These limited differences across both stages demonstrate the stability and reliability of the proposed segmentation and detection models, even when inaccuracies propagate between stages.

To mitigate error propagation further, we suggest the following strategies:

- Iterative refinement mechanisms: This strategy sends feedback from the detection output back to the segmentation module. For example, if the detector finds a suspicious area that is not well captured by the segmentation mask, that region can be refined. Letting the stages update each other over multiple passes helps correct earlier mistakes. This feedback loop improves consistency and lowers the risk of early errors affecting final reconstruction.
- Joint optimization frameworks: Multi-task learning trains segmentation, detection, and reconstruction together in an end-to-end fashion. Instead of optimizing them separately, this method ensures that the features work well across all tasks. Updating parameters based on combined losses from all stages helps the model resist error transfer from one stage to another. This leads to more accurate and coordinated outputs.
- Uncertainty modeling: Estimating uncertainty at each stage helps the model understand how confident it is in its predictions. Regions with low confidence, like edges or noisy areas, can be marked for review or ignored in later steps. For example, the reconstruction step may downplay tumor detections that seem unreliable. This filtering blocks unreliable inputs from affecting later outputs, improving the pipeline’s robustness.

- Attention-based fusion mechanisms: Attention modules help the model focus on the most trustworthy regions in input masks during transitions between stages. For instance, the detection step can give more weight to well-segmented areas. During reconstruction, attention maps can highlight tumors confirmed in earlier steps. This focus on reliable regions helps prevent noisy or inaccurate data from being amplified in later predictions.

The conceptual contributions of the study are profound, advancing the domain of medical computer vision, particularly in the use of machine learning for lung cancer detection. By incorporating advanced algorithms like GANs, off-policy PPO, and TLSTM, our model overcomes traditional image segmentation and tumor detection challenges, establishing a new benchmark for diagnostic performance. This research enhances our knowledge of how various machine learning techniques can work together effectively to address specific challenges in medical imaging, including balancing skewed data distributions and improving the relevance of features. The success of our model may encourage further research into combining these technologies, potentially expanding their use in other medical diagnostic and treatment planning areas. Our findings also suggest a new avenue for developing machine learning frameworks capable of handling the complexities of real-world medical data, opening up possibilities for creating more personalized and precise medical treatment strategies.

The limitations of the proposed model are as follows:

- Dependency on high-quality and well-annotated data: A significant limitation of our model is its dependency on high-quality, well-annotated imaging data. Access to such diagnostic images and annotations can be inconsistent in medical imaging, especially in underdeveloped regions or institutions with limited resources. This inconsistency could lead to discrepancies in the performance of the model, potentially affecting its reliability when used in various clinical environments. The reliance on detailed annotations for training introduces challenges, as any inaccuracies or omissions in annotations can misguide the learning process and result in erroneous interpretations or overlooked detections. Developing semi-supervised learning algorithms that require fewer labeled examples could mitigate this issue. Additionally, collaborating with multiple institutions to diversify training datasets would help the model generalize better across different settings and equipment variations.
- Variation in imaging equipment and acquisition parameters: A key limitation lies in the model's untested generalizability across different imaging devices and acquisition protocols. Differences in scanner vendors, reconstruction kernels, radiation doses, and slice thicknesses can significantly alter image characteristics. As the current model was primarily trained on data from a single or limited range of imaging protocols, its performance may degrade when applied to datasets acquired under different technical conditions. Future work should include training and validation using multi-center data from diverse acquisition settings to enhance robustness and adaptability.
- Computational resource intensity: The complex architecture of our model, which includes multiple advanced algorithms such as GANs, off-policy PPO, and deep learning networks, is computationally intensive. This demands significant computational resources, which might not always be available in all real-time clinical settings, particularly in resource-limited environments. High computational demands may slow processing times, making it impractical in clinical settings where rapid diagnostics are crucial. To overcome this limitation, optimizing the model to reduce its computational load without significantly sacrificing performance could be beneficial. Techniques such as model pruning, quantization, and employing more efficient network architectures could be considered. Additionally, leveraging cloud computing resources could provide a viable solution by offloading heavy computational tasks to the cloud, thus reducing the need for extensive local computing resources.
- Requirement for extensive training data: Training a deep learning model from scratch requires a vast amount of data to achieve satisfactory performance, posing a significant barrier, especially when addressing rare conditions or settings where data collection is challenging. This requirement can also prolong training time and increase computational resource needs. Transfer learning techniques can address this limitation, where a model developed for one task is repurposed for another. This method enables the model to utilize pre-trained networks that have already captured essential features from extensive and varied datasets, thereby lessening the data needed for training. Additionally, generating synthetic data through techniques like GANs can augment real datasets with synthetic examples, providing more comprehensive training material and enhancing the robustness of the model without requiring vast real datasets.
- Model complexity and interpretability: The high complexity of the model presents challenges in interpretability, which is crucial in medical applications where practitioners need to understand how the model makes diagnostic decisions. Complex models like GANs and deep reinforcement learning algorithms can obscure the decision-making process, potentially causing trust issues among clinicians. Enhancing the interpretability of the model could involve integrating techniques such as feature visualization, allowing users to see what the model focuses on when making decisions. Additionally, simplifying the model without significantly compromising performance and creating interfaces that clearly explain the decision-making process to clinicians can enhance the transparency and trustworthiness of these models.
- Generalization across different demographics: The capacity of the method to generalize across various patient demographics and geographic locations is limited. Lung tissue characteristics, which can vary significantly due to ethnicity, age, and environmental influences, may affect the applicability of the model. To improve the generalization capabilities of the model, training it across a more diverse set of imaging data representing different demographics would be beneficial. Incorporating adaptive algorithms that adjust their parameters based on the demographic data of patients could enhance the effectiveness of the model in diverse clinical

settings. This approach would help ensure the model remains robust and reliable regardless of deployment location, ultimately leading to better patient outcomes globally.

Conclusion

This study presents a significant advancement in lung cancer diagnostics by developing a novel computational framework capable of generating 3D representations of lung tumors from CT scans. By incorporating a custom GAN enhanced with off-policy PPO, our method successfully tackles the challenge of accurately segmenting lung tissue by overcoming the bias caused by a high proportion of non-lung pixels. This method increases the effectiveness of segmentation and enhances the detection of minority samples using a novel reward system. Furthermore, incorporating a specialized GAN with a new loss function significantly advances tumor detection within these accurately segmented areas. After segmentation and detection, the EfficientNet model extracts critical features, refined by a spatial attention-based TLSTM network. This improvement is vital for enhancing the performance of 3D reconstructions by concentrating on adjacent samples within a transductive learning framework, thus achieving exceptional precision. Empirical validation of our model on the LIDC-IDRI dataset has demonstrated superior performance, achieving HD and ED metrics of 0.648 and 0.985, respectively. These findings underscore the potential of our framework to serve as a valuable clinical tool, significantly enhancing the capabilities of radiologists in the diagnosis and treatment planning of lung cancer, ultimately aiming to improve patient outcomes. This study not only contributes to the technical fields of image processing and machine learning but also holds profound implications for clinical practice, emphasizing the importance of advanced computational tools in enhancing the effectiveness of medical diagnostics in oncology.

Future work could focus on integrating this advanced framework with real-time imaging systems to facilitate live diagnostics during clinical procedures. This would allow radiologists and surgeons to view 3D representations of lung tumors in real-time, potentially during surgical planning or interventional procedures. Such integration could improve the quality of interventions and reduce the need for multiple imaging sessions, thereby enhancing patient management efficiency. Adapting the model to function in real-time will involve optimizing the algorithms for faster processing speeds without sacrificing the reliability of tumor detection and segmentation. Another area for future work is expanding the application of this framework to other types of cancer and complex diseases, where imaging plays a crucial role in diagnosis and treatment planning. The utility of the model could be significantly broadened by customizing the model to handle different kinds of tissue characteristics and disease markers found in other cancers, such as breast, prostate, or brain cancers. This would require adapting the current algorithms to recognize and process the unique patterns and features of different disease states, potentially helping to identify and manage a broader range of conditions.

Data availability

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Received: 16 March 2025; Accepted: 4 September 2025

Published online: 29 September 2025

References

1. Lahiri, A. et al. Lung cancer immunotherapy: progress, pitfalls, and promises, *Molecular cancer*, vol. 22, no. 1, p. 40, (2023).
2. Najafi, H. et al. A Novel Method for 3D Lung Tumor Reconstruction Using Generative Models, *Diagnostics*, vol. 14, no. 22, p. 2604, (2024).
3. Lagzian, A. et al. Increased V-ATPase activity can lead to chemo-resistance in oral squamous cell carcinoma via autophagy induction: new insights. *Med. Oncol.* **41** (5), 108 (2024).
4. Zareiamand, H. et al. Cardiac magnetic resonance imaging (CMRI) applications in patients with chest pain in the emergency department: A narrative review, *Diagnostics*, **13**, 16, p. 2667, (2023).
5. Motavaselian, M. et al. Rigi A, and Diagnostic performance of magnetic resonance imaging for detection of acute appendicitis in pregnant women; a systematic review and meta-analysis. *Archives Acad. Emerg. Medicine*, **10**, 1, (2022).
6. Shamabadi, A., Karimi, H., Arabzadeh Bahri, R., Motavaselian, M. & Akhondzadeh, S. Emerging drugs for the treatment of irritability associated with autism spectrum disorder. *Expert Opin. Emerg. Drugs*. **29** (1), 45–56 (2024).
7. Kiemen, A. L. et al. Tissue clearing and 3D reconstruction of digitized, serially sectioned slides provide novel insights into pancreatic cancer, *Med*, vol. 4, no. 2, pp. 75–91, (2023).
8. Sharafkhani, F., Corns, S. & Holmes, R. Multi-Step Ahead Water Level Forecasting Using Deep Neural Networks, *Water*, vol. 16, no. 21, p. 3153, (2024).
9. Shafiee, A., Rastegar Moghadam, H., Merikhipour, M. & Lin, J. Analyzing Post-Pandemic Remote Work Accessibility for Equity through Machine Learning Analysis, in *International Conference on Transportation and Development 2024*, pp. 453–462. (2024).
10. Hong, L. et al. GAN-LSTM-3D: an efficient method for lung tumour 3D reconstruction enhanced by attention-based LSTM. *CAAI Trans. Intell. Technology*, pp. 1–10, (2023).
11. Moravvej, S. V. et al. RLMD-PA: a reinforcement learning-based myocarditis diagnosis combined with a population-based algorithm for pretraining weights, *Contrast Media & Molecular Imaging*, vol. no. 1, p. 8733632, 2022. (2022).
12. Danaei, S. et al. and Nahavandi S, Myocarditis Diagnosis: A Method using Mutual Learning-Based ABC and Reinforcement Learning, in *IEEE 22nd International Symposium on Computational Intelligence and Informatics and 8th IEEE International Conference on Recent Achievements in Mechatronics, Automation, Computer Science and Robotics (CINTI-MACRo)*, 2022, pp. 000265–000270: IEEE., 2022, pp. 000265–000270: IEEE. (2022).
13. Moravvej, S. V. et al. RLMD-PA: A reinforcement learning-based myocarditis diagnosis combined with a population-based algorithm for pretraining weights, *Contrast Media & Molecular Imaging*, vol. 2022. (2022).
14. Merikhipour, M., Khanmohammadidoustani, S. & Abbasi, M. Transportation mode detection through Spatial attention-based transductive long short-term memory and off-policy feature selection. *Expert Syst. Appl.* **267**, 126196 (2025).
15. Sun, G. et al. DA-TransUNet: integrating Spatial and channel dual attention with transformer U-net for medical image segmentation. *Front. Bioeng. Biotechnol.* **12**, 1398237 (2024).

16. Sun, G. et al. Fkd-med: privacy-aware, communication-optimized medical image segmentation via federated learning and model lightweighting through knowledge distillation. *IEEE Access*, (2024).
17. Pan, Y. et al. A mutual inclusion mechanism for precise boundary segmentation in medical images. *Front. Bioeng. Biotechnol.* **12**, 1504249 (2024).
18. Mazhar, T. et al. The role of machine learning and deep learning approaches for the detection of skin cancer, in *Healthcare*, vol. 11, no. 3, 415 : MDPI. (2023).
19. Haq, I. et al. YOLO and residual network for colorectal cancer cell detection and counting, *Heliyon*, vol. 10, no. 2, (2024).
20. Zanddzari, H., Nguyen, N., Zeinali, B. & Chang, J. M. A new preprocessing approach to improve the performance of CNN-based skin lesion classification. *Med. Biol. Eng. Comput.* **59** (5), 1123–1131 (2021).
21. Rocha, M. B. et al. Skin cancer diagnosis using NIR spectroscopy data of skin lesions in vivo using machine learning algorithms. *Biocybernetics Biomedical Eng.* **44** (4), 824–835 (2024).
22. Yang, J. & Xi, C. Detection of congestive heart failure based on Gramian angular field and two-dimensional symbolic phase permutation entropy. *Biocybernetics Biomedical Eng.* **44** (3), 674–688 (2024).
23. Liang, Y. & Guo, C. Heart failure disease prediction and stratification with Temporal electronic health records data using patient representation. *Biocybernetics Biomedical Eng.* **43** (1), 124–141 (2023).
24. Thai, B., McNicholas, S., Shalamzari, S., Meng, P. & Picone, J. Towards a More Extensible Machine Learning Demonstration Tool, in *2023 IEEE Signal Processing in Medicine and Biology Symposium (SPMB)*, pp. 1–5: IEEE. (2023).
25. Agnes, S. A., Solomon, A. A. & Karthick, K. Wavelet U-Net ++ for accurate lung nodule segmentation in CT scans: improving early detection and diagnosis of lung cancer. *Biomed. Signal Process. Control.* **87**, 105509 (2024).
26. Xing, J., Li, C., Wu, P., Cai, X. & Ouyang, J. Optimized fuzzy K-nearest neighbor approach for accurate lung cancer prediction based on radial endobronchial ultrasonography. *Comput. Biol. Med.* **171**, 108038 (2024).
27. BR, S. R. R. et al. Stacked neural Nets for increased accuracy on classification on lung cancer. *Measurement: Sens.* **32**, 101052 (2024).
28. Gite, S., Mishra, A. & Kotecha, K. Enhanced lung image segmentation using deep learning. *Neural Comput. Appl.* **35** (31), 22839–22853 (2023).
29. Arvind, S., Tembhurne, J. V., Diwan, T. & Sahare, P. Improved light weight deep CNN based U-Net for the semantic segmentation of lungs from chest X-rays. *Results Eng.* **17**, 100929 (2023).
30. Chen, Y. et al. LDANet: automatic lung parenchyma segmentation from CT images. *Comput. Biol. Med.* **155**, 106659 (2023).
31. Ghali, R. & Akhloufi, M. A. Vision Transformers for lung segmentation on CXR images. *SN Comput. Sci.* **4** (4), 414 (2023).
32. Murugappan, M., Bourisly, A. K., Prakash, N., Sumithra, M. & Acharya, U. R. Automated semantic lung segmentation in chest CT images using deep neural network. *Neural Comput. Appl.* **35** (21), 15343–15364 (2023).
33. Ambesange, S., Annappa, B. & Koolagudi, S. G. Simulating federated transfer learning for lung segmentation using modified UNet model. *Procedia Comput. Sci.* **218**, 1485–1496 (2023).
34. Hasan, D. & Abdulazeez, A. M. Lung segmentation from chest X-Ray images using Deeplabv3plus-Based CNN model. *Indonesian J. Comput. Science*, **13**, 1, (2024).
35. Swaminathan, V. P. et al. GAN based image segmentation and classification using Vgg16 for prediction of lung cancer. *J. Adv. Res. Appl. Sci. Eng. Technol.* **35** (1), 45–61 (2024).
36. Suji, R. J., Godfrey, W. W. & Dhar, J. Exploring pretrained encoders for lung nodule segmentation task using LIDC-IDRI dataset. *Multimedia Tools Appl.* **83** (4), 9685–9708 (2024).
37. Thangavel, C. & Palanichamy, J. Effective deep learning approach for segmentation of pulmonary cancer in thoracic CT image. *Biomed. Signal Process. Control.* **89**, 105804 (2024).
38. Cai, J., Zhu, H., Liu, S., Qi, Y. & Chen, R. Lung image segmentation via generative adversarial networks. *Front. Physiol.* **15**, 1408832 (2024).
39. Ramos, L. & Pineda, I. A semiautomatic image Processing-Based method for binary segmentation of lungs in computed tomography images. *SN Comput. Sci.* **5** (6), 689 (2024).
40. Zheng, J., Wang, L., Gui, J. & Yussuf, A. H. Study on lung CT image segmentation algorithm based on threshold-gradient combination and improved convex hull method. *Sci. Rep.* **14** (1), 17731 (2024).
41. Guo, L., Liu, L., Zhao, Z. & Xia, X. An improved RIME optimization algorithm for lung cancer image segmentation. *Comput. Biol. Med.* **174**, 108219 (2024).
42. Pandey, S. K. & Bhandari, A. K. Morphological active contour based SVM model for lung cancer image segmentation. *Multimedia Tools Applications*, pp. 1–24, (2024).
43. Vijayakumar, S., Aarth, S., Deepa, D. & Suresh, P. Sustainable framework for automated segmentation and prediction of lung cancer in CT image using CapsNet with U-net segmentation. *Biomed. Signal Process. Control.* **99**, 106873 (2025).
44. Qiao, Z., Wu, L., Heidari, A. A., Zhao, X. & Chen, H. An enhanced tree-seed algorithm for global optimization and neural architecture search optimization in medical image segmentation. *Biomed. Signal Process. Control.* **104**, 107457 (2025).
45. Li, H., Ren, Z., Zhu, G. & Wang, J. AE-UNet: a composite lung CT image segmentation framework using attention mechanism and edge detection. *J. Supercomputing.* **81** (1), 331 (2025).
46. Murugaraj, S. S., Vadeivelu, K., Sambandam, P. T. & Kumar, B. S. Lung vessel segmentation and abnormality classification based on hybrid mobile-Lenet using CT image. *Biomed. Signal Process. Control.* **100**, 107072 (2025).
47. Fu, J., Peng, H., Li, B., Wang, J. & Liu, Z. LDN-SNP: SNP-based lightweight deep network for CT image segmentation of COVID-19. *Expert Syst. Appl.* **263**, 125793 (2025).
48. Erciyes, K., Soydan, M. & Gumus, O. Deep-Learning Based 3D Lung Segmentation, in *2025 13th International Symposium on Digital Forensics and Security (ISDFS)*, pp. 1–6: IEEE. (2025).
49. Yang, J. et al. A novel 3D lightweight model for COVID-19 lung CT lesion segmentation. *Medical Eng. & Physics*, p. 104297, (2025).
50. Nguyen, Q. H., Hoang, D. A. & Van Pham, H. Combination of 2D and 3D nnU-Net for ground glass opacity segmentation in CT images of Post-COVID-19 patients. *Comput. Biol. Med.* **195**, 110376 (2025).
51. Vincy, V. A. G., Byeon, H., Mahajan, D., Tonk, A. & Sunil, J. A 3D residual network-based approach for accurate lung nodule segmentation in CT images. *J. Radiation Res. Appl. Sci.* **18** (2), 101407 (2025).
52. Jannat, M. et al. Lung Segmentation with Lightweight Convolutional Attention Residual U-Net, *Diagnostics*, vol. 15, no. 7, p. 854, (2025).
53. Vijh, S., Gaurav, P. & Pandey, H. M. Hybrid bio-inspired algorithm and convolutional neural network for automatic lung tumor detection. *Neural Comput. Appl.* **35** (33), 23711–23724 (2023).
54. Alshayegi, M. H., Abed & Se Lung cancer classification and identification framework with automatic nodule segmentation screening using machine learning. *Appl. Intell.* **53** (16), 19724–19741 (2023).
55. Halder, A. & Dey, D. Atrous Convolution aided integrated framework for lung nodule segmentation and classification. *Biomed. Signal Process. Control.* **82**, 104527 (2023).
56. Venkatesh, C. et al. Deep Learning and Patch Processing Based Lung Cancer Detection on CT Images, in *International Conference on Communications and Cyber Physical Engineering*, 2024, pp. 575–590: Springer., 2024, pp. 575–590: Springer. (2018).
57. Sundarajan, M., Perumal, S., Sasikala, S., Ramachandran, M. & Pradeep, N. Lung Cancer Detection Using Explainable Artificial Intelligence in Medical Diagnosis, in *Advances in Explainable AI Applications for Smart Cities* pp. 352–370 (IGI Global, 2024).
58. Srinivasulu, A. et al. Lung malignant tumor data analytics using fusion ECNN and ERNN, in *Handbook of Artificial Intelligence Applications for Industrial Sustainability*: CRC, 47–63. (2024).

59. Tang, T. & Zhang, R. A Multi-Task model for pulmonary nodule segmentation and classification. *J. Imaging*. **10** (9), 234 (2024).
60. Kongkham, D., Beenarani, B., Nirmala, P., Kumaresan, S. J. & Senthilkumar, C. Deep Learning Approaches for Segmentation of Lung Nodules in CT Scans, in *2nd International Conference on Computer, Communication and Control (IC4)*, 2024, pp. 1–5: IEEE., 2024, pp. 1–5: IEEE. (2024).
61. Gautam, N., Basu, A. & Sarkar, R. Lung cancer detection from thoracic CT scans using an ensemble of deep learning models. *Neural Comput. Appl.* **36** (5), 2459–2477 (2024).
62. Sadremomtaz, A. & Zadnorouzi, M. Improving the quality of pulmonary nodules segmentation using the new proposed U-Net neural network. *Intelligence-Based Med.* **10**, 100166 (2024).
63. Pan, X. et al. LesionMix data enhancement and entropy minimization for semi-supervised lesion segmentation of lung cancer. *Appl. Soft Comput.* **167**, 112244 (2024).
64. Rathana, N. & Lokesh, S. Enhanced lung cancer diagnosis and staging with hrnet: A deep learning approach. *Int. J. Imaging Syst. Technol.* **34** (6), e23193 (2024).
65. Shi, X., Zhang, Z., MAST-UNet. & More adaptive semantic texture for segmenting pulmonary nodules. *Biomed. Signal Process. Control.* **99**, 106804 (2025).
66. Miao, S. et al. Deep learning-based CT image for pulmonary nodule classification with intrathoracic fat: A multicenter study. *Biomed. Signal Process. Control.* **100**, 106938 (2025).
67. Hajihosseini, M., Maghsoudi, A. & Ghezelbash, R. Regularization in machine learning models for MVT Pb-Zn prospectivity mapping: applying Lasso and elastic-net algorithms. *Earth Sci. Inf.* **17** (5), 4859–4873 (2024).
68. Yu, J. et al. ETMO-NAS: an efficient two-step multimodal one-shot NAS for lung nodules classification. *Biomed. Signal Process. Control.* **104**, 107479 (2025).
69. Jiang, C. et al. Kolmogorov–Arnold networks and multi slice partition channel priority attention in convolutional neural network for lung nodule detection. *Biomed. Signal Process. Control.* **103**, 107358 (2025).
70. Xue, X., Ma, Y., Du, W. & Peng, Y. Squeeze-and-excitation vision transformer for lung nodule classification. *IEEE Access*, (2025).
71. Elhassan, S. M., Darwish, S. M. & Elkaffas, S. M. An enhanced lung cancer detection approach using Dual-Model deep learning technique. *CMES-Computer Model. Eng. & Sciences*, **142**, 1, (2025).
72. Abdulqader, A. F. et al. Multi-objective deep learning for lung cancer detection in CT images: enhancements in tumor classification, localization, and diagnostic efficiency. *Discover Oncol.* **16** (1), 529 (2025).
73. Liu, Z., Wei, L. & Song, T. Optimized YOLOv11 model for lung nodule detection. *Biomed. Signal Process. Control.* **107**, 107830 (2025).
74. Li, Y., Hui, L., Wang, X., Zou, L. & Chua, S. Lung nodule detection using a multi-scale convolutional neural network and global channel Spatial attention mechanisms. *Sci. Rep.* **15** (1), 12313 (2025).
75. Pérez-López, A., Torres-Suárez, A. I., Martín-Sabroso, C. & Aparicio-Blanco, J. An overview of in vitro 3D models of the blood-brain barrier as a tool to predict the in vivo permeability of nanomedicines. *Advanced Drug Delivery Reviews*, p. 114816, (2023).
76. Lin, H. et al. Volumetric medical image segmentation via fully 3D adaptation of segment anything model. *Biocybernetics Biomedical Eng.* **45** (1), 1–10 (2025).
77. Zhu, Y., Zheng, F., Gong, Y., Yin, D. & Liu, Y. Biomechanical behavior of customized splint for the patient with temporomandibular disorders: A three-dimensional finite element analysis. *Biocybernetics Biomedical Eng.* **44** (1), 83–94 (2024).
78. Mehrnia, M. et al. Novel Self-Calibrated Threshold-Free probabilistic fibrosis signature technique for 3D late gadolinium enhancement MRI. *IEEE Trans. Biomedical Engineering*, (2024).
79. Ju, J. et al. Improving small intestinal stromal tumor detection using 3D context and auxiliary priori cues. *Biomed. Signal Process. Control.* **102**, 107231 (2025).
80. Saxton, S. H. & Stevens, K. R. 2D and 3D liver models. *J. Hepatol.* **78** (4), 873–875 (2023).
81. Nashtahosseini, Z., Aghamaali, M. R., Sadeghi, F., Heydari, N. & Parsian, H. Circulating status of MicroRNAs 660-5p and 210-3p in breast cancer patients. *J. Gene. Med.* **23** (4), e3320 (2021).
82. Oh, N. et al. Automated 3D liver segmentation from hepatobiliary phase MRI for enhanced preoperative planning. *Sci. Rep.* **13** (1), 17605 (2023).
83. Gu, C., Gao, H. & Combining GAN and LSTM models for 3D reconstruction of lung tumors from CT scans. *International J. Adv. Comput. Sci. Applications*, **14**, 5, (2023).
84. Rezaei, S. R. & Ahmadi, A. A GAN-based method for 3D lung tumor reconstruction boosted by a knowledge transfer approach. *Multimedia Tools Appl.* **82** (28), 44359–44385 (2023).
85. Fawzy, S., Moustafa, H. E. D., AbdelHay, E. H. & Ata, M. M. Proposed optimized active contour based approach for accurately skin lesion segmentation. *Multimedia Tools Appl.* **83** (2), 5745–5797 (2024).
86. Karrar, A., Mabrouk, M. S. & Abdel Wahed, M. Accurate quantification of small pulmonary nodules using 3D reconstruction of 2D computed tomography lung images. *J. Adv. Eng. Trends.* **42** (2), 1–15 (2023).
87. Dlamini, S., Chen, Y.-H. & Kuo, C.-F.-J. Complete fully automatic detection, segmentation and 3D reconstruction of tumor volume for non-small cell lung cancer using YOLOv4 and region-based active contour model. *Expert Syst. Appl.* **212**, 118661 (2023).
88. Shi, Z. et al. Reconstructing 3-D lung surfaces from a single 2-D chest x-ray image via vision transformer. *Med. Phys.* **51** (4), 2806–2816 (2024).
89. Huang, T., Yin, X. & Jiang, E. A novel 3-step technique for 3D tumor reconstruction using generative adversarial networks and an attention-based long short-term memory. *Res. Biomedical Eng.* **41** (2), 30 (2025).
90. Kordani, M. et al. Improving long-term flood forecasting accuracy using ensemble deep learning models and an attention mechanism. *J. Hydrol. Eng.* **29** (6), 04024042 (2024).
91. Khanmohammadidoustani, S., HassanzadehKermanshahi, K., Mohammadi, A. & Kermanshahi, S. Evaluating acceptance of a more strict plate control policy among motorcycle riders in Tehran. *Int. J. Transp. Eng.* **11** (1), 1301–1311 (2023).
92. Naderi, F. et al. Towards chemical source tracking and characterization using physics-informed neural networks. *Atmos. Environ.* **334**, 120679 (2024).
93. Moravvej, S. V., Mousavirad, S. J., Oliva, D., Schaefer, G. & Sobhaninia, Z. An improved de algorithm to optimise the learning process of a bert-based plagiarism detection model, in *2022 IEEE Congress on Evolutionary Computation (CEC)*, Padua, Italy, 1–7 : IEEE. (2022).
94. EskandariNasab, M., Raeisi, Z., Lashaki, R. A. & Najafi, H. A GRU–CNN model for auditory attention detection using microstate and recurrence quantification analysis. *Sci. Rep.* **14** (1), 8861 (2024).
95. Shakya, A. K., Pillai, G. & Chakrabarty, S. Reinforcement learning algorithms: A brief survey. *Expert Syst. Appl.* **231**, 120495 (2023).
96. Sivamayilvelan, K., Rajasekar, E., Vairavasundaram, S., Balachandran, S. & Suresh, V. Flexible recommendation for optimizing the debt collection process based on customer risk using deep reinforcement learning. *Expert Syst. Appl.* **256**, 124951 (2024).
97. Banar, S., Mohammadi, R., SeismoNet: & A proximal policy optimization-based earthquake early warning system using dilated Convolution layers and online data augmentation. *Expert Syst. Appl.* **253**, 124337 (2024).
98. McNeel, R. 10 September 2024). *Rhinoceros 3d*. (2024). Available: <https://www.rhino3d.com>

Acknowledgements

This research is supported by Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2025R897), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia.

Author contributions

Zhisen He: Conceptualization, methodology, study design, computational model implementation, manuscript writing and editing. Leila Jamel: Conceptualization, methodology development, study design, result interpretation, manuscript writing and editing. Danyi Huang: Methodology development, computational model implementation, performance evaluation, manuscript editing. Gaozhe Jiang: Methodology development, computational model implementation, data curation, validation, manuscript editing. Zaffar Ahmed Shaikh: Data curation, validation, study design, visualization, performance evaluation, manuscript editing. Khan, M. A. A.: Data curation, validation, performance evaluation. Seyed Jaleleddin Mousavirad: Methodology, manuscript writing and editing, result interpretation.

Funding

Open access funding provided by Mid Sweden University. Princess Nourah bint Abdulrahman University Researchers Supporting Project number PNURSP2025R897.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to S.J.M.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2025