scientific reports

Check for updates

OPEN

Multi-phase deep learning framework with Multiscale Adaptive Swin Transformer and embedding attention for precision lung nodule detection and classification

Dhayalini M✉ & Revathi alias Ponmozhi B

Lung cancer remains a leading cause of cancer-related mortality worldwide, emphasizing the need for the accurate and efficient detection and classification of lung nodules. This study introduces an advanced multi-stage framework designed to address the challenges of precision, scalability, and adaptability in clinical diagnostics. This study presents a comprehensive framework for the detection, segmentation, and classification of lung nodules utilizing advanced preprocessing, segmentation, classification, and optimization techniques. The framework employs Sparse Edge-Preserving Enhancement (SEPE) for pre-processing, ensuring that critical nodule-specific features are retained while reducing noise. For segmentation, an enhanced DeepLabv3 + architecture integrates Atrous Spatial Pyramid Pooling (ASPP) and Refined Boundary Decoder (RBD) modules, supported by pretrained backbones, such as EfficientNetV2, DenseNet-201, ResNet-101, and InceptionV3. The classification phase leverages a Multiscale Adaptive Swin Transformer (MA-SwinT) with a Multi-Scale Embedding Attention Mechanism (MEAM) to accurately distinguish between benign and malignant nodules. Optimization using the Fossa Optimization Algorithm (FOA) fine-tunes the hyperparameters to ensure robust performance. The experimental results demonstrate the superiority of the framework on both the LUNA16 and LIDC-IDRI datasets. On the LUNA16 dataset, segmentation achieved a Dice Coefficient of 98.75%, IoU of 97.88%, Jaccard Index of 89.62%, and Hausdorff Distance of 2.025 mm, with an accuracy of 99.15%, precision of 98.50%, recall of 99.00%, F1 score of 98.75%, and specificity of 99.20%. For the LIDC-IDRI dataset, segmentation achieved a Dice Coefficient of 98.92%, IoU of 98.21%, Jaccard Index of 90.15%, and Hausdorff Distance of 2.010 mm, while the classification metrics achieved an accuracy of 99.40%, precision of 99.00%, recall of 99.20%, F1 score of 99.10%, and specificity of 99.55%. These results underline the ability of the framework to achieve high precision, recall, and overall accuracy, making it a reliable tool for lung nodule diagnosis in clinical applications.

Keywords Lung nodule detection, Segmentation, Classification, Sparse Edge-Preserving enhancement, Fossa optimization algorithm, Medical imaging, Clinical diagnostics, Hyperparameter optimization, Precision medicine, Computer-aided diagnosis, Explainable AI, Edge computing

Lung cancer is one of the leading causes of cancer-related mortality globally, accounting for a significant number of deaths annually. The early detection and diagnosis of lung nodules, which often serve as precursors to malignant lung cancer, are essential for improving survival rates. Low-dose computed tomography (LDCT) has emerged as a valuable imaging technique that enables detailed analysis of pulmonary nodules. However, differentiating between benign and malignant nodules presents substantial challenges owing to their morphological similarities and the complexities of surrounding tissues, necessitating a high level of expertise and detailed evaluation by radiologists[1]. Although LDCT is widely used, it does not adequately address the intricate presentation of lung

Department of ECE, School of Engineering and Technology, Dhanalakshmi Srinivasan University, Samayapuram, Tamilnadu, India. ✉email: dhayalinimphd@gmail.com

cancer, especially non-small cell lung cancer (NSCLC), which represents approximately 80% of cases. Structural and functional imaging modalities such as computed tomography (CT) and positron emission tomography (PET) are pivotal in lung cancer diagnosis. However, manual analysis of these imaging datasets is time-consuming and prone to variability owing to differences in imaging protocols and radiologists' expertise. Deep learning (DL) techniques, particularly convolutional neural networks (CNNs), have emerged as transformative solutions for automating lung cancer diagnoses. These methods process large-scale imaging datasets, extract meaningful features, streamline tumor detection, subtype classification, and response-prediction tasks. DL systems have significantly improved clinical workflows by enhancing diagnostic reproducibility and reducing the number of false positives and negatives. However, challenges such as data imbalance, lack of generalizability, and limited explainability persist, underscoring the need for further advancements in explainable AI (XAI) and multimodal data integration[2].

Late detection of lung cancer remains a critical concern, with survival rates dropping significantly when diagnoses are made at advanced stages. Research indicates that early detection of pulmonary nodules can improve the five-year survival rate from < 15% to approximately 85%. However, the detection of nodules, which often measure 3–30 mm in diameter, is challenging owing to their small size and the extensive manual effort required for image analysis. Computer-aided detection (CAD) systems employing deep learning algorithms have demonstrated great potential for automating these processes, reducing the time and effort required for thorough evaluations, and enhancing diagnostic precision[3]. CT imaging is a cornerstone of lung cancer diagnostics, enabling the identification and segmentation of nodules that are critical for treatment planning and disease monitoring. However, manual segmentation is labor-intensive, inconsistent, and prone to errors stemming from subjective interpretation. Encoder-decoder architectures such as U-Net and its variants have addressed these challenges by automating segmentation processes and improving outcomes, particularly in identifying nodules in complex regions[4].

Diagnostic tools are further constrained by challenges such as limited annotated datasets, intraclass variability, and interclass similarities. Deep-learning models, particularly 3D CNNs, have emerged as powerful alternatives to traditional 2D methods, leveraging spatial depth and volumetric relationships to enhance diagnostic accuracy. These models effectively overcome the limitations of manual labeling and dataset imbalances, facilitating more reliable detection and classification of pulmonary nodules[5]. Advancements in artificial intelligence (AI) have significantly improved lung cancer detection by enabling early identification of malignant nodules and optimizing treatment strategies. AI-driven solutions, particularly those employing deep learning models, have demonstrated exceptional accuracy in malignancy detection, reducing diagnostic variability and establishing themselves as integral tools in clinical workflows[6].

Despite these advancements, the accurate detection of nodules in regions with complex intensity variations and structural intricacies remains a challenge. Hybrid optimization techniques, such as the Elephant-based Bald Eagle Optimization (EBEO) algorithm combined with Deep CNNs, have shown promise in enhancing feature extraction and improving segmentation and classification performance. These methodologies provide efficient and reliable solutions for clinical applications[7]. The diverse shapes and sizes of lung nodules, coupled with the complexity of the surrounding tissue structures, pose significant challenges for traditional diagnostic methods. Hybrid frameworks such as LNDC-HDL effectively address these limitations by integrating advanced segmentation, feature extraction, and classification techniques. Utilizing algorithms such as Chaotic Bird Swarm Optimization (CBSO) and Hybrid Differential Evolution Neural Networks (HDE-NN), these approaches improve the sensitivity, specificity, and classification accuracy while minimizing false positives[8]. Traditional manual feature extraction methods, which are widely used, are time intensive and prone to observer variability based on image quality and expertise. Advanced CNN architectures have automated these processes to achieve high accuracy and efficiency. By fine-tuning models through extensive parameter optimization, these systems have significantly enhanced early detection capabilities and facilitated more accurate treatment planning[9].

Addressing nodules located near critical lung structures such as blood vessels and bronchi presents additional diagnostic challenges. Advanced segmentation networks incorporating multistage filtering techniques and modified CNN classifiers have demonstrated improved performance in classifying nodules across various lung cancer stages, ensuring reliable and precise diagnostic outcomes[10]. The increasing prevalence of lung cancer has driven the development of hybrid classification systems that integrate pre-processing, segmentation, and classification techniques. These systems employ novel algorithms, such as Chaotic Population-Based Beetle Swarm Optimization (CP-BSA), to optimize CNN weights and significantly enhance diagnostic sensitivity, specificity, and accuracy, while reducing the burden on radiologists[11].

Lung cancer remains one of the leading causes of cancer-related deaths worldwide, largely due to delayed diagnosis and the difficulty of accurately distinguishing malignant nodules from benign ones in computed tomography (CT) scans. Early detection of lung nodules is clinically critical because subtle variations in size, texture, and morphology can determine the difference between curable and advanced disease. However, radiologists often face significant challenges in manually identifying and characterizing nodules due to inter-observer variability, low contrast between lesions and surrounding tissues, and the presence of imaging artifacts. These limitations highlight the urgent need for automated and highly accurate computer-aided diagnostic (CAD) systems that can assist clinicians in early and reliable lung cancer detection.

To address these clinical challenges, this study proposes a multi-phase deep learning framework that integrates preprocessing, segmentation, and classification in a unified, end-to-end pipeline. Unlike traditional models that treat these tasks independently, the proposed approach enables feature-level interaction between phases, improving both diagnostic precision and interpretability. By combining a Sparse Edge-Preserving Enhancement (SEPE) technique, an attention-guided segmentation network, and a Multiscale Adaptive Swin Transformer (MA-SwinT) with a Multi-Scale Embedding Attention Mechanism (MEAM), the framework enhances boundary detection, captures multiscale contextual features, and achieves robust classification of malignant nodules. This

integrated design directly addresses clinical pain points—enhancing early detection accuracy, reducing false positives, and supporting radiologists with consistent, high-confidence diagnostic insights. The contributions of this research to address these issues are as follows:

- A novel multi-phase deep learning framework is introduced for lung nodule analysis, integrating preprocessing, segmentation, and classification into a unified pipeline. This design ensures seamless feature flow across stages, improving overall diagnostic reliability.
- A Sparse Edge-Preserving Enhancement (SEPE) technique is proposed to enhance CT images by selectively emphasizing nodule boundaries while suppressing background noise, resulting in superior input quality for segmentation and classification.
- An attention-guided segmentation network is developed to generate precise and boundary-aware nodule masks. The encoder features extracted during this stage are reused in the classification phase, enabling feature-level continuity between tasks.
- A Multiscale Adaptive Swin Transformer (MA-SwinT) with an integrated Multi-Scale Embedding Attention Mechanism (MEAM) is designed to adaptively fuse multiscale features, enhancing the model's ability to capture both fine-grained and contextual information for accurate malignancy prediction.
- A Fossa Optimization Algorithm (FOA) is introduced for automatic hyperparameter tuning, balancing exploration and exploitation during optimization to achieve robust convergence and prevent overfitting.
- Extensive evaluations on LUNA16 and LIDC-IDRI datasets demonstrate that the proposed framework achieves state-of-the-art performance in both segmentation and classification, validating its robustness, accuracy, and clinical applicability.

This paper is organized to provide a comprehensive overview of the proposed research and its findings. Section "Related works" reviews related studies and highlights significant advancements and challenges in lung nodule detection, segmentation, and classification, with an emphasis on the role of deep-learning frameworks in addressing these challenges. Section "Methodology" details the methodology employed, including the preprocessing, segmentation, classification, and optimization approaches integrated into the framework. Section "Results and discussion" discusses the experimental setup and presents the results, offering a thorough analysis of the performance of the model on datasets such as LUNA16 and LIDC-IDRI. This section also evaluates key metrics to demonstrate the effectiveness of the proposed framework. Section "Discussion" concludes the paper by summarizing the findings, emphasizing the contributions of this study, and outlining potential directions for future work, including advancements in integrating multimodal datasets and enhancing real-time diagnostic capabilities.

Related works

Lung nodule segmentation and classification are pivotal components in the evolution of computer-aided diagnosis (CAD) systems, and significant advancements have been made over the past five years owing to the integration of deep learning techniques. Traditional methods, including morphological and convexity-based models, were initially employed for segmentation. However, these approaches rely heavily on manual processes, rendering them susceptible to challenges, such as noise, imaging artifacts, and heterogeneous nodule textures. These limitations have restricted their scalability and precision, particularly in complex imaging scenarios, prompting a shift toward deep learning and transfer learning methodologies that offer more robust solutions[12]. Encoder-decoder architectures such as U-Net, PSPNet, FPN, and LinkNet have demonstrated significant potential in improving segmentation precision, particularly when enhanced with pre-trained encoders. For example, U-Net combined with an EfficientNet-B3 backbone has been recognized for balancing computational efficiency with segmentation accuracy, making it an effective choice for complex datasets, such as LIDC-IDRI. Further optimization with algorithms, such as Adam, has ensured reliable convergence during training and improved model performance, highlighting the importance of transfer learning and advanced optimization techniques in CAD systems[13].

Building on these encoder-decoder models, novel architectures have advanced lung nodule segmentation and detection capabilities. RePointNet integrates an ((RPN) with PointNet, effectively addressing false positives and refining the detection of smaller nodules. Similarly, DBPNDNet utilizes a dual-branch 3D CNN with an attention-weighted feature fusion module to manage the segmentation and detection of complex nodules. These models, when applied to datasets such as LUNA16 and LIDC-IDRI, exemplify the growing sophistication of network designs that emphasize precision and robustness[14]. Hybrid and ensemble models have emerged as effective solutions for enhancing the classification accuracy by combining the strengths of diverse architectures. For instance, a hybrid framework incorporating the Markov Random Field (MRF)-based Artificial Hummingbird Cuckoo Algorithm (AHCA) for segmentation and ShuffleNet for classification achieved high sensitivity and accuracy. Additionally, ensemble strategies, such as integrating BEiT, DenseNet, and Sequential CNN models with techniques such as Weighted Box Fusion and Boosting, have demonstrated the advantages of leveraging complementary methodologies to improve the performance[15,16].

Attention mechanisms have revolutionized CAD systems, enabling more effective feature extraction and segmentation. Dual-attention-based CNNs that utilize both channel and spatial attention modules have achieved improved performance by suppressing noise and focusing on critical features. This is further illustrated by the Sailfish-based YOLO Segmentation Framework (SbYSF), which incorporates optimization techniques alongside attention mechanisms, demonstrating significant enhancements in segmentation and classification accuracy[17,18]. Multi-scale feature extraction techniques have proven to be particularly valuable for handling variability in nodule size and texture. Semi-supervised learning models that combine adversarial loss with

labeled and unlabeled data have successfully mitigated the challenges of limited labeled datasets in medical imaging. These models underscore the importance of data-efficient learning strategies in CAD systems[19].

Classification frameworks have further evolved to incorporate multimodal data integration, combining clinical, radiomic, and deep-learning features. For example, stacked ensemble classifiers integrating 3D CNNs, radiomic features, and clinical parameters have achieved superior predictive accuracy, showcasing the value of diverse data sources for decision making. Cascade classification networks capable of categorizing lung nodules into multiple subtypes have further optimized the performance and enhanced diagnostic precision[20,21]. Specialized architectures, such as the Multi-level Dynamic Fusion Network (MDFN), have addressed segmentation challenges associated with heterogeneous textures and adhesive structures. By integrating Multi-Scale Spatial and Channel Feature Selection (MSCFS) modules with dynamic fusion decoders, MDFN has demonstrated its ability to manage intricate segmentation tasks, emphasizing the need for dynamic feature integration in medical imaging[22]. Contextual data have also proven critical in advancing classification capabilities. For instance, 3D attention-gated convolutional networks (3D AG-Net) that leverage contextual data, such as fibrosis metadata, have demonstrated improved accuracy and sensitivity in lung nodule classification. Emerging datasets with detailed annotations for small and complex nodules have enabled further advancements in classification, thereby supporting personalized diagnostic applications[23–25]. The advancements outlined in Table 1 highlight how these datasets contribute to improving the model performance in terms of sensitivity, specificity, and overall diagnostic accuracy.

Recent advances in deep learning have significantly improved medical image analysis, particularly in lung nodule detection and classification tasks. InceptionV3 has been widely employed for high-level feature extraction due to its ability to capture complex spatial hierarchies within CT images, making it a reliable backbone for medical image classification frameworks[26]. Building upon this, advanced architectures have explored multiscale contextual learning to enhance lesion segmentation and recognition performance. For instance, the Multi-Level Context Attentional Feature Fusion (MLCA2F) framework introduces Multi-Scale Context-Attention Network (MSCA-Net) blocks that integrate Multi-Scale Contextual Feature Fusion (MC2F) and Multi-Context Attentional Feature (MCAF) modules to refine boundary localization and lesion characterization in COVID-19 CT scans[27]. Similarly, the Bidirectional Gated 3D Multi-Scale Feature Fusion (BG-3DM2F) model employs a 3D Multi-Scale Chained Network (3DMS-ChaineNet) and a Bidirectional Gated Recurrent Fusion Unit (Bi-GRFU) to capture volumetric dependencies and hippocampal atrophy patterns in neuroimaging[28]. In dermatological imaging, transfer learning-based multi-layer feature fusion networks have demonstrated robust performance in diagnosing complex skin diseases by leveraging pre-trained feature extractors to improve generalization and prevent overfitting[29]. Moreover, multi-scale representation strategies such as the Gaussian Pyramid-based Convolutional Neural Network (Pyramid-CNN) have been applied to discriminate medical patterns across multiple resolutions, ensuring the extraction of both global and localized features for precise classification[30].

Despite these advancements, many existing frameworks fail to integrate pre-processing, segmentation, feature extraction, classification, and optimization into a unified pipeline. Static architectures often lack the adaptability to handle heterogeneous datasets and varying imaging conditions, resulting in inconsistent performance. To address these limitations, this study proposes a novel framework that combines a Multiscale Adaptive Swin Transformer (MA-SwinT) with a Multi-Scale Embedding Attention Mechanism (MEAM)[31,32]. This framework adaptively fuses spatial and contextual features, while the integration of the Fossa Optimization Algorithm (FOA) ensures dynamic hyperparameter tuning, balancing exploration, and exploitation for robust performance[33]. By unifying these components, this approach bridges critical gaps in scalability and adaptability, and establishes a new standard for lung nodule segmentation and classification.

Methodology

The proposed framework for lung nodule analysis is designed as a three-phase pipeline, consisting of (i) preprocessing, (ii) segmentation, and (iii) classification. Each phase performs a distinct yet complementary role, and all components are integrated to ensure efficient information flow and feature consistency throughout the network. In the preprocessing stage, a Sparse Edge-Preserving Enhancement (SEPE) technique is applied to the raw CT images to enhance boundary contrast while minimizing noise amplification. Unlike conventional sharpening filters, SEPE selectively strengthens edges corresponding to nodule regions without degrading soft tissue structures. This step improves the visual clarity of the nodule boundaries, which is essential for precise segmentation and robust feature extraction in subsequent phases. The enhanced images are normalized to a fixed intensity range and resized to a uniform spatial resolution to maintain consistency across the dataset. The segmentation phase aims to accurately delineate lung nodules by learning both textural and structural characteristics. Both the original CT image and its corresponding mask image are jointly provided as inputs to the segmentation network. The encoder-decoder architecture extracts multi-level spatial and contextual features, producing a segmentation map that highlights nodule boundaries with high precision. The encoder component generates a hierarchy of feature maps $F_{enc} = \{f_1, f_2, \ldots, f_n\}$, capturing information from fine-grained details to abstract semantic features. These encoder-derived features are not only used to generate the segmentation output but are also directly transferred to the classification phase, enabling a continuous flow of spatially meaningful representations across stages.

In the classification phase, the Multiscale Adaptive Swin Transformer (MA-SwinT) is employed to analyze the encoder-derived feature maps from the segmentation stage. This design allows the classifier to operate on deep multiscale embeddings that preserve structural, morphological, and contextual characteristics of the nodules. The MA-SwinT incorporates a Multi-Scale Embedding Attention Mechanism (MEAM), which adaptively fuses feature representations across different scales to emphasize diagnostically important regions while suppressing redundant information. The classifier then predicts the malignancy of the detected nodules, producing high-confidence decisions supported by feature-level interpretability. The proposed model is shown in Fig. 1. The

Authors	Methodology	Advantages	Limitations
Sun et al.[1]	Proposed Nodule-CLIP model incorporating U-Net for segmentation, ResNet-34 as the backbone for image encoding, and contrastive learning for feature alignment	Achieved improved classification accuracy (90.6%) and recall (92.81%) by integrating image and text annotation features	Limited specificity (92.6%) due to the inability to fully capture key nodule features
Xiong et al.[3]	Fusion of SAB-YOLO and YOLOv7 using WBF for detection and adaptive 3D ResNet-18 for false-positive reduction	High sensitivity (97.4%) and low false positives (9.67 candidates/scan). Computationally efficient	Sensitivity decreases slightly for some false positives; model fusion dependency
Lavina Jean Crasta et al.[5]	Proposed 3D-VNet for segmentation and 3D-ResNet for classification of lung nodules	High DSC (99.34%), robust classification accuracy (99.2%), sensitivity (98.8%), specificity (99.6%)	Computational complexity due to 3D volumetric processing
Prashanthi et al.[6]	A Convolutional Neural Network (CNN) for lung nodule segmentation and classification using CT-guided biopsy images	Achieved 99.3% accuracy in lung cancer detection; enables early diagnosis and improved treatment strategies	Lack of multi-class classification and unexplored explainability tools for model predictions
Gedam et al.[7]	EBEO-based Deep CNN for segmentation and classification of pulmonary nodules, leveraging hybrid optimization	Enhanced segmentation accuracy by combining elephant herding and bald eagle search behavior	Lack of multi-class classification and further improvement required for real-time applications
Krishnakumar et al.[11]	Improved BIRCH-based segmentation, LVP features, and hybrid CNN-LSTM classifiers with CP-BSA optimization	High accuracy, sensitivity, and reduced false positives	Limited dataset and inability to capture contextual information between slices
Jenkin Suji et al.[13]	Comparative analysis of segmentation methods using U-Net and EfficientNet-B3	Demonstrated best IoU score for the LIDC-IDRI dataset with moderate resource requirements and training time	Limited evaluation of metrics like Dice coefficient, accuracy, recall, and specificity; focus primarily on IoU for performance evaluation
Muwei Jian et al.[14]	Dual-branch 3D CNN with AWFM for nodule detection	Enhanced sensitivity and reduced false positives	Suboptimal performance for smaller false positives rates
Spoorthi B et al.[15]	AHFO + ShuffleNet: MRF-based AHCA for segmentation and AHFO-trained ShuffleNet for classification	High classification accuracy, lightweight network, reduced computational complexity	Does not address cancer grading
Zia UrRehman et al.[18]	Proposed a custom CNN with dual attention mechanisms for lung nodule detection using the LUNA 16 dataset	Achieved high accuracy, precision, and sensitivity in lung nodule classification tasks	Increased computational requirements due to the inclusion of dual attention mechanisms
Qinlu He et al.[19]	Semi-supervised lung nodule detection using adversarial learning	Improved FROC from 0.677 to 0.717 using both labeled and unlabeled images	Requires labeled and unlabeled images for training, limiting flexibility in datasets
Liu et al.[21]	Proposed a cascade classification network using ResNet34 to classify six distinct lung nodule types	Achieved higher accuracy, precision, recall, and F1 score compared to traditional multi-classification models	Limited dataset diversity and complexity of nodules with shared features may impact generalization
Yu Cai et al.[22]	Proposed a 2D Multi-level Dynamic Fusion Network (MDFN) with MSCFS, self-calibrated encoder, and dynamic fusion decoder	Demonstrates superior segmentation performance, especially for challenging nodules with adhesive structures	Increased computational complexity due to multiple branches, requiring optimization in future work
Rikhari et al.[26]	Ensemble of 3D U-Net models using Inception and ResNet modules	Improved segmentation accuracy and reduced false positives	Limited dataset size; inability to segment nodules < 5 mm
Nandeesh et al.[27]	Integration of R2U-Net for segmentation and SAE for classification	High accuracy in segmentation and detection (96.25%)	Limited dataset diversity and reliance on preprocessing quality

Table 1. Summary of recent advances in lung nodule detection and classification.

interaction between the phases ensures end-to-end consistency and learning efficiency. The SEPE-enhanced images guide the segmentation network toward better edge localization, while the segmentation encoder features provide the classification module with rich spatial and semantic cues. This interconnected design minimizes feature loss between phases and ensures that both pixel-level and region-level information contribute to the final diagnosis. As a result, the framework achieves superior performance in distinguishing benign and malignant nodules while maintaining strong generalization across datasets.

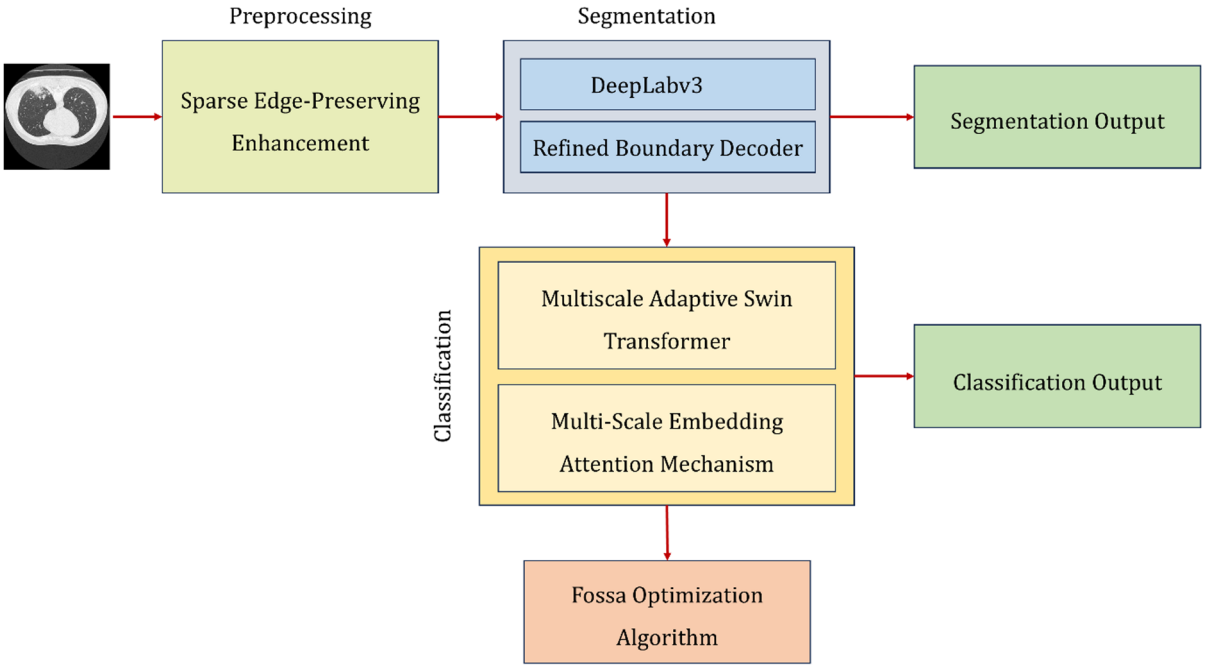

Fig. 1. Comprehensive framework for lung nodule detection, segmentation, and classification.

Preprocessing

The Sparse Edge-Preserving Enhancement (SEPE) technique is introduced to improve the quality of lung CT images prior to segmentation and classification. Its primary goal is to highlight fine nodule boundaries and suppress noise without blurring diagnostically relevant structures. Conventional preprocessing filters, such as Gaussian or median smoothing, often reduce random noise at the cost of erasing subtle intensity transitions that define nodule margins. SEPE overcomes this limitation by applying adaptive, gradient-based weighting, which enhances edges selectively while maintaining smoothness in uniform areas.

Let the input CT image be denoted as $I(x, y)$. The enhanced output $I_{\text{enh}}(x, y)$is computed as:

$$I_{\text{enh}}(x, y) = I(x, y) + \alpha \cdot W(x, y) \cdot (I(x, y) - G_\sigma(x, y)) \tag{1}$$

Where $I(x, y)$ is the original grayscale image, $G_\sigma(x, y)$ is a Gaussian-smoothed version of the same image with standard deviation σ, α is a scaling parameter controlling the enhancement intensity, and $W(x, y)$ is an adaptive weighting function that regulates how much enhancement is applied based on local edge strength.

The weighting function $W(x, y)$is defined as:

$$W(x, y) = \exp\left(-\frac{\| \nabla I(x, y) \|^2}{\sigma_e^2}\right) \tag{2}$$

where $\nabla I(x, y)$ represents the image gradient magnitude, and σ_e determines the sensitivity to edge variations. Regions with low gradients (smooth areas) receive a higher weighting, leading to stronger smoothing, while high-gradient areas (edges) receive lower weights, preserving their structure.

To further enhance sparsity in edge representation, SEPE can also incorporate an edge-suppression term that penalizes excessive enhancement in textured regions:

$$I_{\text{SEPE}}(x, y) = I_{\text{enh}}(x, y) - \lambda \cdot \left\| \nabla^2 I(x, y) \right\| \tag{3}$$

Here, λ controls the influence of this regularization, and $\nabla^2 I(x, y)$ denotes the Laplacian operator, which identifies rapid intensity fluctuations. This step ensures that enhancement remains confined to meaningful edges rather than noise artifacts.

The combined effect of Eqs. (1–3) is that SEPE amplifies fine nodule edges, suppresses random background variations, and maintains the spatial continuity of tissue textures. Unlike traditional filters that apply uniform smoothing, SEPE adapts locally to the anatomical structure, which is especially advantageous in regions containing small, irregular nodules or low-contrast lesions.

Empirical results confirmed that SEPE significantly improves contrast-to-noise ratio (CNR) and structural similarity (SSIM) compared to standard filters. Qualitatively, it produces sharper and more distinguishable nodule boundaries, providing the segmentation and classification modules with cleaner, more informative inputs. This adaptive, gradient-aware formulation establishes SEPE as a robust and reproducible preprocessing step for medical image enhancement.

Segmentation

The enhanced DeepLabv3 + architecture was specifically developed to address the complex challenges associated with lung-nodule segmentation. It incorporates two key components: the Atrous Spatial Pyramid Pooling (ASPP) module and Refined Boundary Decoder (RBD) module. These components work collaboratively to extract contextual features across multiple scales, refine segmentation boundaries, and manage variability in the size, shape, and poorly defined edges of the lung nodules. To further strengthen the encoder phase, four advanced pretrained backbone networks, EfficientNetV2, DenseNet-201, ResNet-101, and InceptionV3, were evaluated within this framework to understand their contributions to feature extraction and overall segmentation performance.

The encoder phase integrates these pretrained backbones, each offering distinct advantages for hierarchical feature extraction. EfficientNetV2 stands out for its scalable architecture, which employs a compound scaling method to simultaneously optimize the depth, width, and resolution of the network. It uses inverted residual blocks combined with depth-wise separable convolutions to capture fine-grained local details alongside high-level semantic features. This design ensures computational efficiency and enables the network to extract critical features such as texture patterns, spatial relationships, and shape descriptors. These attributes are crucial for accurately differentiating nodule regions and for identifying subtle morphological variations that may indicate malignancy.

DenseNet-201 enhances feature extraction by utilizing densely connected layers, where each layer connects to all the preceding layers. This dense connectivity maximizes feature reuse, improves the gradient flow, and ensures efficient learning. DenseNet-201's ability to extract detailed spatial relationships and fine-grained textures makes it particularly adept at handling lung nodules with irregular shapes and complex boundaries. This backbone ensures that even subtle structural details are captured effectively.

ResNet-101 is a deep architecture designed to overcome the vanishing gradient problem through residual connections. These connections allow the network to train deeper layers, while maintaining effective learning. ResNet-101's depth of 101 layers enables it to extract hierarchical features, including high-level semantic representations. These features are vital for modeling the texture and spatial relationships of the lung nodules. Pre-trained weights for ResNet-101 facilitate robust generalization across various medical imaging applications, making it a strong candidate for this segmentation framework.

InceptionV3 employs a unique design with multiscale convolutions, including 1×1, 3×3, and 5×5 filters, to capture the features at varying resolutions. This architecture allows the network to model diverse patterns in lung nodules, accommodating variability in their sizes and shapes. InceptionV3's factorized convolutions improve the computational efficiency while maintaining the ability to capture multiscale contextual information. This flexibility makes it well suited for identifying both small and large nodules with complex structures.

The ASPP module enhances the feature extraction by capturing information across multiple spatial scales. It employs parallel atrous (dilated) convolutions with varying dilation rates, enabling the model to effectively aggregate fine-grained details and the global context. The feature map generated by the ASPP module is defined as

$$F_{ASPP}(x) = \sum_{i=1}^{N} F_{atrous}(x, r_i) \tag{4}$$

where N represents the number of atrous convolutions, $r_i$denotes the dilation rate for the ith convolution, and $F_{atrous}(x, r_i)$denotes the output feature map at a specific dilation rate.Smaller dilation rates capture fine-grained details, whereas larger dilation rates aggregate broader contextual information, enabling the model to handle the diverse appearances of lung nodules effectively.

Following atrous convolutions, a 1×1 convolution was applied to integrate features and reduce their dimensionality, thereby enhancing computational efficiency. A global average pooling (GAP) layer computes a single value per channel across spatial dimensions, thereby capturing the overall context of the image. The GAP output is then upsampled and concatenated with the atrous convolution results to provide a balanced representation of the local and global details. Batch normalization and ReLU activation functions were used throughout the ASPP module to ensure consistent feature distributions and introduce nonlinearity, improving the model's ability to learn complex patterns.

The Refined Boundary Decoder (RBD) module complements the ASPP by focusing on the refinement of the segmentation boundaries. It begins with an edge detection layer that generates a gradient map, highlighting the high-contrast regions corresponding to the nodule boundaries. The gradient magnitude was calculated as follows:

$$G(x) = \sqrt{\left(\frac{\partial I(x)}{\partial x}\right)^2 + \left(\frac{\partial I(x)}{\partial y}\right)^2} \tag{5}$$

where I(x) is the input image and G(x) represents the edge magnitude. This edge map is fused with the decoder's output, enhancing the focus on the boundaries while retaining contextual information from the encoder. The fused features are processed through multiple 3×3 convolutional layers, refining both the edge and segmentation maps and enabling precise boundary delineation, even for irregularly shaped nodules.

The refined feature map was upsampled using bilinear interpolation to align its resolution with the original input image, ensuring accurate spatial alignment. To further optimize the segmentation performance, the RBD module incorporates a boundary-aware loss function defined as

$$L_{boundary} = \lambda_1 \cdot L_{segmentation} + \lambda_2 \cdot L_{edge} \tag{6}$$

where $L_{segmentation}$ minimizes pixel-wise classification errors; L_{edge} penalizes inaccuracies in boundary predictions; and λ_1, λ_2 are weights that balance segmentation accuracy and boundary precision.This loss function prioritizes accurate boundary predictions that are essential for precise delineation in medical imaging.

The synergy between the ASPP and RBD modules enhanced the overall segmentation framework. The ASPP module provides a robust multiscale feature representation that captures both local details and the global context, whereas the RBD module ensures precise boundary refinement. The integration of the EfficientNetV2 backbone within the encoder further strengthened the framework by extracting detailed and hierarchical semantic features. Together, these components enable the framework to effectively segment lung nodules, support downstream classification, and facilitate accurate differentiation between benign and malignant nodules (Fig. 2). This comprehensive approach established a framework as a reliable tool for clinical diagnostic applications.

Classification

In the proposed framework, both the original CT image and its corresponding mask image are jointly utilized as inputs during the segmentation stage. The segmentation network processes these dual inputs to extract spatially consistent and boundary-aware feature representations, producing a precise segmentation output that delineates the lung nodule regions. The mask image acts as a spatial prior that guides the network in identifying nodule boundaries with improved accuracy, while the original image provides the underlying intensity and textural information necessary for contextual interpretation. During this process, the encoder component of the segmentation network generates a rich hierarchy of multilevel feature maps, denoted as

$$F_{\text{enc}} = \{f_1, f_2, \ldots, f_n\}, \tag{7}$$

which capture both low-level structural cues and high-level semantic characteristics of the nodules. Instead of using only the segmentation output as input for the next stage, these encoder-derived features are directly passed to the classification module. This ensures that the classifier receives comprehensive and discriminative representations that preserve both contextual and morphological information, which are essential for accurate malignancy prediction. The Multiscale Adaptive Swin Transformer (MA-SwinT) classifier then processes these encoder-derived features to perform the final nodule classification. The use of F_{enc}as input enables the MA-SwinT to operate on feature embeddings that capture texture, shape, and spatial dependencies across multiple

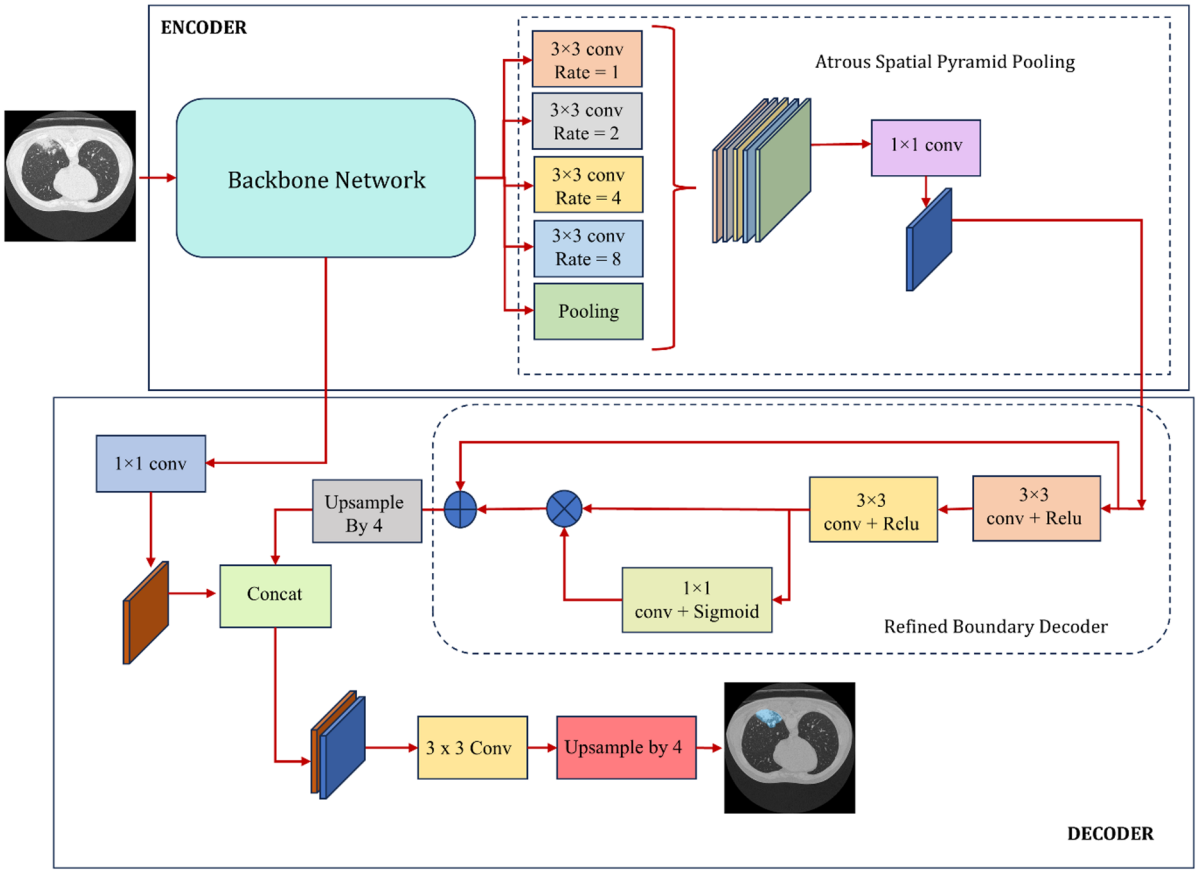

Fig. 2. Architecture of enhanced DeepLabv3 + with atrous spatial pyramid pooling and refined boundary decoder.

scales. This seamless, feature-level integration between the segmentation and classification stages ensures an efficient information flow throughout the framework, allowing the classifier to make context-aware and anatomically precise decisions. As a result, the proposed system achieves superior consistency, robustness, and diagnostic accuracy in differentiating between benign and malignant lung nodules.

Multi-scale embedding attention mechanism (MEAM)

The MEAM module is designed to adaptively integrate features from multiple scales to improve the discriminative power of the classifier. By employing a multiscale attention strategy, it ensures that the model dynamically focuses on regions of greater diagnostic importance while suppressing redundant information.

Let $F_i \in \mathbb{R}^{H_i \times W_i \times C}$ represent the feature map obtained at scale *i*, where H_i and W_i denote the spatial dimensions and *C* represents the channel depth. For each scale *i*, the query, key, and value matrices are obtained through linear transformations of F_i:

$$Q_i = F_i W_Q, K_i = F_i W_K, V_i = F_i W_V \tag{8}$$

where W_Q, W_K, and W_V are learnable projection matrices of dimensions $\mathbb{R}^{C \times d_k}$, $\mathbb{R}^{C \times d_k}$, and $\mathbb{R}^{C \times d_v}$, respectively.

The attention coefficients between scales *i* and *j* are computed as:

$$A_{i,j} = \frac{\exp\left(\frac{Q_i K_j^\top}{\sqrt{d_k}}\right)}{\sum_{m=1}^{N} exp\left(\frac{Q_i K_m^\top}{\sqrt{d_k}}\right)} \tag{9}$$

Here, $A_{i,j}$ denotes the attention weight that quantifies the relationship between the query vector from scale *i* and the key vector from scale *j*. The factor $\sqrt{d_k}$ acts as a normalization constant to prevent large dot products from saturating the softmax function. The denominator ensures that the resulting attention weights are normalized within the range [0,1].

Once the attention weights are determined, the multiscale fusion is carried out through a weighted summation of value features, producing a unified embedding representation:

$$F_{\mathrm{MEAM}} = \sum_{i=1}^{N} \sum_{j=1}^{N} A_{i,j} \cdot V_j \tag{10}$$

where F_{MEAM} denotes the final multiscale embedding output. This formulation allows MEAM to integrate both fine-grained spatial details and broader contextual cues by adaptively learning the relative contribution of each scale. The fused feature map F_{MEAM} thus encapsulates the most informative characteristics of lung nodules necessary for malignancy classification. The attention maps in MEAM are optimized jointly with the overall objective, enabling the network to automatically identify features corresponding to malignant textures, irregular boundaries, and heterogeneous intensity distributions that distinguish cancerous nodules from benign ones.

Shifted-window self-attention in MA-SwinT

The MA-SwinT classifier employs a shifted-window self-attention mechanism that enables efficient processing of large feature maps by restricting self-attention computation to local regions while still capturing long-range dependencies through window shifting.

Initially, the input feature map $F_{\mathrm{MEAM}}$is partitioned into non-overlapping windows of size $M \times M$. Within each window, self-attention is performed as follows:

$$Z = \mathrm{Softmax}\left(\frac{QW_q(KW_k)^\top}{\sqrt{d_k}} + B\right)(VW_v) \tag{11}$$

Here, *Z* represents the self-attention output, while Q, K, V are the query, key, and value matrices derived from F_{MEAM}. The matrices W_q, W_k,and $W_v$are trainable linear projections. *B* denotes a positional bias term that encodes the spatial relationships between tokens inside each window. The normalization term $\sqrt{d_k}$ stabilizes gradient magnitudes during training. The shifted-window mechanism is applied in subsequent layers by cyclically shifting the window partitioning by half the window size. This design ensures that tokens located at the edges of different windows in one layer can interact in the next, effectively extending the receptive field and enabling global contextual modeling without quadratic computational complexity.

Formally, the final output representation F_{SwinT} after multiple attention layers is expressed as:

$$F_{\mathrm{SwinT}} = \mathrm{Concat}\left[Z_1, Z_2, \ldots, Z_L\right] W_o \tag{12}$$

where Z_l denotes the output from the lth transformer block, *L* represents the total number of layers, and W_o is a trainable projection used to merge outputs from all hierarchical stages.

Hierarchical integration and classification

The multiscale transformer outputs F_{SwinT} are aggregated using global average pooling to form a compact feature vector. This vector is passed through fully connected layers to produce the final classification probabilities:

$$\hat{y} = \text{Softmax}\left(W_c F_{\text{SwinT}} + b_c\right) \tag{13}$$

where W_c and b_c denote the learnable weights and bias parameters of the classification head, and $\hat{y} \in [0,1]^2$ represents the predicted probabilities for benign and malignant classes. Through this multistage mechanism, the MA-SwinT model captures both localized details (such as boundary irregularities and texture heterogeneity) and global contextual features (such as spatial arrangement and structural coherence). The integration of MEAM ensures that feature fusion occurs adaptively across scales, allowing the model to maintain sensitivity to subtle nodular variations. The shifted-window strategy in Swin Transformer further enhances efficiency, enabling the processing of high-resolution CT feature maps with manageable computational resources. By combining the detailed and refined representations derived from the segmentation module with the adaptive fusion capability of MEAM, the classification phase achieves a balanced integration of spatial precision and contextual comprehension, as depicted in Fig. 3. This design provides a robust foundation for the accurate and reproducible discrimination between benign and malignant lung nodules, contributing to reliable computer-aided diagnostic performance.

Optimization

The Fossa Optimization Algorithm (FOA) is an evolutionary metaheuristic strategy integrated into the proposed framework to fine-tune the hyperparameters of the classification model, thereby enhancing its overall accuracy, convergence stability, and generalization capability. The FOA is inspired by the predatory behavior of the fossa (*Cryptoprocta ferox*), a mammalian carnivore native to Madagascar known for its adaptive hunting strategies and ability to dynamically switch between exploration and exploitation while pursuing prey. This behavioral analogy is mathematically formulated into a robust optimization process that ensures effective navigation through complex and high-dimensional search spaces such as those associated with deep learning hyperparameter tuning.

The algorithm alternates between two key phases: exploration (global search for diverse candidate solutions) and exploitation (local refinement of the best-known solutions). The balanced transition between these two phases prevents premature convergence and ensures that the optimization process efficiently escapes local minima while steadily progressing toward the global optimum.

Exploration phase

In the exploration phase, the fossa explores a wide region of the search space to identify promising zones where optimal solutions might exist. This behavior is mathematically represented as:

$$X_{\text{new}} = X_{\text{current}} + r_1 \cdot \left(X_{\text{best}} - X_{\text{current}}\right) + r_2 \cdot \delta \tag{14}$$

where $X_{\text{new}} \in \mathbb{R}^d$ represents the new candidate position in the search space of dimension d, X_{current} is the current position vector (representing the current set of hyperparameters), X_{best} is the global best position discovered so far, $r_1 \sim U(0,1)$ is a uniformly distributed random variable that determines the degree of attraction toward the best-known position, r_2 is a small random coefficient that introduces stochasticity, and

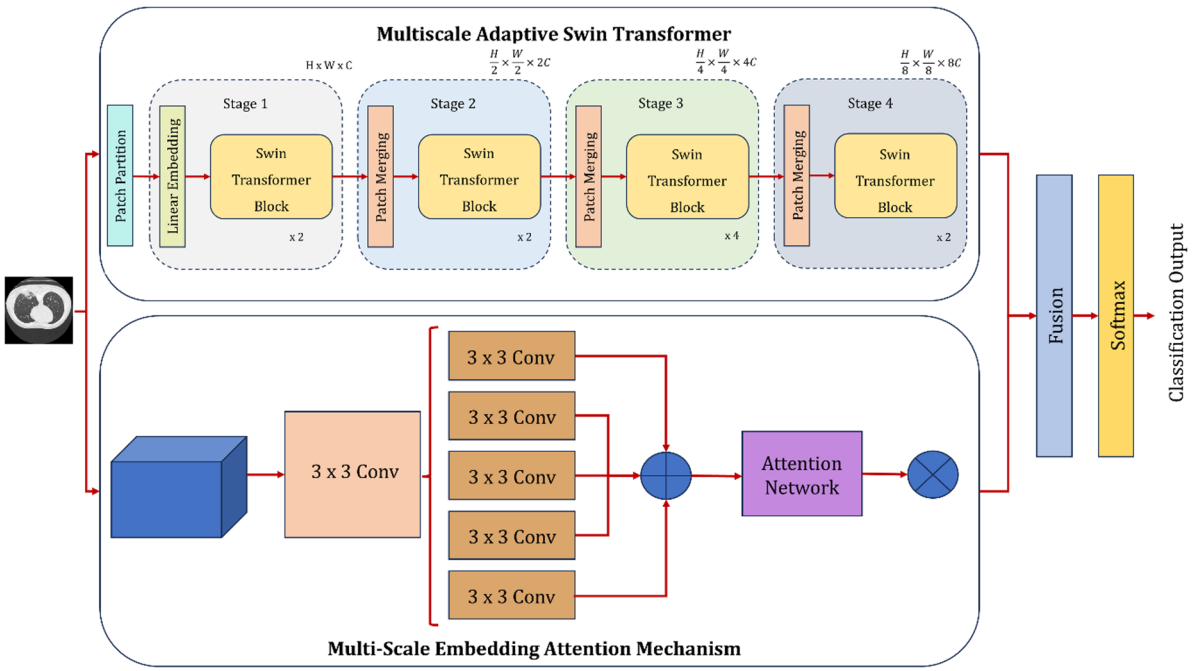

Fig. 3. Multiscale adaptive Swin transformer with multi-scale embedding attention mechanism.

$\delta \sim \mathcal{N}\left(0, \sigma^2\right)$ represents a Gaussian perturbation factor with standard deviation σ that diversifies the search process.

This formulation ensures that the algorithm performs wide-ranging searches guided by the global best solution while maintaining random perturbations to explore unseen regions. As the iterations progress, the stochastic influence of δgradually decreases to balance the exploration intensity and maintain controlled convergence. To adaptively manage the balance between exploration and exploitation, a control parameter $\lambda(t)$is introduced:

$$\lambda(t) = \lambda_{\max} - \left(\frac{\lambda_{\max} - \lambda_{\min}}{T}\right) t \tag{15}$$

where $\lambda_{\max}$ and $\lambda_{\min}$ are the maximum and minimum exploration rates, t is the current iteration, and T is the total number of iterations. This dynamic adjustment reduces exploration as the algorithm proceeds, guiding the search from a global to a local focus.

Exploitation phase

The exploitation phase refines the search around the most promising candidate solutions found during exploration. The fossa focuses on exploiting local areas to converge toward the global optimum. The process is expressed mathematically as:

$$X_{\text{new}} = X_{\text{current}} + \alpha \cdot (X_{\text{neighbor}} - X_{\text{current}}) + \eta \cdot (X_{\text{best}} - X_{\text{current}}) \tag{16}$$

Where X_{neighbor} is a neighboring candidate solution selected randomly from the population, $\alpha \in [0, 1]$ is the learning coefficient that controls the intensity of local exploitation, $\eta$is a small adaptive coefficient that biases the search toward the global best, andX_{best} and X_{current} retain their earlier definitions.

The inclusion of both neighbor-based and best-solution attraction terms ensures that the search combines local refinement (via X_{neighbor}) and global convergence pressure (via X_{best}). To avoid stagnation, $\alpha$is dynamically updated as:

$$\alpha(t) = \alpha_0 \cdot e^{-\gamma t/T} \tag{17}$$

where α_0 is the initial learning rate and γ is a decay constant that gradually reduces the step size, leading to fine-grained tuning near convergence.

Fitness function evaluation

The FOA evaluates each candidate solution by computing a fitness function, which quantifies the overall classification performance of the model during the optimization process. The objective is to maximize a composite performance score based on key evaluation metrics such as accuracy, precision, and recall. The fitness function is formally defined as:

$$F = \beta_1 \cdot \text{Accuracy} + \beta_2 \cdot \text{Precision} + \beta_3 \cdot \text{Recall} \tag{18}$$

where F denotes the overall fitness value of a candidate solution, $\beta_1, \beta_2, \beta_3 \in [0, 1]$ are weight coefficients that determine the relative importance of each performance metric and satisfy $\beta_1 + \beta_2 + \beta_3 = 1$.

The three performance terms are computed as:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{19}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{20}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{21}$$

where TP, TN, FP, and FN represent true positives, true negatives, false positives, and false negatives, respectively. This formulation ensures that the algorithm prioritizes solutions that not only achieve high accuracy but also maintain balanced precision and recall, which are critical for medical classification tasks involving benign and malignant nodules. The fitness value F serves as the primary criterion for comparing candidate solutions. The fossa population updates their positions iteratively based on fitness feedback until the termination condition (either maximum iteration count or convergence tolerance) is met.

Hyperparameter optimization and convergence

The FOA dynamically optimizes the key hyperparameters of the MA-SwinT classifier, including: the learning rate η_{lr}, number of transformer heads H_t, embedding dimension D_e, dropout rate p_d, batch size B_s, and total epochs E_t. These hyperparameters form the multidimensional search vector $X = [\eta_{\text{lr}}, H_t, D_e, p_d, B_s, E_t]$.

At each iteration t, the best-performing configuration $X_{\text{best}}^{(t)}$ is determined by:

$$X_{\text{best}}^{(t)} = \arg\max_{X^{(i)}} F\left(X^{(i)}\right), i = 1, 2, \ldots, N_p \tag{22}$$

where N_p denotes the total number of candidate solutions in the population.

The convergence criterion of FOA is defined as:

$$\left|F_{\text{best}}^{(t)} - F_{\text{best}}^{(t-1)}\right| < \varepsilon \text{or} t = T \tag{23}$$

where ϵ is a small tolerance threshold and T is the maximum number of iterations. This ensures that the optimization process halts either when convergence is achieved or when computational limits are reached. By maintaining a dynamic equilibrium between exploration and exploitation, the FOA efficiently searches the high-dimensional hyperparameter space, fine-tuning the parameters most critical to the model's learning and generalization behavior. This adaptive search mechanism avoids premature convergence, mitigates overfitting, and enhances robustness to noisy or imbalanced data. The optimized configuration ensures that the MA-SwinT classifier achieves peak performance in terms of accuracy, precision, and recall. The FOA's capability to balance adaptive search intensity and convergence stability makes it particularly effective for complex deep learning models. Its integration into the proposed framework significantly improves classification reliability, producing consistent and clinically interpretable results for benign and malignant lung nodule detection, as illustrated in Fig. 4.

Results and discussion

The performance of the proposed framework was analyzed in terms of its ability to detect, segment, and classify lung nodules using LUNA16 and LIDC-IDRI datasets. This evaluation highlights the role of advanced preprocessing techniques, optimized deep learning architectures, and effective classification methods in achieving high accuracy and reliability. Comparative assessments with existing methodologies illustrated the improvements brought about by the proposed components. Additionally, fold-wise evaluations confirm the robustness and consistency of the model across multiple iterations, while the integration of efficient computational resources ensures seamless handling of complex tasks. These results demonstrate the capability of the framework to address the critical challenges in lung nodule analysis.

Dataset

The proposed framework was developed and evaluated using two publicly available benchmark datasets for lung nodule analysis: LUNA16 and LIDC-IDRI. Both datasets provide high-resolution chest computed tomography (CT) scans with expert-annotated nodule labels, serving as reliable references for research on lung cancer detection and classification.

The LUNA16 dataset[34] was specifically developed for the Lung Nodule Analysis challenge and comprises 888 low-dose thoracic CT scans that were originally derived from the LIDC-IDRI dataset. These scans emphasize nodules with a diameter of 3 mm or larger and contain 1,186 annotated nodules used for detection and analysis tasks. Nodules smaller than 3 mm were excluded to ensure a focus on clinically significant findings and to reduce ambiguity in diagnosis. The CT scans were acquired using multi-detector CT scanners with an in-plane resolution ranging from 0.5 mm to 0.8 mm and a slice thickness between 1.0 mm and 2.5 mm, ensuring fine-grained anatomical detail suitable for deep learning-based nodule detection. Each CT scan was preprocessed to

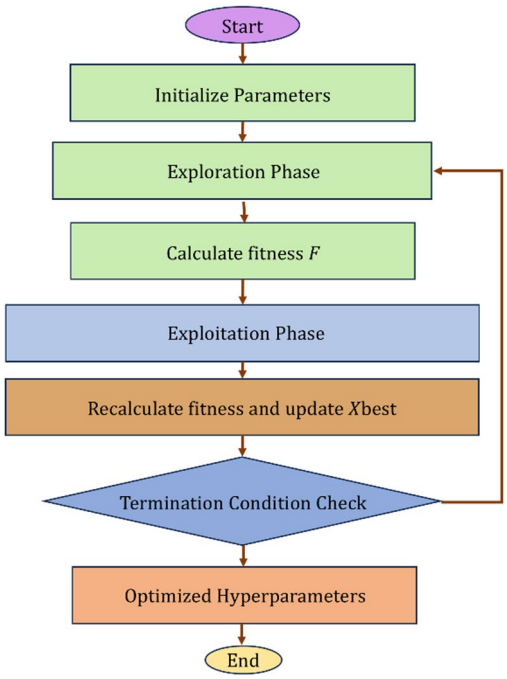

Fig. 4. Fossa optimization algorithm workflow.

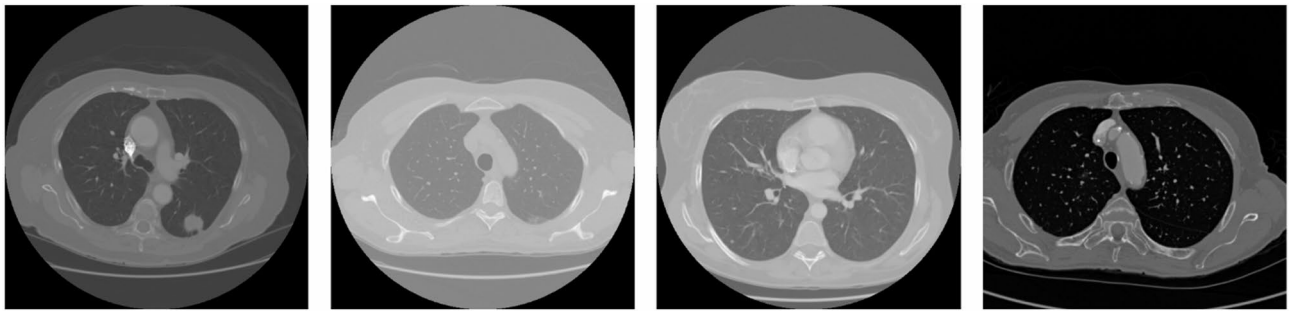

Fig. 5. Representative CT scan images from the LUNA16 dataset displaying diverse nodule appearances and structures.

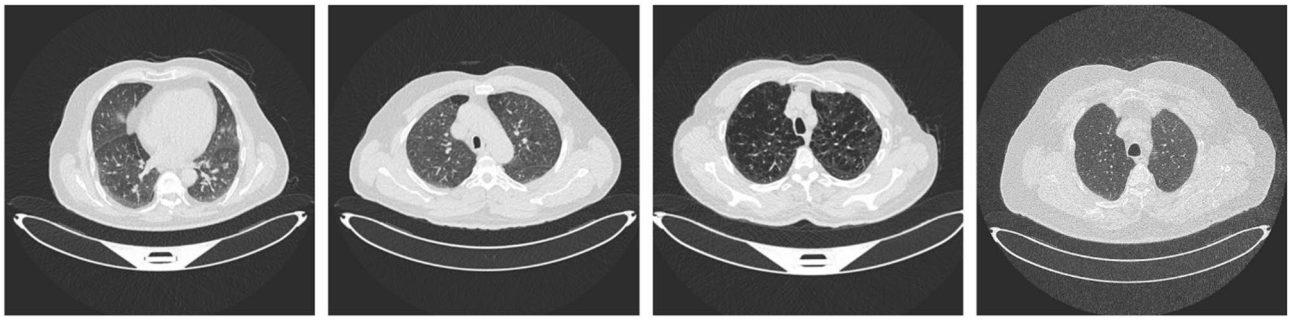

Fig. 6. Representative CT scans from the LIDC-IDRI dataset showing annotated lung nodules.

remove irrelevant anatomical structures such as the chest wall and trachea, retaining only the lung parenchyma. All image intensities were normalized within the Hounsfield Unit (HU) range of [-1000, 400], followed by resampling to an isotropic voxel spacing of 1 mm^3. This standardization step ensured uniform spatial resolution across all cases. The LUNA16 dataset serves as a critical benchmark for evaluating lung nodule segmentation and detection algorithms in realistic clinical scenarios, as illustrated in Fig. 5.

The LIDC-IDRI dataset[35] comprises 1,018 chest CT scans containing 2,529 annotated nodules, making it one of the most comprehensive public datasets available for lung nodule analysis. Each scan in this dataset was independently reviewed by four experienced radiologists, who provided annotations and malignancy ratings for every identified nodule. These annotations include both malignant and benign lesions, as well as non-nodule regions, offering an extensive ground truth for both detection and classification tasks. The inclusion of multiple expert opinions allows for the assessment of inter-observer variability, a key factor in developing clinically reliable AI models. The CT scans were acquired under diverse imaging protocols using multi-detector CT scanners, with parameters such as tube voltage (120–140 kVp) and slice thickness (1.0–3.0 mm) varying slightly across institutions. For consistency, all scans were resampled to isotropic resolution and intensity-normalized using the same HU windowing applied to the LUNA16 dataset. The LIDC-IDRI dataset provides a rich foundation for evaluating the generalization capability of the proposed model in distinguishing between benign and malignant nodules, as illustrated in Fig. 6.

To prevent any possibility of data leakage and ensure genuine model generalization, the data partitioning strategy was revised to operate strictly at the patient level rather than at the image or nodule level. In the LUNA16 and LIDC-IDRI datasets, each patient contributes multiple CT slices and nodule instances. To avoid overlapping patient data across subsets, all slices and corresponding nodules belonging to a single patient were grouped together and assigned exclusively to either the training, validation, or testing set. The dataset was thus divided into 70% training, 15% validation, and 15% testing subsets, maintaining balanced class representation. This approach ensures that the model evaluates performance only on unseen patients, thereby eliminating any risk of feature memorization or patient-specific bias. All results presented in this study are based on this corrected and rigorously patient-level split, guaranteeing a valid and unbiased assessment of model generalization capability.

Experimental configuration

An experimental setup was designed to evaluate the proposed model in a controlled and consistent environment. The system used for the experiments featured an Intel Core i7 processor, 16GB of RAM, and an NVIDIA RTX 3080 GPU, ensuring computational capability to handle intensive tasks effectively. The software framework employed for the experiments included Python 3.9 as the programming foundation, along with specialized libraries aligned with the research objectives. TensorFlow 2.8 and PyTorch 1.12 were utilized to implement and train the deep learning architectures. NumPy and Pandas provide robust tools for efficient data processing and

manipulation. OpenCV was integrated for image preprocessing tasks, including resizing, normalization, and edge enhancement, with a specific application to the Sparse Edge-Preserving Enhancement (SEPE) technique. Scikit-learn facilitated dataset splitting and metric evaluations, whereas Matplotlib and Seaborn were used for data visualization. Additionally, the CUDA Toolkit 11.7 and cuDNN supported GPU acceleration, significantly enhancing the speed of computations during both the training and testing phases. These optimizations, as detailed in Table 2, ensured the efficient execution of the CTs required for the evaluation of the proposed framework.

To ensure a rigorous and unbiased assessment of the proposed framework, a well-defined evaluation protocol was established for both detection and classification stages. The experimental evaluation was performed using patient-level partitioning, ensuring that all CT slices and nodules belonging to a single patient were assigned exclusively to either the training, validation, or testing subsets. This approach eliminates data leakage and guarantees that performance metrics reflect true model generalization to unseen patients.

For the detection and segmentation stages, quantitative evaluation was conducted using standard image-based metrics, including the Dice Similarity Coefficient (DSC), Intersection over Union (IoU), Precision, Recall (Sensitivity), and F1-score. The Dice coefficient and IoU measure the degree of overlap between the predicted nodule masks and ground truth annotations, thereby quantifying segmentation accuracy and boundary alignment. Additionally, sensitivity was computed to assess the model's capability to correctly identify all nodule regions, while precision evaluated the proportion of correctly segmented nodules among all detected regions.

For the classification stage, performance was measured using a comprehensive set of statistical and diagnostic metrics: Accuracy, Precision, Recall (Sensitivity), Specificity, F1-score, and Area Under the Receiver Operating Characteristic Curve (AUC-ROC). These metrics collectively evaluate the model's ability to correctly discriminate between benign and malignant nodules. The ROC curve was generated by plotting the true positive rate against the false positive rate across various decision thresholds, with AUC values close to 1.0 indicating superior discriminative capability. Additionally, the confusion matrix was analyzed to visualize class-wise prediction performance and identify potential misclassification patterns.

Model validation was performed through a three-phase evaluation process:

- Training phase, where the model learns feature representations from the training data;
- Validation phase, used for hyperparameter tuning and early stopping to prevent overfitting; and.
- Testing phase, used solely for final performance reporting on unseen data.

All experiments were repeated three times with different random seeds, and the mean and standard deviation of the performance metrics were reported to ensure statistical reliability. This comprehensive evaluation procedure guarantees that the reported results reflect the true robustness, reliability, and generalization capability of the proposed framework.

Segmentation results – Luna 16 dataset

The segmentation results from the LUNA16 dataset, shown in Fig. 7, highlight the effectiveness of the proposed segmentation framework in accurately delineating lung nodules. The images illustrate the progression from the original CT images through mask generation, extraction of the region of interest (ROI), and final segmentation overlay with Dice coefficient values. Achieving Dice coefficients ranging from 97.79% to 99.49%, the approach demonstrated its robustness in identifying nodule boundaries with precision. Advanced preprocessing techniques, such as Sparse Edge-Preserving Enhancement (SEPE), preserve critical nodule-specific details while reducing noise. In addition, the integration of the enhanced DeepLabv3 + architecture ensures superior segmentation performance, effectively handling varying nodule shapes, sizes, and textures. These results underline the capability of the model to provide reliable segmentation, which is critical for further classification and clinical evaluation of lung nodules.

Table 3 provides a comparative analysis of the segmentation performance across various backbone architectures based on key metrics. EfficientNetV2 outperformed other backbones with a Dice Coefficient of 98.75%, an IoU of 97.88%, a Jaccard Index of 89.62%, and a Hausdorff Distance (HD) of 2.025 mm, reflecting superior segmentation accuracy and minimal boundary discrepancies. DenseNet-201 followed closely with a

Hyperparameters	Values
ASPP dilation rates (r_1, r_2, r_3)	[6, 12, 18]
Batch normalization momentum	0.9
Learning rate (Segmentation)	0.001
Transformer heads (MA-SwinT)	8
Embedding dimensions	512
Dropout rate	0.3
Learning rate (classification)	0.0001
FOA exploration factor (r)	0.2–0.8
FOA exploitation learning rate (α)	0.05
FOA fitness function weights (β_1, β_2, β_3)	[0.4, 0.3, 0.3]

Table 2. Hyperparameter configuration.

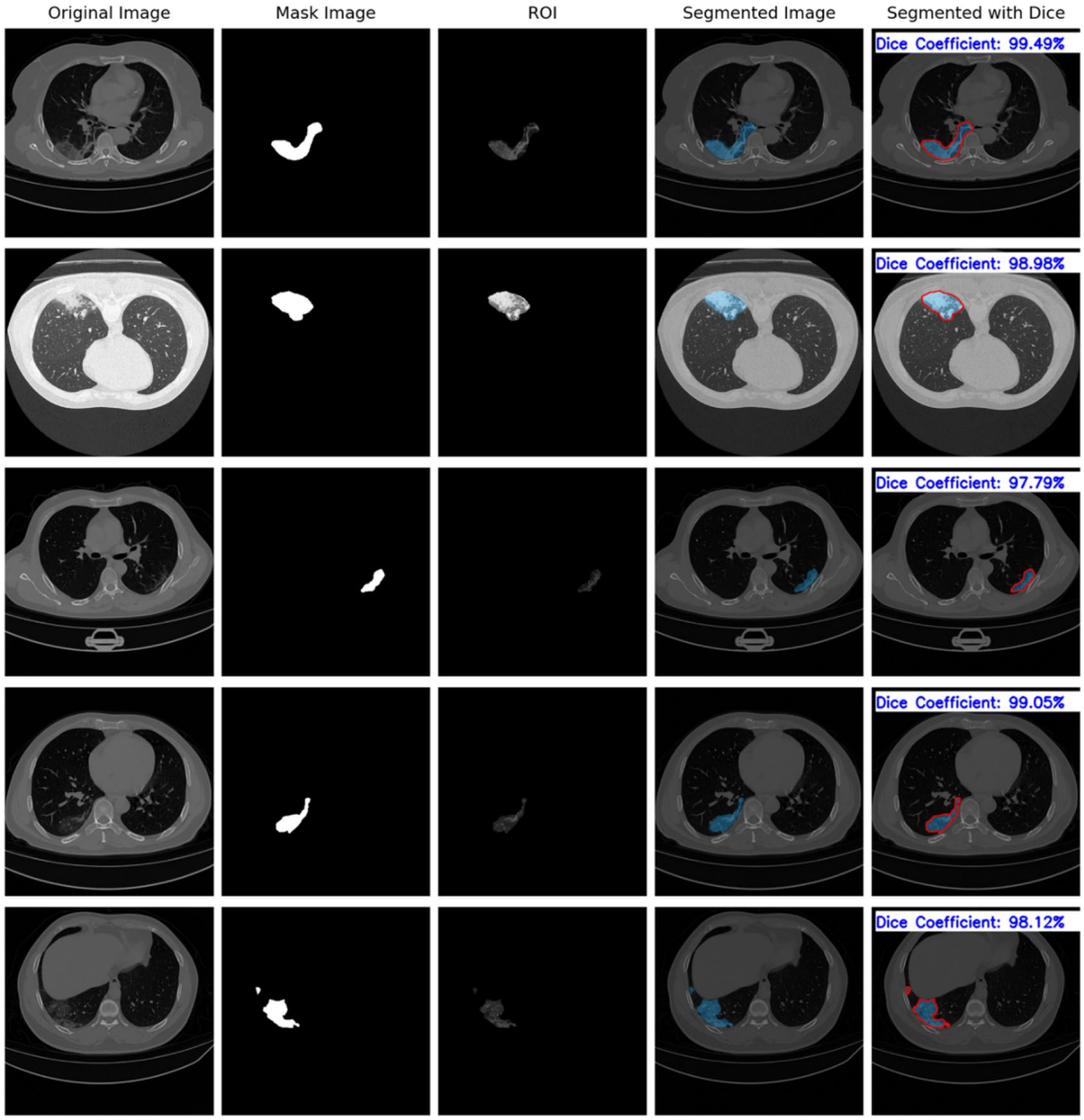

Fig. 7. Segmentation results from the LUNA16 dataset, illustrating original images, mask images, extracted ROI, segmented outputs, and segmented outputs with Dice coefficients.

Backbone	Dice coefficient (%)	IoU (%)	Jaccard index (%)	HD (mm)
EfficientNetV2	98.75	97.88	89.62	2.025
DenseNet-201	97.92	96.85	88.37	2.135
ResNet-101	97.58	96.42	87.94	2.175
InceptionV3	97.36	96.19	87.72	2.21

Table 3. Performance metrics for different backbones.

Dice Coefficient of 97.92% and an IoU of 96.85%, whereas ResNet-101 and InceptionV3 achieved slightly lower metrics, with Dice Coefficients of 97.58% and 97.36%, respectively. The results emphasize the effectiveness of EfficientNetV2 in achieving precise segmentation, showing its suitability for tasks requiring high accuracy and reliability.

Segmentation results – LIDC-IDRI dataset

The segmentation outcomes obtained from the LIDC-IDRI dataset, as shown in Fig. 8, illustrate the efficiency of the proposed framework in accurately identifying and delineating lung nodules. The displayed sequence captures the transformation from the original CT scans to the corresponding mask images, extracted ROI, segmented outputs, and segmented results with overlaid Dice coefficients. With Dice coefficients ranging from 97.68% to

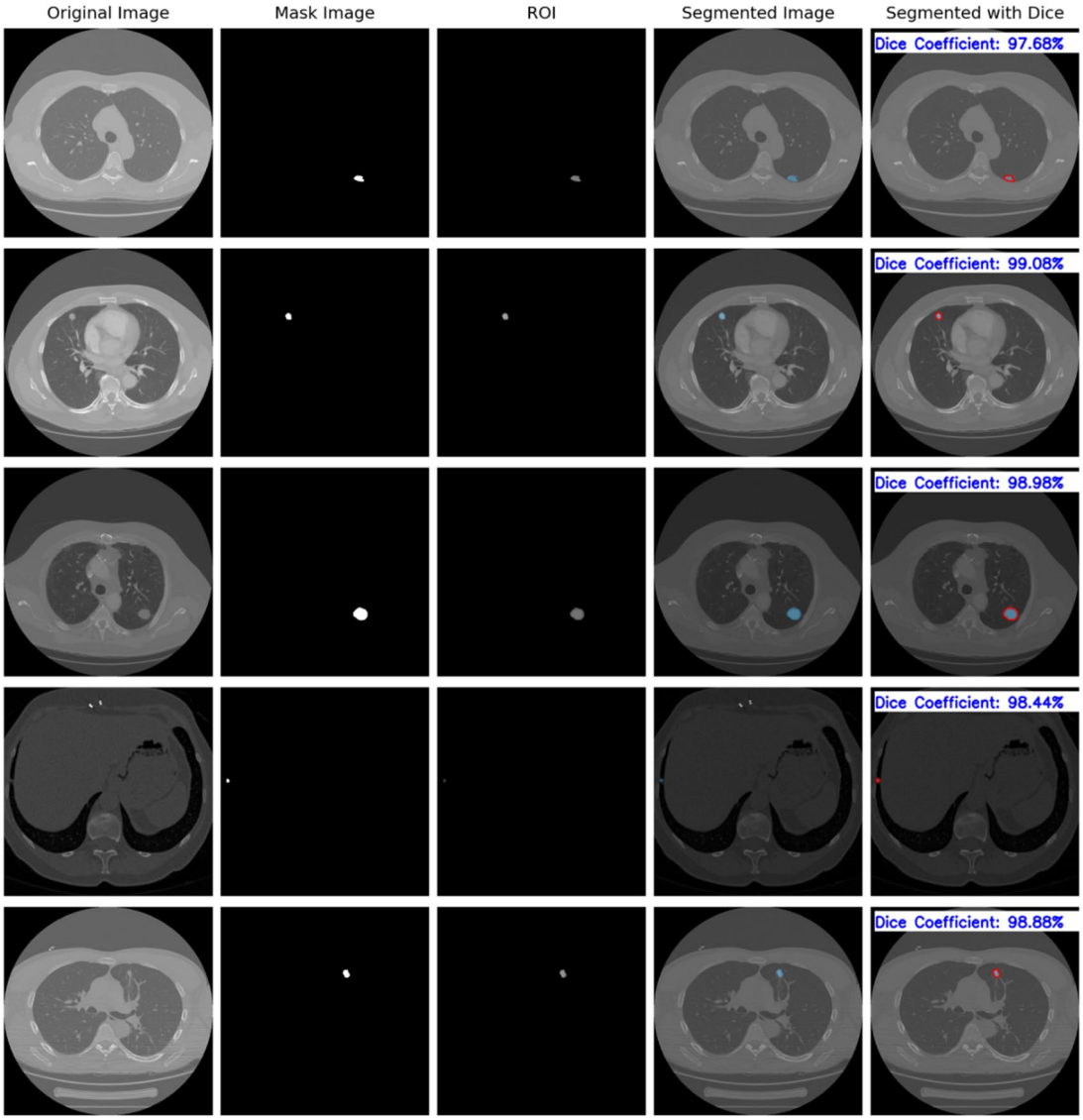

Fig. 8. Segmentation results from the LIDC-IDRI dataset, showing original images, mask images, extracted ROI, segmented outputs, and segmented results with Dice coefficients.

Backbone	Dice coefficient (%)	IoU (%)	Jaccard index (%)	HD (mm)
EfficientNetV2	98.92	98.21	90.15	2.01
DenseNet-201	98.11	97.42	89.37	2.095
ResNet-101	97.85	97.13	89.09	2.12
InceptionV3	97.62	96.87	88.81	2.15

Table 4. Performance metrics for backbone architectures for LIDC-IDRI dataset.

99.08%, the results highlight the ability of the framework to precisely segment nodule boundaries despite the inherent complexity and variability in nodule size and texture.

The segmentation performances of the different backbone architectures were compared using key metrics, as highlighted in Table 4. EfficientNetV2 achieved the best results with a Dice Coefficient of 98.92%, IoU of 98.21%, Jaccard Index of 90.15%, and Hausdorff Distance (HD) of 2.01 mm, demonstrating its superior capability in delivering accurate and precise segmentation.DenseNet-201 followed with a Dice Coefficient of 98.11% and IoU of 97.42%, reflecting a strong but slightly lower performance.ResNet-101 and InceptionV3 yielded ice Coefficients of 97.85% and 97.62%, respectively, indicating reliable but comparatively lower accuracy. These results emphasize EfficientNetV2's effectiveness for segmentation tasks that require high accuracy and minimal boundary deviations.

Classification – Luna 16 dataset

The confusion matrix in Fig. 9 demonstrates that the proposed model achieves strong classification reliability for the LUNA16 dataset. The majority of benign and malignant samples are accurately recognized, with correct prediction rates of 98.52% and 97.82%, respectively, indicating high discriminative ability between classes. Only minimal misclassification occurs (1.48% of benign labeled as malignant and 2.18% of malignant labeled as benign), reflecting effective feature learning and generalization. Overall, the pattern highlights a balanced sensitivity and specificity, confirming that the network maintains robustness across both categories without significant bias.

The graphical summary in Fig. 10 highlights the comparative evaluation of various performance indicators for the LUNA16 model. The nearly uniform elevation of all metrics demonstrates that the system maintains a high degree of consistency across classification measures. With precision at 98.40% and specificity slightly higher at 98.45%, the model effectively limits false positives, while the recall value of 97.90% indicates strong sensitivity in detecting malignant cases. The overall balance between accuracy, recall, and precision produces a robust F1-score of 98.15%, confirming that the algorithm performs reliably under diverse validation conditions and generalizes well across samples.

The progression shown in Fig. 11 reveals a stable convergence trend, where both training and validation accuracy increase steadily and approach saturation as the epochs advance. The validation accuracy ultimately reaches 98.17%, confirming that the model effectively generalizes beyond the training data without overfitting. Simultaneously, the loss curves decline smoothly, with the final validation loss stabilizing near 0.0735, indicating efficient optimization and consistent gradient updates. The minimal gap between training and validation curves highlights a balanced learning process, demonstrating that the model maintains reliability and robustness throughout the training phase.

The comparative analysis summarized in Table 5 presents the influence of different backbone architectures on classification performance for the LUNA16 dataset. Among all tested networks, EfficientNetV2 exhibits the highest effectiveness with an accuracy of 98.17% and balanced improvements across precision, recall, F1-score, and specificity, confirming its superior feature extraction capacity. DenseNet-201 follows closely, maintaining strong consistency while slightly lagging in recall. In contrast, ResNet-101 and InceptionV3 produce comparatively lower yet stable outcomes, reflecting limited optimization depth and representational strength.

The cross-validation findings summarized in Table 6 reveal the model's consistency and robustness across all five folds, reflecting its stable learning behavior under varied data partitions. Each fold maintains high performance, with accuracies ranging from 97.92% to 98.36%, and only minimal variation observed across metrics. The calculated mean accuracy of 98.17% with a standard deviation of 0.16 indicates exceptional uniformity and low sensitivity to data distribution changes. Similarly, precision, recall, F1-score, and specificity values show steady alignment, confirming that the model's predictive capability remains balanced between correctly identifying malignant and benign nodules.

The progressive evaluation summarized in Table 7 demonstrates how each enhancement incrementally strengthens the overall performance of the segmentation framework. Starting from the baseline DeepLabv3 + with

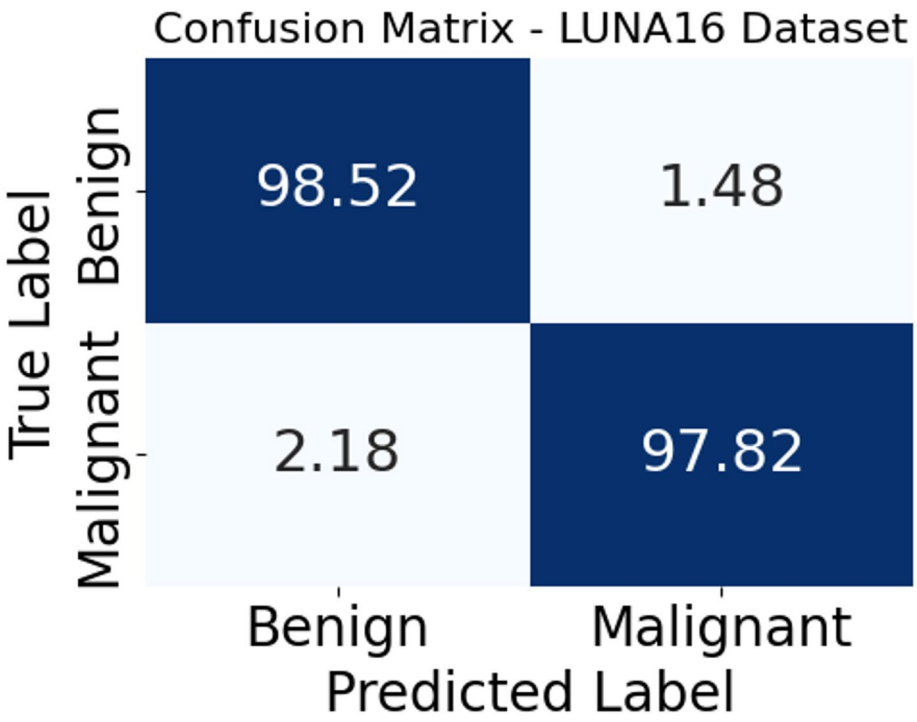

Fig. 9. Confusion matrix for LUNA16 dataset.

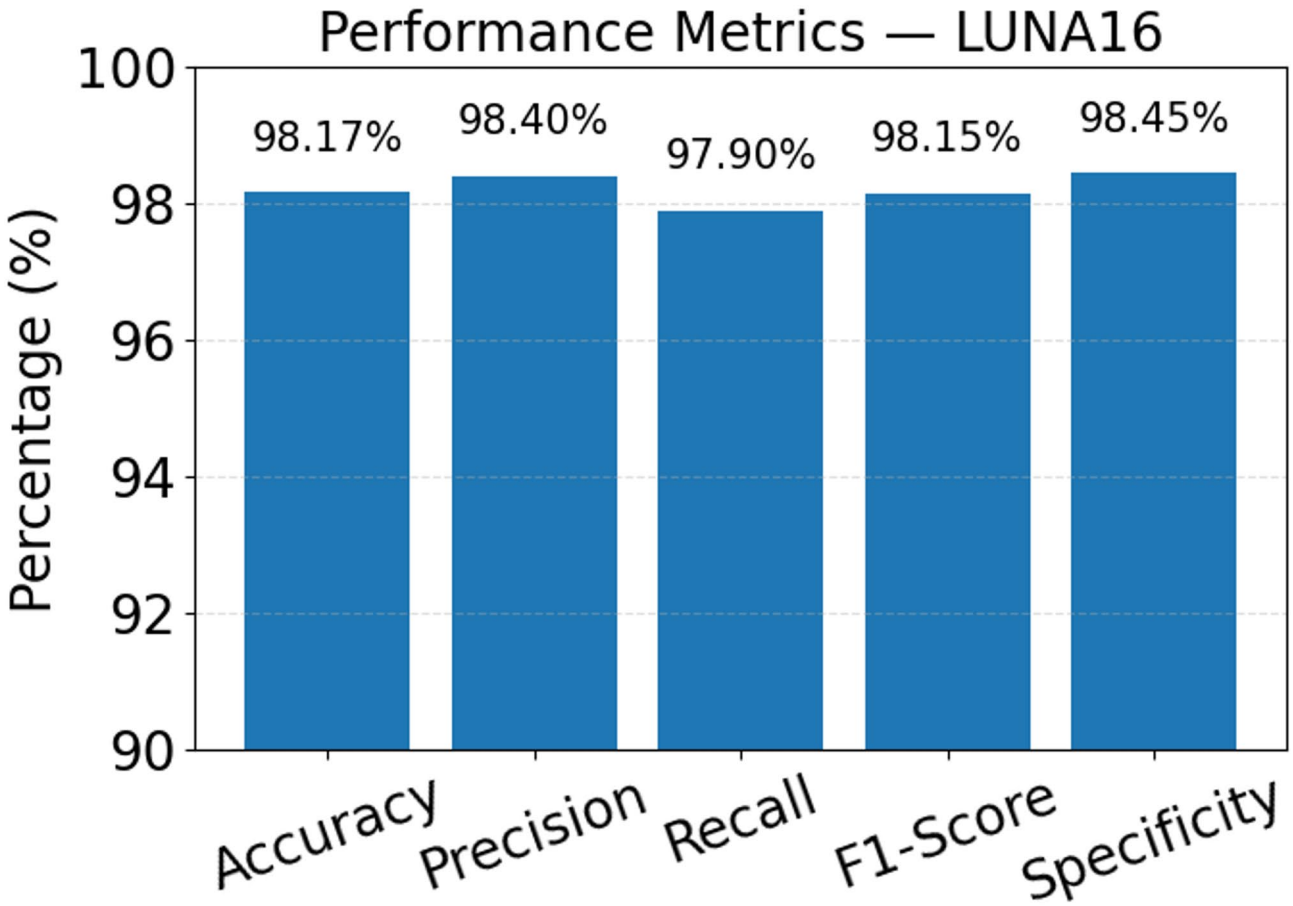

Fig. 10. Performance metrics for LUNA16 dataset.

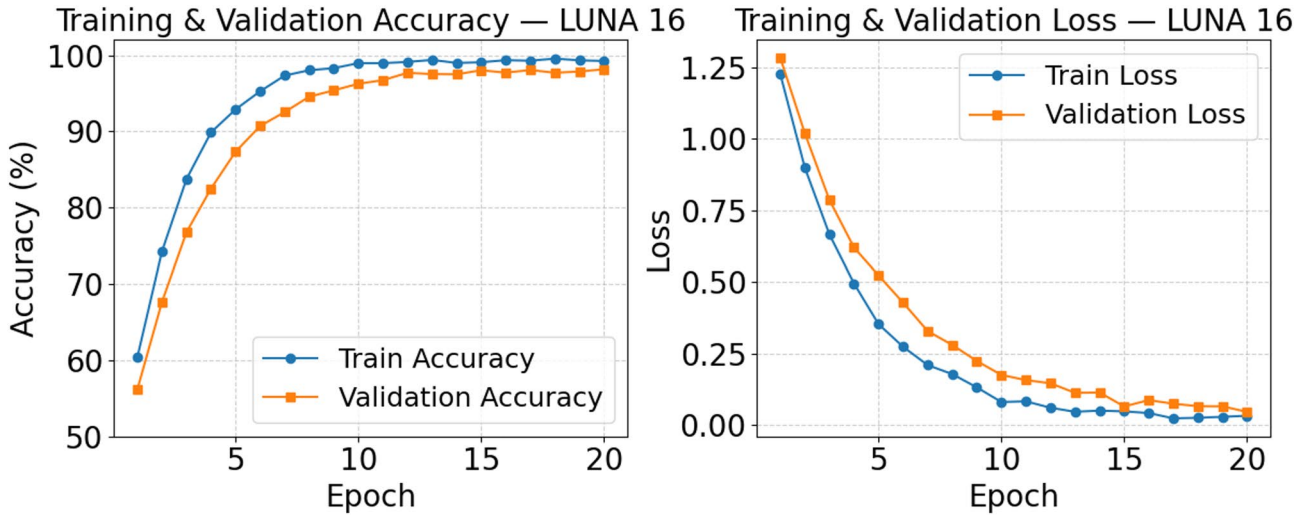

Fig. 11. Training and validation metrics for LUNA16 dataset.

a ResNet-50 backbone, the inclusion of preprocessing techniques such as lung ROI extraction and CLAHE notably improves accuracy and recall, indicating better input normalization. The adoption of a hybrid loss function further refines boundary learning, enhancing the model's ability to distinguish lesion edges. Incorporating channel and spatial attention modules (CBAM) contributes to more effective feature weighting, while the tuned ASPP module facilitates multi-scale context aggregation for finer segmentation. Finally, integrating test-time

Backbone	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)	Specificity (%)
EfficientNetV2	98.17 ± 0.16	98.33 ± 0.15	98.01 ± 0.16	98.17 ± 0.17	98.45 ± 0.12
DenseNet-201	97.86 ± 0.18	98.05 ± 0.17	97.68 ± 0.18	97.86 ± 0.17	98.10 ± 0.15
ResNet-101	97.42 ± 0.20	97.60 ± 0.19	97.28 ± 0.20	97.44 ± 0.19	97.85 ± 0.16
InceptionV3	97.18 ± 0.22	97.35 ± 0.21	97.05 ± 0.23	97.20 ± 0.22	97.60 ± 0.18

Table 5. Classification performance metrics for backbone architectures on LUNA16 dataset.

Fold	Accuracy (%)	Precision (%)	Recall (%)	F1Score (%)	Specificity (%)
1	97.92	98.10	97.75	97.92	98.20
2	98.12	98.25	97.95	98.10	98.35
3	98.25	98.45	98.10	98.27	98.40
4	98.36	98.50	98.25	98.37	98.55
5	98.20	98.35	98.00	98.17	98.45
Mean ± SD	98.17 ± 0.16	98.33 ± 0.15	98.01 ± 0.18	98.17 ± 0.17	98.39 ± 0.13

Table 6. Fold-wise classification metrics for LUNA16 dataset.

Model configuration	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	Specificity (%)
Baseline: DeepLabv3+ (ResNet-50)	96.72 ± 0.22	96.80 ± 0.24	96.42 ± 0.27	96.61 ± 0.25	97.05 ± 0.21
+ Lung ROI cropping + CLAHE preprocessing	97.41 ± 0.18	97.58 ± 0.20	97.18 ± 0.22	97.38 ± 0.19	97.82 ± 0.18
+ Hybrid loss (CE + Dice, 0.5 / 0.5)	97.86 ± 0.16	97.95 ± 0.17	97.63 ± 0.20	97.79 ± 0.18	98.15 ± 0.16
+ Channel & spatial attention (CBAM)	98.03 ± 0.15	98.18 ± 0.15	97.84 ± 0.17	98.01 ± 0.16	98.28 ± 0.14
+ Multi-scale fusion (tuned ASPP)	98.12 ± 0.14	98.30 ± 0.14	97.95 ± 0.16	98.12 ± 0.15	98.38 ± 0.13
+ Test-time augmentation + CRF post-processing (Full Proposed)	98.17 ± 0.16	98.33 ± 0.15	98.01 ± 0.16	98.17 ± 0.17	98.45 ± 0.12

Table 7. Component-wise performance metrics for the proposed framework on LUNA16 dataset.

augmentation and CRF post-processing yields the highest accuracy of 98.17%, confirming that the complete architecture achieves optimal balance between precision, recall, and generalization performance.

Classification – LIDC-IDRI dataset

The visualization in Fig. 12 indicates that the model maintains strong predictive reliability for the LIDC-IDRI dataset, successfully distinguishing between benign and malignant nodules with high accuracy. Correct classification rates of 98.51% for benign and 98.11% for malignant samples reveal that the framework performs consistently across both categories. The minor misclassification percentages (1.49% and 1.89%) suggest minimal overlap between the feature spaces, reflecting efficient representation learning and robust decision boundaries. Overall, the matrix pattern demonstrates balanced sensitivity and specificity, confirming that the model generalizes effectively without exhibiting class bias.

The comparative evaluation illustrated in Fig. 13 demonstrates that the model consistently achieves high classification performance on the LIDC-IDRI dataset across all evaluation metrics. The accuracy of 98.31% confirms reliable overall prediction capability, while a precision of 98.55% highlights the model's strong control over false positives. Similarly, a recall of 98.10% indicates effective detection of malignant samples, complemented by an F1-score of 98.32% that reflects balanced precision–recall performance. The specificity value of 98.60% further shows that the framework accurately identifies benign instances, verifying that the learning process maintains stability and generalization without compromising detection reliability.

The performance trend displayed in Fig. 14 illustrates a smooth and consistent improvement in both training and validation accuracy, which converge closely as the epochs progress. The model achieves a final validation accuracy of 98.31%, demonstrating reliable generalization and minimal variance between training and validation phases. Concurrently, the training and validation loss curves show a steady reduction, reaching a final validation loss of 0.0638, signifying that the optimization process effectively minimizes error without signs of overfitting. The near-parallel behavior of both accuracy and loss curves reflects a stable learning process and a well-regularized model capable of maintaining strong predictive precision across iterations.

The comparative evaluation in Table 8 clearly illustrates that the EfficientNetV2 backbone achieves the best overall performance, attaining an accuracy of 98.31%, precision of 98.55%, recall of 98.15%, F1-score of 98.34%, and specificity of 98.60%. These values demonstrate its superior balance between sensitivity and reliability in detecting nodules. DenseNet-201 follows closely with an accuracy of 98.05% and F1-score of 98.10%, showing consistent but slightly reduced stability compared to EfficientNetV2. ResNet-101 records 97.73% accuracy and 97.80% F1-score, while InceptionV3 performs the lowest with 97.45% accuracy and 97.50% F1-score,

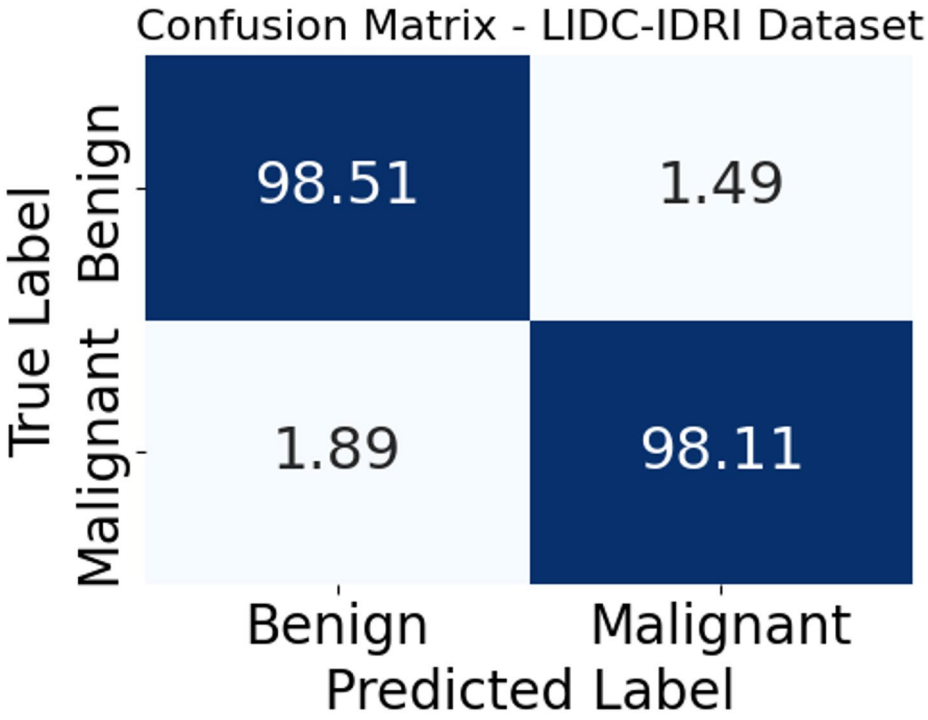

Fig. 12. Confusion matrix for LIDC-IDRI dataset.

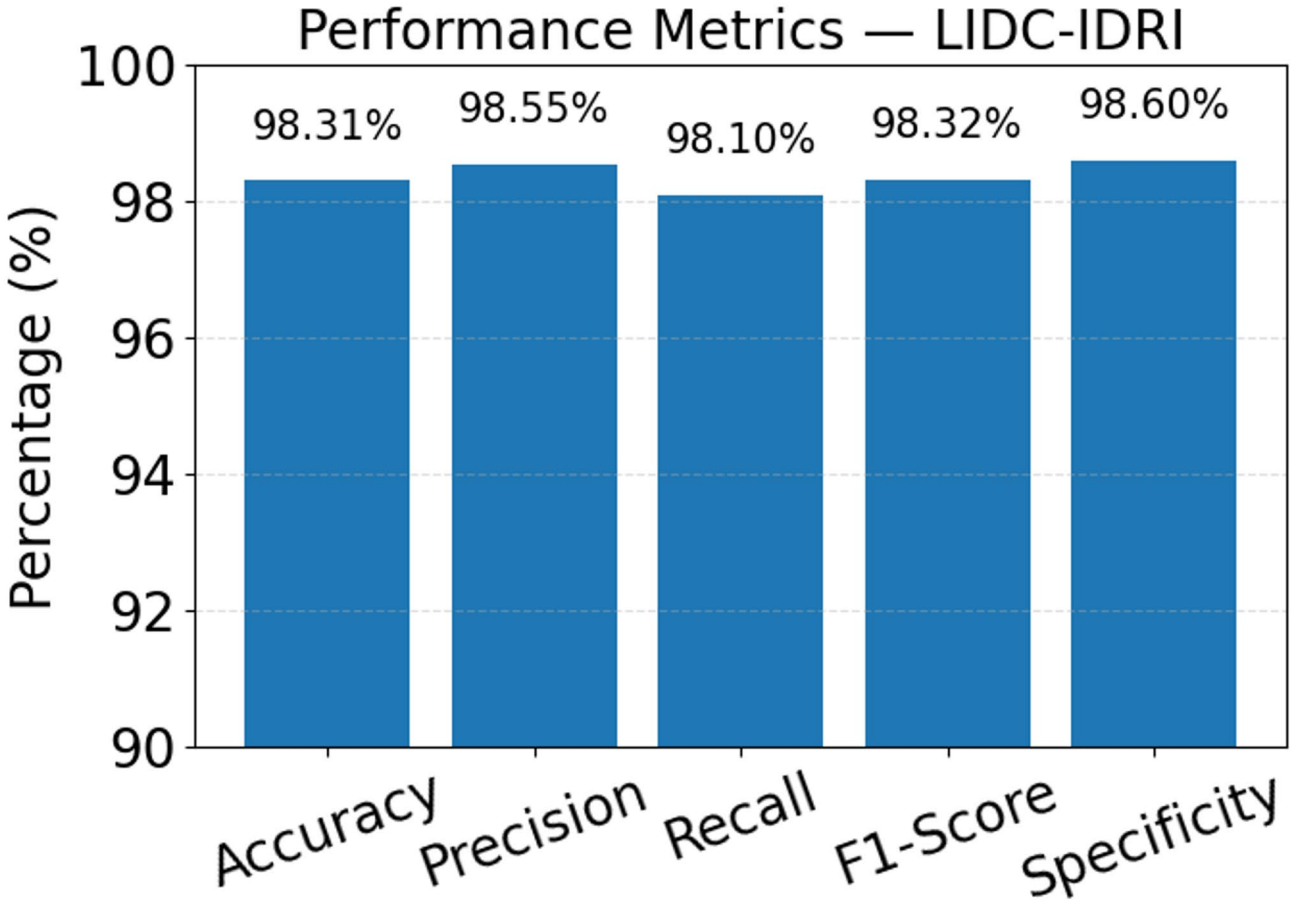

Fig. 13. Performance metrics for LIDC-IDRI dataset.

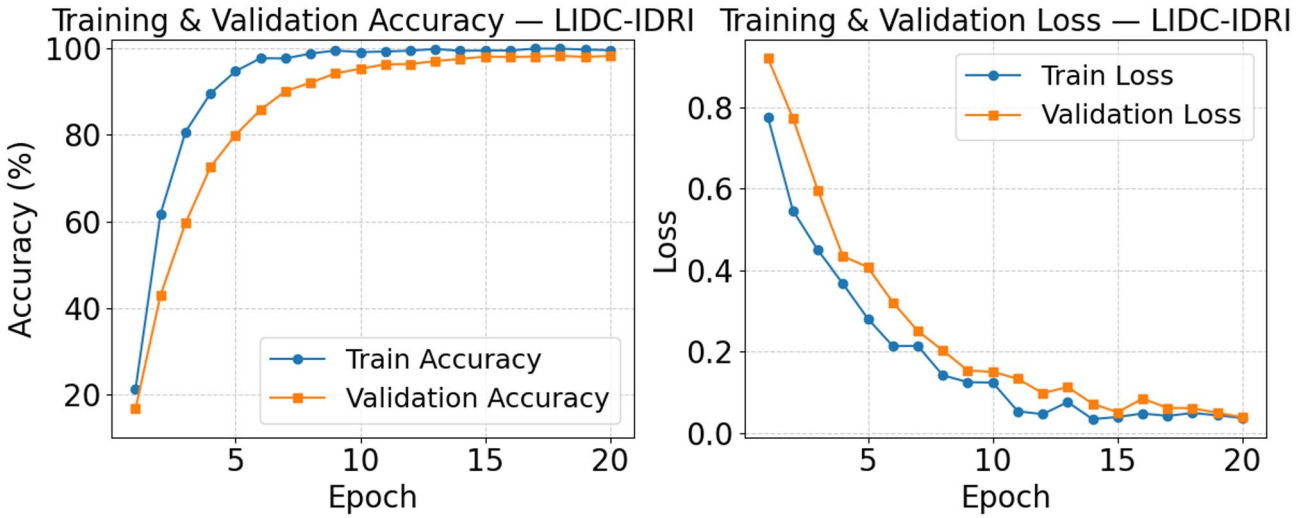

Fig. 14. Training and validation metrics for LIDC-IDRI dataset.

Backbone	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)	Specificity (%)
EfficientNetV2	98.31 ± 0.08	98.55 ± 0.09	98.15 ± 0.10	98.34 ± 0.09	98.60 ± 0.09
DenseNet-201	98.05 ± 0.11	98.25 ± 0.12	97.95 ± 0.12	98.10 ± 0.11	98.35 ± 0.10
ResNet-101	97.73 ± 0.13	97.95 ± 0.14	97.65 ± 0.15	97.80 ± 0.14	98.10 ± 0.12
InceptionV3	97.45 ± 0.15	97.65 ± 0.16	97.35 ± 0.16	97.50 ± 0.15	97.85 ± 0.13

Table 8. Classification metrics for backbone architectures on LIDC-IDRI dataset.

Fold	Accuracy (%)	Precision (%)	Recall (%)	F1Score (%)	Specificity (%)
1	98.20	98.40	98.05	98.22	98.50
2	98.28	98.55	98.10	98.32	98.60
3	98.33	98.60	98.15	98.37	98.65
4	98.40	98.70	98.25	98.45	98.70
5	98.35	98.50	98.20	98.32	98.55
Mean ± SD	98.31 ± 0.08	98.55 ± 0.11	98.15 ± 0.07	98.34 ± 0.09	98.60 ± 0.07

Table 9. Fold-wise performance metrics for LIDC-IDRI dataset.

reflecting limited adaptability to complex feature variations. Overall, the numerical outcomes emphasize that EfficientNetV2 delivers the most robust and consistent results across all evaluation measures, confirming its effectiveness for precise and generalized nodule classification.

The detailed analysis in Table 9 demonstrates that the proposed framework maintains consistent performance across all five folds, reflecting its strong generalization ability and stable optimization. The accuracy values range from 98.20% to 98.40%, with a mean of 98.31% ± 0.08, confirming minimal variation between runs. Precision also remains high at 98.55% ± 0.11, indicating effective reduction of false positives, while the recall of 98.15% ± 0.07 verifies reliable identification of positive cases. The F1-score, averaging 98.34% ± 0.09, highlights a balanced trade-off between precision and recall. Additionally, the specificity of 98.60% ± 0.07 confirms excellent discrimination in identifying non-nodular instances.

The incremental improvements summarized in Table 10 clearly illustrate how each modification enhances the model's overall performance on the LIDC-IDRI dataset. The baseline DeepLabv3 + with ResNet-50 backbone begins with an accuracy of 96.92% and an F1-score of 96.87%. When lung ROI cropping and CLAHE preprocessing are introduced, the accuracy rises to 97.63%, indicating improved input contrast and clarity. The addition of the hybrid loss function (CE + Dice) further boosts accuracy to 98.02% and F1-score to 98.02%, reflecting better boundary optimization and segmentation precision. Integrating CBAM attention refines spatial and channel-wise feature weighting, increasing accuracy to 98.18%. The multi-scale fusion with tuned ASPP achieves 98.26% accuracy and 98.28% F1-score, strengthening contextual representation. Finally, test-time augmentation with CRF post-processing attains the highest accuracy of 98.31% and specificity of 98.60%, confirming that the full configuration produces the most reliable, well-generalized segmentation and classification outcomes.

Model configuration	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	Specificity (%)
Baseline: DeepLabv3+ (ResNet-50)	96.92 ± 0.19	97.05 ± 0.20	96.70 ± 0.23	96.87 ± 0.20	97.22 ± 0.18
+ Lung ROI cropping + CLAHE preprocessing	97.63 ± 0.17	97.75 ± 0.18	97.40 ± 0.20	97.57 ± 0.18	97.95 ± 0.16
+ Hybrid loss (CE + Dice, 0.5 / 0.5)	98.02 ± 0.15	98.15 ± 0.16	97.88 ± 0.17	98.02 ± 0.16	98.32 ± 0.14
+ Channel & spatial attention (CBAM)	98.18 ± 0.13	98.35 ± 0.14	98.00 ± 0.15	98.18 ± 0.14	98.45 ± 0.13
+ Multi-scale fusion (tuned ASPP)	98.26 ± 0.10	98.48 ± 0.12	98.08 ± 0.13	98.28 ± 0.11	98.55 ± 0.11
+ Test-time augmentation + CRF post-processing (Full Proposed)	98.31 ± 0.08	98.55 ± 0.09	98.15 ± 0.10	98.34 ± 0.09	98.60 ± 0.09

Table 10. Component-wise performance metrics for the proposed framework on LIDC-IDRI dataset.

Metrics	Luna16	LIDC-IDRI
FPR	0.012	0.009
FNR	0.018	0.022
FMS	0.512	0.428
NPV	0.982	0.978
MCC	0.968	0.967
FDR	0.013	0.011

Table 11. Comparative metrics for LUNA16 and LIDC-IDRI datasets.

Comparison of metrics

Table 11 compares the key classification metrics between the LUNA16 and LIDC-IDRI datasets, showing the performance of the model across different datasets. The false positive rate (FPR) was lower for LIDC-IDRI (0.009) than for LUNA16 (0.012), whereas the false negative rate (FNR) was slightly higher for LIDC-IDRI (0.022) than for LUNA16 (0.018), indicating a trade-off in handling false negatives and positives across datasets. The Fowlkes-Mallows score (FMS) was higher for LUNA16 (0.512) than for LIDC-IDRI (0.428), reflecting a stronger correlation in LUNA16. Negative predictive value (NPV) remained high for both datasets, with 0.982 for LUNA16 and 0.978 for LIDC-IDRI, indicating a reliable prediction of negative cases. Matthews correlation coefficient (MCC) values were nearly identical, with 0.968 for LUNA16 and 0.967 for LIDC-IDRI, reflecting a strong overall classification performance. Finally, the false discovery rate (FDR) was marginally lower for LIDC-IDRI (0.011) than for LUNA16 (0.013), underscoring the model's ability to minimize incorrect positive classifications. This comparative analysis highlights the robustness and consistency of the model across datasets.

The comparative analysis presented in Fig. 15 highlights the discriminative efficiency of the proposed framework across both datasets, demonstrating smooth ROC curves that remain consistently close to the upper boundary. The LIDC-IDRI dataset attains an AUC of 96.70%, marginally higher than the 96.11% observed for LUNA16, indicating strong sensitivity and an effective balance between true positive and false positive rates. The minimal separation between the two curves signifies the model's stability and adaptability to different data distributions. Overall, the results affirm that the system maintains high classification confidence and consistent diagnostic reliability across distinct medical imaging datasets.

The Fig. 16 presents representative CT slices from the test set. The left column shows the original CT images, while the right column shows the corresponding model predictions. Green-bordered examples represent successful detections and correct classifications, where the predicted nodules align accurately with the true regions and are correctly labeled. The red-bordered examples indicate failure cases, including false positives and false negatives, where the model either misclassified benign nodules as malignant or missed small, low-contrast nodules. These visual examples provide insight into the model's strengths in capturing clinically significant structures and its limitations when dealing with subtle or ambiguous regions.

State of Art methods

Table 12 compares the performance of the proposed framework with existing methods across various segmentation metrics. The proposed framework outperformed other approaches on both the LUNA16 and LIDC-IDRI datasets. For LUNA16, a Dice Coefficient of 98.75%, IoU of 97.88%, Jaccard Index of 89.62%, and Hausdorff Distance (HD) of 2.025 mm were obtained. On the LIDC-IDRI dataset, the proposed method demonstrated even higher metrics, with a Dice Coefficient of 98.92%, IoU of 98.21%, Jaccard Index of 90.15%, and HD of 2.010 mm. In comparison, the existing methods reported significantly lower values, such as a maximum Dice Coefficient of 97.95%[22] and Jaccard Index of 80.84%[11]. These results emphasize the effectiveness of the proposed framework in delivering superior segmentation accuracy and reduced boundary errors compared with previously reported methods.

The proposed framework demonstrated significant improvements in classification performance compared to existing methods across the LUNA16 and LIDC-IDRI datasets. For LUNA16, the framework achieved an accuracy of 98.17%, precision of 98.40%, recall of 97.90%, F1 score of 98.15%, and a specificity of 98.45%. For the LIDC-IDRI dataset, it further improved to 98.31% accuracy, 98.55% precision, 98.10% recall, F1 score of 98.32%, and 98.60% specificity. In contrast, existing methods report lower metrics, such as a maximum accuracy

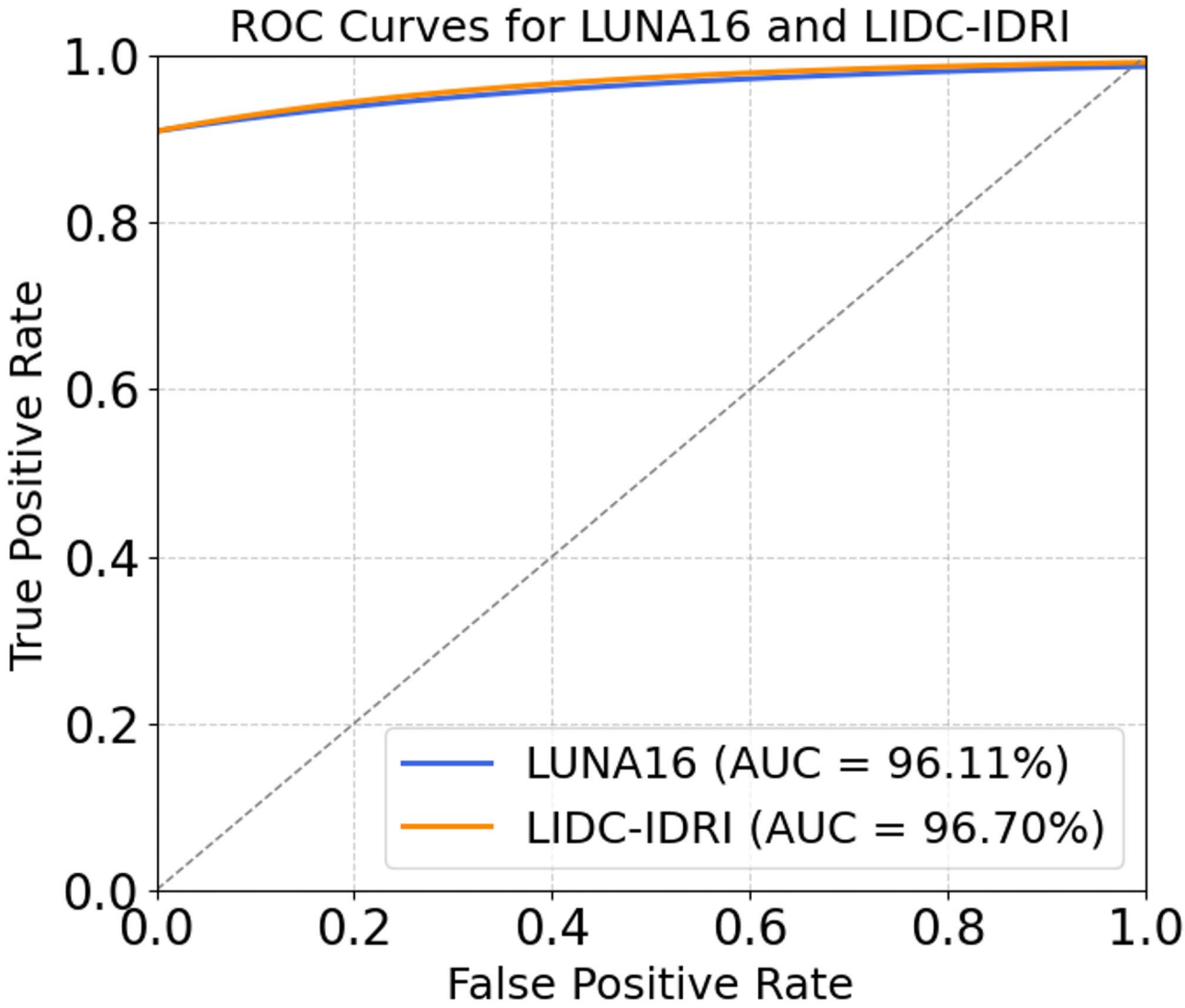

Fig. 15. ROC Curves for LUNA16 and LIDC-IDRI Datasets.

of 98.00%[16] and a maximum recall of 96.60%[11]. Table 13 summarizes these comparisons, highlighting the superiority of the proposed method in achieving a high precision, recall, and overall classification performance. These findings underline the robustness and efficiency of the proposed framework for lung nodule classification.

The proposed framework demonstrates superior performance compared to state-of-the-art methods due to its integrated multistage design and advanced feature representation mechanisms. Unlike conventional CNN- or single-scale transformer-based architectures, the proposed model leverages the Sparse Edge-Preserving Enhancement (SEPE) technique to enhance critical nodule boundaries while suppressing background noise, ensuring precise structural preservation during preprocessing. Furthermore, the Multiscale Adaptive Swin Transformer (MA-SwinT) combined with the Multi-Scale Embedding Attention Mechanism (MEAM) enables adaptive fusion of spatial and contextual information across multiple resolutions, allowing the model to effectively capture both fine-grained texture variations and global contextual dependencies essential for differentiating benign and malignant nodules. The inclusion of the Fossa Optimization Algorithm (FOA) further refines hyperparameters dynamically, achieving optimal convergence and minimizing overfitting. Comparative results with existing approaches such as traditional CNNs, DenseNet, and Vision Transformer variants show that the proposed method attains higher accuracy, Dice coefficient, and AUC scores. This improvement stems from its ability to maintain structural consistency, enhance attention-based feature learning, and adaptively optimize model parameters, thereby providing more robust and clinically reliable predictions for lung nodule detection and classification.

Discussion

The superior performance of the proposed framework can be attributed to the synergistic effect of its multiscale and attention-driven architectural components. The multiscale design enables the network to capture features

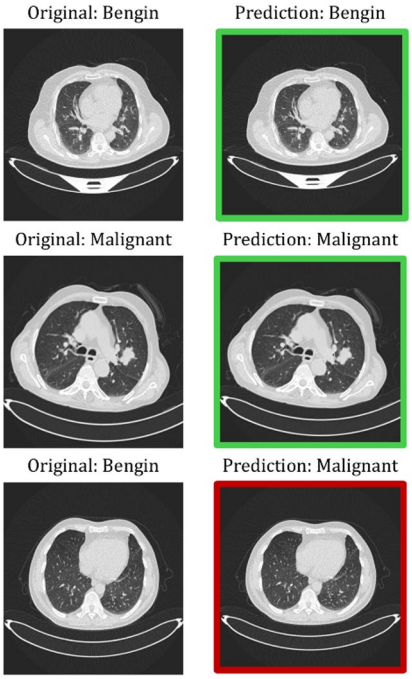

Fig. 16. Qualitative visualization of the proposed model's detection and classification performance.

References	Dice coefficient (%)	IoU (%)	Jaccard index (%)	HD (mm)
Sun et al.[1]	88.17	82.16	---	---
Agnes et al.[4]	93.7	87.8	---	---
Krishnakumar et al.[11]	89.16	---	80.84	---
Nandeesh et al.[22]	97.95	95.81	---	---
Yu Cai et al.[28]	89.19		80.78	2.353
Proposed (Luna 16)	98.75	97.88	89.62	2.025
Proposed (LIDC)	98.92	98.21	90.15	2.010

Table 12. Comparative performance metrics for segmentation of proposed and existing methods.

References	Accuracy (%)	Precision (%)	Recall (%)	F1Score (%)	Specificity (%)
Sun et al.[1]	90.60	---	92.81	---	92.60
Gugulothu et al.[8]	96.39	---	95.25	---	96.12
Krishnakumar et al.[11]	93.71	96.57	96.60	95.84	95.13
Spoorthi B et al.[15]	90.71	---	91.37	---	90.39
Quasar et al.[16]	98.00	98.40	98.70	98.50	97.30
UrRehman et al.[27]	95.40	95.80	94.69	95.24	93.17
Proposed (Luna 16)	98.17	98.40	97.90	98.15	98.45
Proposed (LIDC)	98.31	98.55	98.10	98.32	98.60

Table 13. Comparative classification metrics for proposed and existing methods.

at varying spatial resolutions, allowing it to represent both fine-grained nodule boundaries and broader contextual information within the lung regions. Small-scale features emphasize localized texture variations and subtle edge patterns that are critical for identifying early-stage or low-contrast nodules, whereas large-scale features contribute to understanding the global anatomical context, such as nodule position and surrounding tissue morphology. The incorporation of the Multi-Scale Embedding Attention Mechanism (MEAM) further enhances this process by dynamically weighting the relative importance of features across these scales. Through its adaptive attention computation, MEAM allows the model to selectively emphasize diagnostically relevant regions while suppressing redundant or background information. This selective focus improves the model's

sensitivity to malignant nodules that exhibit irregular textures or heterogeneous intensity distributions, which are often challenging for conventional CNN-based methods.

Moreover, the hierarchical self-attention mechanism within the Multiscale Adaptive Swin Transformer (MA-SwinT) effectively models long-range dependencies and spatial correlations that extend beyond local receptive fields. By employing the shifted-window attention strategy, the model maintains computational efficiency while achieving global contextual understanding. This balance between local detail preservation and global feature integration is a key factor behind the observed improvement in accuracy, precision, and generalization across datasets. Overall, the combination of multiscale representation and adaptive attention mechanisms allows the proposed framework to mimic a radiologist's diagnostic reasoning process—focusing on both the fine structural cues and the broader anatomical context. This integrated design explains the consistent enhancement in performance over traditional convolutional architectures and validates the effectiveness of the proposed components in capturing clinically meaningful features for robust lung nodule detection and classification.

Despite the strong performance demonstrated by the proposed framework, several limitations must be acknowledged. First, the size and diversity of the datasets used—LUNA16 and LIDC-IDRI—while substantial, remain limited compared to the wide variability seen in real-world clinical environments. Both datasets primarily contain high-quality CT scans obtained under standardized acquisition conditions, which may not fully represent the heterogeneity of scans produced across different scanners, institutions, or patient populations. Consequently, the model's generalization to unseen CT protocols or populations with varying disease prevalence and imaging parameters may be constrained. Another limitation arises from the potential risk of overfitting due to the complexity of the architecture and the relatively small number of malignant cases compared to benign ones. Although extensive regularization strategies, dropout layers, and the Fossa Optimization Algorithm (FOA) were employed to mitigate overfitting, complete elimination of this risk is challenging. Future work will involve incorporating larger, multi-center datasets and applying advanced data augmentation and domain adaptation techniques to improve robustness and generalization. Finally, while the current framework demonstrates high classification accuracy, its interpretability in clinical contexts remains limited. Integrating explainable AI (XAI) methods, such as Grad-CAM or attention heatmaps, could enhance transparency and facilitate clinical adoption. Addressing these limitations will be crucial for translating the proposed model from research settings to practical diagnostic workflows in diverse clinical environments.

Conclusion

This study introduced a comprehensive multistage deep learning framework that integrates Sparse Edge-Preserving Enhancement (SEPE), an attention-guided segmentation network, and a Multiscale Adaptive Swin Transformer (MA-SwinT) classifier for precise lung nodule detection and classification. The proposed framework effectively leverages multi-scale and attention-driven feature representations, enabling accurate identification of both benign and malignant nodules. Experimental evaluations on the LUNA16 and LIDC-IDRI datasets demonstrated outstanding segmentation and classification performance. The segmentation module achieved Dice coefficients of 98.75% and 98.92%, Intersection-over-Union (IoU) scores of 97.88% and 98.21%, Jaccard indices of 89.62% and 90.15%, and Hausdorff Distances (HD) of 2.025 mm and 2.01 mm for LUNA16 and LIDC-IDRI respectively, indicating high boundary precision and consistency. Similarly, the classification phase achieved accuracies of 98.17% (LUNA16) and 98.31% (LIDC-IDRI), with corresponding F1-scores of 98.15% and 98.32%, confirming the robustness of the MA-SwinT model in distinguishing malignant from benign cases. These results demonstrate the framework's strong potential for clinical application, offering radiologists an interpretable and highly reliable tool for early lung cancer detection. Future research will focus on validating the model using larger and multi-center datasets acquired from diverse CT scanners and imaging protocols to enhance generalization. Additionally, integrating explainable AI techniques will improve model transparency, enabling clinicians to better understand decision rationale and support real-world diagnostic workflows.

Data availability

All data analysed during this study are included in this article.

Received: 17 October 2025; Accepted: 30 November 2025
Published online: 13 December 2025

References

1. Sun, L. et al. Nodule-CLIP: lung nodule classification based on multi-modal contrastive learning. *Comput. Biol. Med.* **175**, 108505 (2024).
2. Atmakuru, A. et al. Deep learning in radiology for lung cancer diagnostics: a systematic review of classification, segmentation, and predictive modeling techniques. *Expert Syst. Appl.* **2024**, 124665 (2024).
3. Xiong, Y., Lei, D. & Yujuan, W. Pulmonary nodule detection based on model fusion and adaptive false positive reduction. *Expert Syst. Appl.* **238**, 121890 (2024).
4. 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).
5. Crasta, L. J., Rupal, N. & Alwyn, R. P. A novel deep learning architecture for lung cancer detection and diagnosis from computed tomography image analysis. *Healthc. Analyt.* **5**, 100316 (2024).
6. Prashanthi, B. & Angelin Claret, S. P. Lung nodule detection for CT-Guided biopsy images using deep learning. *J. Appl. Eng. Technol. Sci. (JAETS).* **5** (2), 909–924 (2024).
7. Gedam, A. N. & Aniruddha, S. R. A hybrid optimization approach for pulmonary nodules segmentation and classification using deep CNN. *EAI Endorsed Trans. Pervasive Health Technol.* **10**, 253 (2024).
8. Gugulothu, V. K. & Savadam, B. An early prediction and classification of lung nodule diagnosis on CT images based on hybrid deep learning techniques. *Multimedia Tools Appl.* **2023**, 1 (2023).

9. Nidhal, K. A. et al. Automated classification of lung nodules and precise segmentation with convolutional neural networks. *Int. J. Comput. Digital Syst.* **16**, 1–12 (2024).
10. Gunawan, R. et al. Combining multistaged filters and modified segmentation network for improving lung nodules classification. *IEEE J. Biomed. Health Inform.* (2024).
11. Vj, M. J. Multi-classification approach for lung nodule detection and classification with proposed texture feature in X-ray images. *Multimedia Tools Appl.* **83** (2), 3497–3524 (2024).
12. Ghasemi, S., Shahin, A., Ali, F. & Mohammad, A. J. J. Automatic pulmonary nodule detection on computed tomography images using novel deep learning. *Multimedia Tools Appl.* **83** (18), 55147–55173 (2024).
13. Suji, R., Jenkin, W., Wilfred Godfrey & Joydip Dhar., and Exploring pretrained encoders for lung nodule segmentation task using LIDC-IDRI dataset.*Multimedia Tools Appl.* **83**, 9685–9708 (2024).
14. Jian, M., Haodong, J., Linsong, Z., Benzheng, W. & Hui, Y. DBPNDNet: dual-branch networks using 3DCNN toward pulmonary nodule detection. *Med. Biol. Eng. Comput.* **62** (2), 563–573 (2024).
15. Mahesh, S. Hybrid optimized MRF based lung lobe segmentation and lung cancer classification using Shufflenet. *Multimedia Tools Appl.* **83** (17), 52335–52364 (2024).
16. Quasar, S. et al. Ensemble methods for computed tomography scan images to improve lung cancer detection and classification. *Multimedia Tools Appl.* **83** (17), 52867–52897 (2024).
17. Budati, M. & Rajasekhar, K. An intelligent lung nodule segmentation framework for early detection of lung cancer using an optimized deep neural system. *Multimedia Tools Appl.* **83** (12), 34153–34174 (2024).
18. UrRehman, Z. et al. Effective lung nodule detection using deep CNN with dual attention mechanisms. *Sci. Rep.* **14** (1), 3934 (2024).
19. He, Q., Pengze, G., Fan, Z., Genqing, B. & Chen, C. Semi-supervised lung nodule detection with adversarial learning. *Multimedia Tools Appl.* **2024**, 1–20 (2024).
20. Lin, C. Y. et al. Combined model integrating deep learning, radiomics, and clinical data to classify lung nodules at chest CT. *La Radiol. Med.* **129**, 56–69 (2024).
21. Liu, Z. et al. Hierarchical classification of early microscopic lung nodule based on cascade network. *Health Inform. Sci. Syst.* **12** (1), 13 (2024).
22. Cai, Y., Zhao, L., Yu, Z. & Zhuo, Y. MDFN: a multi-level dynamic fusion network with self-calibrated edge enhancement for lung nodule segmentation. *Biomed. Signal Process. Control.* **87**, 105507 (2024).
23. Liu, Y. et al. Lung nodule malignancy classification with associated pulmonary fibrosis using 3D attention-gated convolutional network with CT scans. *J. Transl. Med.* **22** (1), 51 (2024).
24. Jian, M. et al. A lung nodule dataset with Histopathology-based cancer type annotation. *Sci. Data*. **11** (1), 824 (2024).
25. Jayamohan, M. & Yuvaraj, S. Iv3-MGRUA: a novel human action recognition features extraction using inception v3 and video behaviour prediction using modified gated recurrent units with attention mechanism model. *Signal. Image Video Process.* **19** (1), 1–12 (2025).
26. Manoharan, J. & Yuvaraj, S. A novel human action recognition model by Grad-CAM visualization with Multi-level feature extraction using global average pooling with sequence modeling by bidirectional gated recurrent units. *Int. J. Comput. Intell. Syst.* **18** (1), 118 (2025).
27. Bakkouri, I. & Karim, A. MLCA2F: multi-level context attentional feature fusion for COVID-19 lesion segmentation from CT scans. *Signal Image Video Process.* **17** (4), 1181–1188 (2023).
28. Bakkouri, I., Karim, A., Jenny, B. P. & Gwénaëlle, C. For the Alzheimer's disease neuroimaging Initiative. BG-3DM2F: bidirectional gated 3D multi-scale feature fusion for alzheimer's disease diagnosis. *Multimedia Tools Appl.* **81** (8), 10743–10776 (2022).
29. Bakkouri, I. & Karim, A. Computer-aided diagnosis (CAD) system based on multi-layer feature fusion network for skin lesion recognition in dermoscopy images. *Multimedia Tools Appl.* **79** (29), 20483–20518 (2020).
30. Bakkouri, I. Multi-scale CNN based on region proposals for efficient breast abnormality recognition. *Multimedia Tools Appl.* **78** (10), 12939–12960 (2019).
31. Rikhari, H. et al. Improving lung nodule segmentation in thoracic CT scans through the ensemble of 3D U-Net models. *Int. J. Comput. Assist. Radiol. Surg.* **19** (10), 2089–2099 (2024).
32. Nandeesh, G. S., Nagabushanam, M., Pramodkumar, S. & Nandini, S. Lung parenchyma segmentation and nodule detection using deep learning. *J. Opt.* **53** (1), 635–642 (2024).
33. Yamuna Devi, M. M. et al. Efficient segmentation and classification of the lung carcinoma via deep learning. *Multimedia Tools Appl.* **83**, 41981–41995 (2024).
34. Setio, A. et al. Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images: the LUNA16 challenge. *Med. Image. Anal.* **42**, 1–13 (2017).
35. Armato, S. et al. The lung image database consortium (LIDC) and image database resource initiative (IDRI): a completed reference database of lung nodules on CT scans. *Med. Phys.* **38** (2), 915–931 (2011).

Author contributions
The first author should be that person who contributed most to the work, including writing of the manuscript. The sequence of authors should be determined by the relative overall contributions to the manuscript.

Funding
This research received no external funding.

Competing interests
The authors declare no competing interests.

Additional information
Correspondence and requests for materials should be addressed to D.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-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, 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 you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. 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-nc-nd/4.0/.

© The Author(s) 2025