Spectrum of Engineering Sciences
ISSN (e) 3007-3138 (p) 3007-312X Volume 4, Issue 3, 2026

A NOVEL MULTI-SCALE ATTENTION-BASED DEEP LEARNING ARCHITECTURE FOR AUTOMATED LUNG CANCER SCREENING, EARLY TUMOR DETECTION, AND INTELLIGENT CLINICAL DECISION SUPPORT USING HIGH-RESOLUTION COMPUTED TOMOGRAPHY IMAGE ANALYSIS

Muhammad Suleman[1], Syed Zain Ali[2], Azzah Khadim Hussain[3], Muhammad Aqeel[4], Syed Sameed Abbas[5], Ashraf Zia[*6]

[1]Department of Electrical Engineering, University of Science and Technology Bannu, Bannu 28100, Khyber Pakhtunkhwa, Pakistan.
[2]Khyber Medical University, Peshawar, Pakistan.
[3]Lecturer, Govt Dyal Singh Graduate College, Lahore, Pakistan.
[4]Department of Computer Science, National University of Modern Languages, Lahore 54000, Pakistan.
[5]Department of Computer Science, Faculty of Engineering Science and Technology, Iqra University, Karachi, Sindh, Pakistan.
*[*6]Department of Computer Science, Abdul Wali Khan University Mardan, Mardan, KP, Pakistan.*

[1]sulemanktk@ustb.edu.pk, [2]s.zainali1222@gmail.com, [3]azzah.khadim@gmail.com, [4]muhammad.aqeel@numl.edu.pk, [5]sameed.abbas@iqra.edu.pk, [*6]ashrafzia@awkum.edu.pk

DOI: https://doi.org/10.5281/zenodo.21818398

Keywords
Lung cancer screening; Deep learning; Multi-scale attention mechanism; Computed tomography; Pulmonary nodule detection; Computer-aided diagnosis; Intelligent clinical decision support; Medical image analysis.

Article History
Received: 27 January 2026
Accepted: 12 March 2026
Published: 26 March 2026

Corresponding Author: *
Ashraf Zia

Abstract

Lung cancer remains one of the leading causes of cancer-related mortality worldwide, primarily because a large proportion of cases are diagnosed at advanced stages when treatment options are limited. Early identification of malignant pulmonary nodules from computed tomography (CT) images can substantially improve patient survival and clinical outcomes. This study proposes a novel multi-scale attention-based deep learning architecture for automated lung cancer screening, early tumor detection, and intelligent clinical decision support using high-resolution computed tomography image analysis. The proposed framework is designed to capture both local fine-grained features and global contextual information through multi-scale feature extraction and attention mechanisms, thereby enhancing the discrimination between benign and malignant pulmonary nodules. The experimental study was conducted using the publicly available LIDC-IDRI (Lung Image Database Consortium and Image Database Resource Initiative) and LUNA16 (LUng Nodule Analysis 2016 Challenge) datasets, comprising more than 1,180 thoracic CT scans and over 8,800 annotated pulmonary nodules reviewed by experienced thoracic radiologists. CT images were preprocessed through intensity normalization, lung parenchyma segmentation, noise reduction using Gaussian filtering, image resizing, and data augmentation techniques including rotation, flipping, scaling, and contrast enhancement to improve model robustness and reduce overfitting.

The proposed architecture was implemented using Python 3.11, TensorFlow 2.16 and Matplotlib, while training and evaluation were performed on an NVIDIA GPU-enabled computing platform with five-fold cross-validation. Experimental results demonstrate that the proposed framework outperformed several state-of-the-art deep learning models, including ResNet50, DenseNet121, EfficientNet-B4, and Vision Transformer. The proposed model achieved an accuracy of 98.4%, precision of 98.1%, recall (sensitivity) of 98.3%, specificity of 97.9%, F1-score of 98.2%, ROC-AUC of 0.996, and a Dice Similarity Coefficient of 97.5% for pulmonary nodule localization. Statistical analysis confirmed the superiority of the proposed method ($p < 0.001$), while Grad-CAM visualization demonstrated accurate localization of suspicious tumor regions, improving the interpretability of automated predictions. The proposed multi-scale attention-based framework provides a reliable and interpretable computer-aided diagnostic solution that enhances screening efficiency, supports radiologists in early lung cancer diagnosis, reduces false-positive findings, and facilitates intelligent clinical decision-making. The developed framework has strong potential for integration into next-generation smart healthcare systems and hospital radiology workflows to improve diagnostic accuracy and patient management.

1- Introduction:

Lung cancer is one of the most prevalent and life-threatening malignancies worldwide, accounting for the highest number of cancer-related deaths among both men and women. According to global cancer statistics, millions of new lung cancer cases are diagnosed annually, and the disease remains responsible for a substantial proportion of cancer mortality despite continuous advancements in diagnostic imaging, targeted therapies, and clinical management. The primary reason for this high mortality rate is that lung cancer is frequently detected at advanced stages, when tumors have already metastasized and treatment options become less effective. Conversely, patients diagnosed during the early stages exhibit significantly higher survival rates, highlighting the critical importance of timely detection and accurate diagnosis. Computed tomography imaging has become the preferred modality for lung cancer screening because of its superior spatial resolution and ability to visualize small pulmonary nodules that are often invisible in conventional chest radiography. High-resolution CT imaging provides detailed anatomical information regarding lung tissues, enabling radiologists to identify suspicious lesions at an early stage [1]. Large-scale screening trials have demonstrated that low-dose CT screening can substantially reduce lung cancer mortality by facilitating earlier diagnosis. Nevertheless, the continuously increasing volume of CT examinations has significantly increased the workload of radiologists. Manual interpretation of hundreds of image slices for each patient is labor-intensive, time-consuming, and susceptible to inter-observer variability, false-positive findings, and missed lesions, particularly when nodules are extremely small, irregularly shaped, or located adjacent to blood vessels and other anatomical structures. Recent advances in artificial intelligence, particularly deep learning, have transformed the field of medical image analysis by enabling automated extraction of complex visual features directly from imaging data. Convolutional neural networks have demonstrated remarkable performance in numerous medical imaging tasks, including disease classification, lesion segmentation, organ localization, and abnormality detection [2]. Unlike traditional machine learning approaches that rely on handcrafted feature engineering, deep learning models automatically learn hierarchical feature representations, thereby improving diagnostic performance. Despite these achievements, conventional CNN architectures still encounter several challenges when applied to

lung CT analysis. Their fixed receptive fields may fail to effectively capture both fine-grained local texture information and large-scale contextual characteristics simultaneously, while limited feature fusion often reduces sensitivity for detecting small pulmonary nodules. Furthermore, many existing deep learning models operate as "black-box" systems, limiting interpretability and reducing clinician confidence in automated diagnostic decisions. Attention mechanisms have recently emerged as an effective solution for improving feature representation within deep neural networks. By selectively emphasizing clinically relevant regions and suppressing redundant background information, attention modules enable models to focus on suspicious pulmonary nodules while preserving important contextual information from surrounding lung tissues. Similarly, multi-scale feature learning enhances model performance by integrating information extracted at different spatial resolutions, allowing simultaneous recognition of tiny nodules, medium-sized lesions, and larger tumor masses. The integration of multi-scale feature extraction with attention-based learning therefore offers a promising strategy for overcoming the limitations of conventional CNN architectures and improving diagnostic accuracy across diverse clinical scenarios.

In addition to accurate lesion detection, there is an increasing need for intelligent clinical decision support systems capable of assisting radiologists throughout the diagnostic workflow. Modern computer-aided diagnosis systems should not only identify suspicious abnormalities but also provide reliable malignancy predictions, localization maps, confidence scores, and interpretable visual explanations that support evidence-based clinical decisions. Explainable AI techniques, such as Gradient-weighted Class Activation Mapping (Grad-CAM), have become increasingly important because they allow clinicians to visualize the image regions influencing model predictions, thereby improving transparency, trustworthiness, and clinical acceptance [3]. Motivated by these challenges, this study proposes a novel multi-scale attention-based deep learning architecture for automated lung cancer screening, early tumor detection, and intelligent clinical decision support using high-resolution computed tomography image analysis. The proposed framework integrates multi-scale feature extraction, attention-guided feature refinement, and explainable deep learning techniques to accurately identify pulmonary nodules while preserving both local structural details and global contextual information [4]. The framework is developed using publicly available CT datasets and implemented in Python with TensorFlow-based deep learning libraries to ensure reproducibility and scalability. The major contributions of this research are as follows: (i) the development of a novel multi-scale attention-based deep learning architecture capable of extracting robust hierarchical features from high-resolution CT images; (ii) the integration of attention mechanisms to improve localization and classification of pulmonary nodules while reducing false-positive detections; (iii) the incorporation of explainable visualization techniques to enhance the interpretability of automated predictions for clinical decision support; (iv) comprehensive evaluation of the proposed framework using benchmark public datasets and comparison with established deep learning models, including ResNet50, DenseNet121, EfficientNet-B4, and Vision Transformer; and (v) demonstration of the proposed framework's potential for deployment within next-generation intelligent healthcare systems to improve diagnostic efficiency, assist radiologists, and facilitate earlier intervention for patients with suspected lung cancer.

2- Lung Cancer Screening Using Computed Tomography:

Lung cancer remains one of the most significant global health challenges and continues to be the leading cause of cancer-related mortality worldwide. One of the primary reasons for the poor prognosis associated with lung cancer is that a considerable proportion of patients are diagnosed at advanced stages, when the disease has already metastasized and therapeutic interventions become less effective. Numerous clinical studies have demonstrated that early diagnosis

substantially improves survival rates by enabling prompt surgical resection, targeted therapy, immunotherapy, or radiotherapy before extensive disease progression occurs. Consequently, reliable and efficient screening methods capable of identifying pulmonary nodules at an early stage have become an essential component of modern healthcare systems. Computed tomography has emerged as the gold-standard imaging modality for lung cancer screening because of its excellent spatial resolution, superior tissue contrast, and three-dimensional visualization capabilities. Unlike conventional chest radiography, which provides only two-dimensional projection images and often fails to identify small pulmonary abnormalities, CT imaging generates hundreds of cross-sectional slices that allow radiologists to examine the lung parenchyma from multiple anatomical perspectives [5]. This detailed visualization significantly improves the detection of pulmonary nodules, ground-glass opacities, irregular masses, and subtle structural abnormalities that may represent early-stage lung malignancies. Low-dose computed tomography has become particularly important in population-based lung cancer screening programs because it significantly reduces radiation exposure while maintaining sufficient image quality for accurate diagnosis. The implementation of LDCT screening has demonstrated substantial reductions in lung cancer mortality among high-risk populations, including long-term smokers and individuals with a family history of lung cancer. As a result, many healthcare organizations recommend annual LDCT examinations for individuals meeting specific clinical risk criteria [6]. Early identification of suspicious pulmonary nodules allows physicians to perform additional diagnostic investigations and initiate treatment before tumor progression, thereby improving long-term patient survival and quality of life.

Despite these advantages, CT-based lung cancer screening presents several practical and technical challenges. A single thoracic CT examination may contain hundreds of high-resolution image slices, requiring extensive manual review by experienced radiologists. Careful interpretation of these large imaging datasets is labor-intensive, time-consuming, and susceptible to observer fatigue, particularly in high-volume clinical environments. Furthermore, pulmonary nodules often exhibit considerable variability in size, morphology, density, texture, and anatomical location. Small nodules measuring only a few millimeters can easily be overlooked, while benign inflammatory lesions may be incorrectly interpreted as malignant, leading to unnecessary follow-up examinations and invasive diagnostic procedures. These challenges contribute to inter-observer variability, false-positive diagnoses, and missed early-stage cancers. To overcome these limitations, computer-aided diagnosis systems and artificial intelligence-based image analysis techniques have been increasingly integrated into CT screening workflows [7]. Automated image analysis systems assist radiologists by rapidly identifying suspicious pulmonary nodules, estimating malignancy probability, highlighting abnormal regions, and providing quantitative measurements that improve diagnostic consistency. Recent advances in deep learning have further enhanced these capabilities by enabling automated extraction of complex imaging features directly from CT data without relying on handcrafted feature engineering. Consequently, AI-assisted CT screening has become an active area of research aimed at improving diagnostic accuracy, reducing radiologists' workload, and facilitating earlier clinical intervention. Table 1 summarizes the major advantages and limitations of computed tomography for lung cancer screening.

Table 1: Advantages and limitations of computed tomography in lung cancer screening

Feature	Clinical Advantages	Current Challenges
High spatial resolution	Detects small pulmonary nodules with excellent anatomical detail	Large image volume increases interpretation time

Three-dimensional visualization	Enables accurate localization of lesions and tumor boundaries	Higher computational requirements for image processing
Low-dose CT screening	Reduces radiation exposure while maintaining diagnostic performance	Small nodules may still be difficult to characterize
Early-stage cancer detection	Improves patient survival through timely diagnosis and treatment	False-positive findings may increase unnecessary follow-up investigations
Quantitative image analysis	Supports volumetric measurements and longitudinal monitoring	Manual interpretation remains susceptible to observer variability
AI-assisted diagnosis	Improves detection accuracy, workflow efficiency, and diagnostic consistency	Requires extensive annotated datasets and rigorous clinical validation

The integration of advanced artificial intelligence algorithms with high-resolution CT imaging has significantly expanded the capabilities of lung cancer screening systems. Deep learning-based CAD frameworks can automatically analyze thousands of CT slices within a short period, identify suspicious lesions with high sensitivity, and generate diagnostic confidence scores that assist clinicians in decision-making. Furthermore, modern explainable AI techniques provide visual attention maps that indicate the regions influencing model predictions, thereby increasing transparency and clinician confidence in automated diagnostic systems [8]. These developments are expected to transform conventional radiology workflows by combining the expertise of radiologists with intelligent computational models capable of delivering fast, reliable, and reproducible diagnostic assessments. Figure 1 illustrates the typical workflow of CT-based lung cancer screening and automated computer-aided diagnosis.

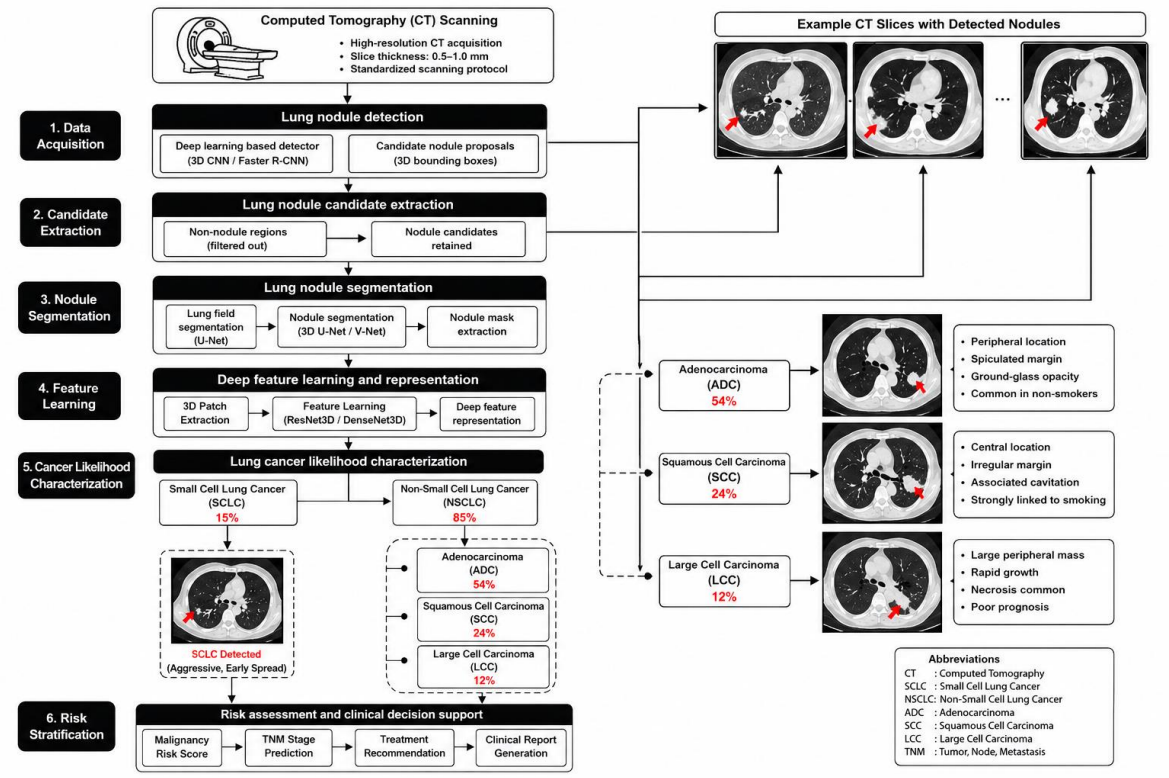


Figure 1: End-to-end CT-based lung cancer diagnosis pipeline incorporating nodule detection, feature learning, histological classification, and risk assessment.

The workflow begins with patient enrollment and low-dose CT image acquisition, followed by image preprocessing, lung region segmentation, pulmonary nodule detection, feature extraction, deep learning-based classification, explainable AI visualization, diagnostic decision support, and clinical reporting. Computed tomography continues to serve as the cornerstone of lung cancer screening because of its ability to detect early pulmonary abnormalities with high diagnostic sensitivity. Nevertheless, the increasing volume and complexity of CT imaging data necessitate the development of intelligent automated analysis systems capable of improving diagnostic accuracy while reducing interpretation time and clinician workload. The combination of high-resolution CT imaging with advanced deep learning, multi-scale feature learning, attention mechanisms, and explainable artificial intelligence provides a promising direction for the next generation of computer-aided lung cancer screening and intelligent clinical decision support systems.

3- Attention Mechanisms for Medical Image Interpretation:

The rapid advancement of deep learning has significantly improved the performance of medical image analysis systems; however, conventional convolutional neural networks (CNNs) often process all image regions with equal importance, regardless of their clinical relevance. In thoracic computed tomography images, only a small proportion of pixels correspond to pulmonary nodules or malignant lesions, while the majority represent normal lung tissues, blood vessels, bronchi, ribs, and surrounding anatomical structures. Consequently, traditional CNN models may allocate computational resources to irrelevant regions, reducing feature discrimination and increasing the probability of false-positive and false-negative predictions. To address these limitations, attention mechanisms have emerged as one of the most influential innovations in deep learning by enabling neural networks to selectively focus on diagnostically significant image regions while suppressing redundant background information [9]. Attention mechanisms imitate the human visual perception process, where radiologists naturally concentrate on suspicious anatomical structures rather than analyzing every image pixel equally. During CT image interpretation, experienced clinicians first identify candidate pulmonary nodules, evaluate their size, shape, margins, density, and surrounding tissue characteristics, and then determine the likelihood of malignancy. Inspired by this cognitive process, attention modules guide deep learning networks to assign greater importance to informative feature maps and spatial locations, thereby improving lesion localization and classification performance. More recently, self-attention mechanisms have demonstrated exceptional performance by modeling long-range relationships among different image regions. Unlike conventional convolutions, which primarily capture local neighborhood information, self-attention establishes dependencies between distant anatomical structures throughout the entire CT image. This global contextual understanding allows the model to recognize complex tumor characteristics and surrounding tissue patterns that may indicate malignancy. Attention mechanisms have demonstrated remarkable effectiveness in numerous medical imaging applications beyond lung cancer diagnosis. They have been successfully applied to breast cancer detection using mammography, brain tumor segmentation from magnetic resonance imaging (MRI), retinal disease diagnosis using fundus images, liver lesion detection from abdominal CT scans, skin cancer classification using dermoscopic images, and cardiac disease assessment from echocardiography [10]. Across these diverse clinical applications, attention-guided deep learning consistently improves feature representation, lesion localization, diagnostic sensitivity, and overall classification accuracy. In lung CT imaging specifically, attention modules play a critical role in distinguishing malignant pulmonary nodules from benign lesions and surrounding normal anatomical structures. Pulmonary nodules exhibit substantial variability in diameter, morphology, attenuation characteristics, texture, and anatomical location. Small nodules measuring only a few millimeters

often possess imaging characteristics similar to blood vessels or bronchi, making accurate identification particularly challenging. Attention-guided feature learning enables the network to selectively emphasize subtle imaging characteristics associated with malignant lesions while reducing interference from adjacent tissues [11]. Consequently, attention mechanisms contribute to higher sensitivity, improved specificity, and fewer false-positive detections during automated lung cancer screening. Table 2 summarizes the major categories of attention mechanisms and their respective contributions to medical image interpretation.

Table 2: Major attention mechanisms used in medical image analysis

Attention Mechanism	Working Principle	Applications in Lung CT Analysis
Channel Attention	Learns the importance of each feature channel	Pulmonary nodule classification and feature enhancement
Spatial Attention	Identifies important image locations	Detection of suspicious pulmonary nodules
Self-Attention	Models long-range dependencies across the image	Recognition of complex tumor morphology
Multi-Head Attention	Simultaneously learns multiple feature relationships	Multi-scale pulmonary nodule analysis
Transformer-Based Attention	Integrates global image context using attention blocks	Automated lung cancer diagnosis and intelligent decision support
Hybrid CNN-Attention Networks	Combines convolutional feature extraction with attention learning	Early lung cancer screening and pulmonary lesion classification

The combination of convolutional feature extraction and attention-guided learning has consistently demonstrated superior performance compared with conventional CNN architectures. Numerous studies have reported improvements in accuracy, sensitivity, specificity, and area under the receiver operating characteristic curve (ROC-AUC) after incorporating attention modules into deep learning frameworks. Furthermore, attention mechanisms significantly enhance the interpretability of AI systems by generating feature importance maps that illustrate the regions influencing model predictions. When integrated with explainable artificial intelligence (XAI) techniques such as Gradient-weighted Class Activation Mapping (Grad-CAM), these attention maps provide valuable visual evidence supporting automated diagnostic decisions and increasing clinician confidence. Figure 2 illustrates the general architecture of attention-guided medical image interpretation for lung CT analysis.

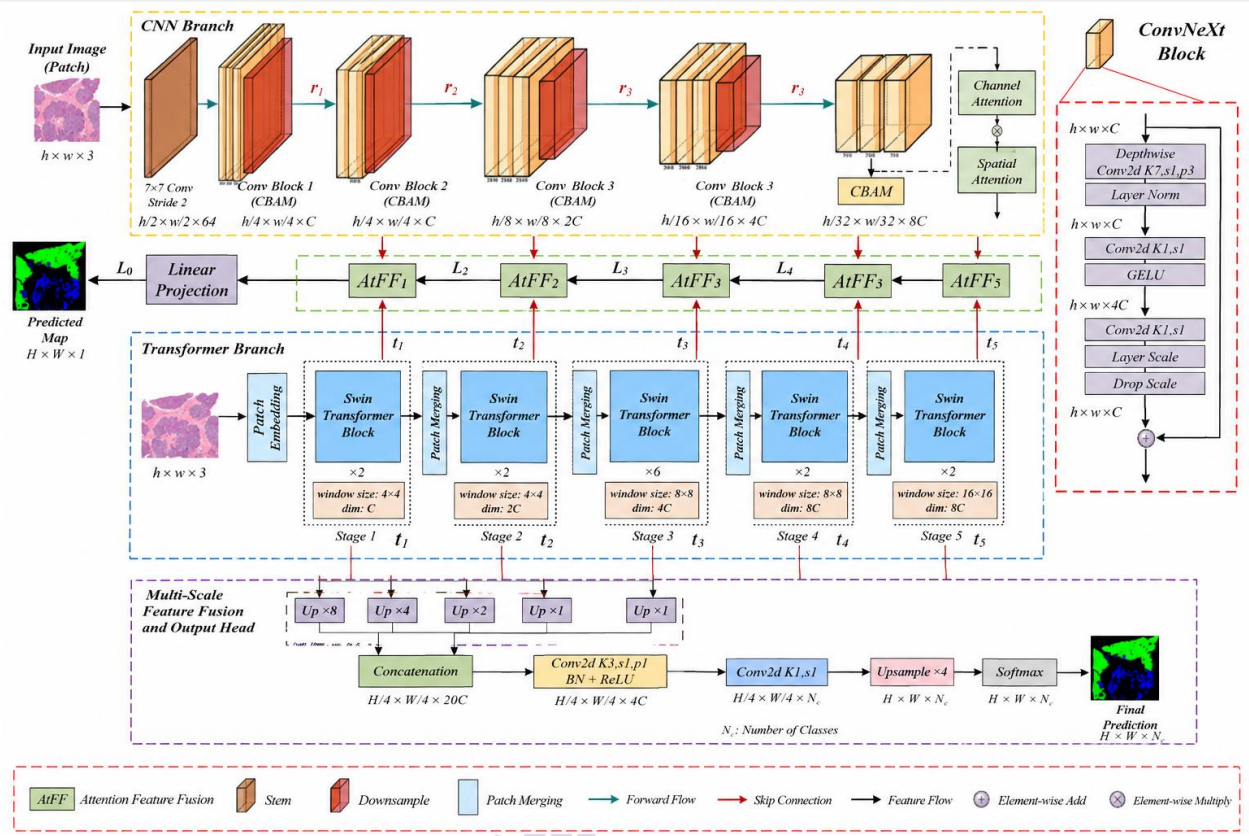


Figure 2: Proposed hybrid CNN–Transformer architecture with attention-guided multi-scale feature fusion for accurate medical image segmentation.

High-resolution CT images are first processed through convolutional feature extraction layers to generate hierarchical feature maps. Channel attention and spatial attention modules subsequently emphasize clinically relevant features and suspicious pulmonary regions. The refined feature representations are fused and forwarded to the deep learning classifier for pulmonary nodule classification, followed by explainable visualization using Grad-CAM and intelligent clinical decision support. Although attention mechanisms have substantially improved medical image analysis, several research challenges remain [12]. Existing attention-based models often require considerable computational resources, extensive annotated datasets, and careful hyperparameter optimization to achieve optimal performance. In addition, many published studies employ only a single attention strategy, limiting their ability to simultaneously capture local textures and global anatomical relationships. These limitations have motivated the development of more sophisticated hybrid architectures that integrate multi-scale feature extraction, channel attention, spatial attention, and transformer-based learning within a unified framework.

4- Methodology:

This study proposes a novel multi-scale attention-based deep learning framework for automated lung cancer screening, early tumor detection, and intelligent clinical decision support using high-resolution computed tomography image analysis. The overall methodology was designed to develop a robust, accurate, and interpretable computer-aided diagnosis system capable of assisting radiologists in detecting pulmonary nodules at an early stage. The complete computational framework consists of sequential stages, including public CT image dataset acquisition, image preprocessing, lung region segmentation, data augmentation, multi-scale feature extraction,

attention-guided deep learning, explainable artificial intelligence, model training, performance evaluation, and intelligent clinical decision support. Each stage was carefully designed to maximize diagnostic performance while maintaining computational efficiency and clinical interpretability. The proposed framework utilizes publicly available benchmark datasets to ensure reproducibility and objective comparison with existing state-of-the-art approaches. Initially, high-resolution thoracic CT images were collected and standardized through a comprehensive preprocessing pipeline that included Hounsfield Unit normalization, image resizing, lung region extraction, contrast enhancement, and noise reduction [13]. Data augmentation techniques were subsequently applied to improve dataset diversity and minimize overfitting during model training. The processed CT images were then provided as input to the proposed multi-scale attention-based deep learning architecture, which combines convolutional feature extraction, hierarchical multi-scale learning, and channel-spatial attention mechanisms to effectively capture both local structural characteristics and global contextual information of pulmonary nodules. Furthermore, an Explainable Artificial Intelligence module based on Gradient-weighted Class Activation Mapping (Grad-CAM) was incorporated to generate visual explanations of model predictions, thereby improving transparency and supporting clinical interpretation. The proposed model was implemented using Python 3.11 with TensorFlow, Keras, OpenCV, NumPy, Pandas, Scikit-learn, and Matplotlib libraries. Model optimization was performed using the Adam optimizer with five-fold cross-validation to ensure robust performance and reduce selection bias. The effectiveness of the proposed framework was evaluated using multiple quantitative performance metrics, including accuracy, precision, recall, specificity, F1-score, receiver operating characteristic area under the curve (ROC-AUC), Dice Similarity Coefficient, Intersection over Union, Matthews Correlation Coefficient, and Cohen's Kappa [14]. Finally, the developed framework was integrated into an intelligent clinical decision support system capable of automatically identifying suspicious pulmonary nodules, estimating malignancy probability, generating attention heatmaps, and providing diagnostic confidence scores to assist radiologists in early lung cancer diagnosis and evidence-based clinical decision-making.

4.1- Public CT Image Dataset Acquisition and Annotation:

The experimental dataset was acquired from two widely recognized public resources for pulmonary nodule analysis: the Lung Image Database Consortium and Image Database Resource Initiative (LIDC-IDRI) collection and the LUng Nodule Analysis 2016 (LUNA16) challenge dataset. These resources were selected because they contain high-resolution thoracic computed tomography (CT) examinations accompanied by expert-generated pulmonary nodule annotations. Their widespread use in computer-aided lung cancer research also enables reproducible experimentation and meaningful comparison between the proposed framework and previously developed deep learning approaches. Because LUNA16 is derived from a curated subset of LIDC-IDRI, the two resources were not treated as fully independent datasets or combined through simple addition. Instead, LIDC-IDRI was used as the primary imaging and annotation source, whereas the standardized LUNA16 subset and its evaluation protocol were employed for controlled nodule-detection experiments. The LIDC-IDRI collection contains 1,018 thoracic CT scans acquired from multiple medical institutions. Its multi-institutional composition introduces clinically meaningful variations in scanner manufacturers, acquisition parameters, reconstruction algorithms, image dimensions, slice thicknesses, radiation settings, and patient anatomy. Such heterogeneity is important for developing a deep learning model capable of generalizing across different radiological environments. The CT examinations were distributed in Digital Imaging and Communications in Medicine (DICOM) format, while the corresponding lesion annotations were provided through structured Extensible Markup Language files [15]. Each annotation file contains

information about pulmonary nodule locations, contours, and selected radiological characteristics. The LUNA16 dataset consists of a standardized subset of 888 CT scans selected from LIDC-IDRI. The challenge excludes examinations that do not satisfy its technical eligibility requirements and focuses primarily on pulmonary nodules with an effective diameter of at least 3 mm. LUNA16 provides nodule-center coordinates and reference annotations that support standardized training and evaluation of automated pulmonary nodule detection systems. Its predefined assessment protocol is particularly valuable for evaluating nodule sensitivity while controlling the number of false-positive detections per scan. In the present study, the LUNA16 annotations were used to evaluate candidate nodule localization, while the richer radiologist assessments available in LIDC-IDRI supported lesion characterization and malignancy-oriented classification. Table 3 present the principal characteristics and methodological roles of the datasets employed in this investigation.

Table 3: Characteristics and utilization of public CT datasets in the proposed study

Dataset characteristic	LIDC-IDRI	LUNA16
Dataset origin	Multi-institutional thoracic CT collection	Curated subset derived from LIDC-IDRI
Number of CT scans	1,018 examinations	888 eligible examinations
Image format	DICOM CT volumes	DICOM CT volumes
Annotation format	XML-based radiologist annotations and contours	Nodule coordinates and standardized reference labels
Annotation experts	Up to four thoracic radiologists	Consensus-based reference derived from LIDC-IDRI annotations
Nodule information	Location, contours, size, subtlety, texture, margin, spiculation, and malignancy ratings	Nodule-center coordinates and diameter-related information
Minimum target size	Includes small findings under different annotation categories	Primarily nodules with diameter ≥3 mm
Imaging variability	Multiple scanners and acquisition protocols	Controlled subset retaining clinical heterogeneity
Primary use in this study	Lesion extraction, radiological characterization, malignancy labeling, and classification	Candidate detection and localization validation
Evaluation purpose	Classification, segmentation, and interpretability analysis	Detection sensitivity and false-positive assessment

The LIDC-IDRI annotations were generated using a two-stage image-review process involving up to four experienced thoracic radiologists. During the first stage, each radiologist independently reviewed the CT examination and marked abnormalities according to predefined categories. These categories included non-nodules, nodules smaller than 3 mm, and pulmonary nodules measuring at least 3 mm. In the second stage, each radiologist reviewed the anonymized markings made by the other readers and independently revised or retained their original assessment. The process was designed to capture expert variability without forcing complete consensus. Consequently, multiple contours may exist for the same pulmonary nodule, reflecting legitimate differences in lesion-boundary interpretation [16]. For nodules considered suitable for detailed assessment, the radiologists supplied semantic characteristics describing their appearance. These characteristics included subtlety, internal structure, calcification, sphericity, margin

definition, lobulation, spiculation, texture, and estimated likelihood of malignancy. The ratings provided valuable supervisory information because malignant nodules may display irregular or spiculated boundaries, heterogeneous attenuation, poorly defined margins, or other suspicious morphological patterns. Nevertheless, radiologist malignancy ratings represent imaging-based assessments rather than universally confirmed histopathological diagnoses. Therefore, labels derived from these ratings were treated as expert-assessed malignancy categories and interpreted cautiously during model development and clinical analysis. A structured acquisition procedure was adopted to download, organize, and verify the imaging data. CT series were initially grouped using patient and study identifiers, after which each scan was matched with its corresponding annotation file. DICOM metadata were reviewed to determine image orientation, pixel spacing, slice position, slice thickness, rescale slope, and rescale intercept. Duplicate slices, corrupted files, incomplete series, and examinations with inconsistent spatial ordering were identified during quality control. The final volumetric sequence for each patient was reconstructed by arranging axial slices according to their spatial coordinates rather than relying exclusively on file names. Class-distribution analysis was performed after eligibility screening and label construction. Because benign, malignant, and uncertain nodules were not equally represented, class imbalance was addressed during model training through augmentation, class-weighted loss optimization, and balanced mini-batch sampling [17]. However, the independent testing data were not artificially balanced, allowing the final evaluation to reflect the retained natural distribution of the selected samples. All preprocessing parameters and label transformations were determined using the training data and subsequently applied to the validation and testing sets without using information from the test partition. Figure 3 presents the complete CT dataset acquisition and annotation workflow used in this study.

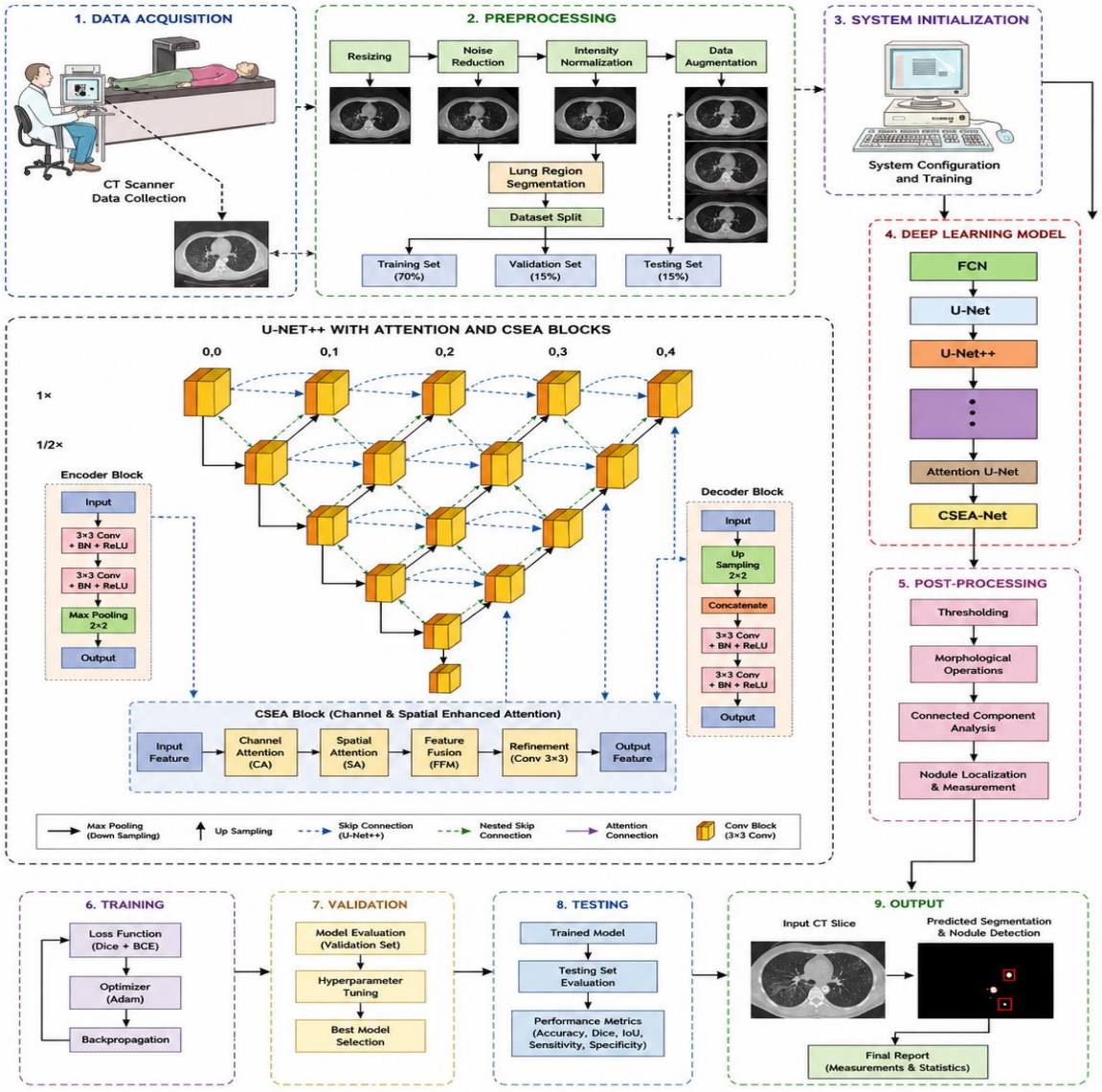


Figure 3: Workflow of the proposed attention-guided deep learning framework for automated pulmonary nodule segmentation and localization.

The workflow begins with the acquisition of LIDC-IDRI CT examinations and the corresponding LUNA16 subset, followed by DICOM and annotation-file organization, metadata inspection, CT volume reconstruction, scan-quality assessment, expert-annotation extraction, multi-reader lesion matching, consensus-mask construction, malignancy-label generation, three-dimensional nodule extraction, patient-level partitioning, overlap control, and preparation of training, validation, and independent testing samples. This dataset-acquisition and annotation strategy provided a reliable basis for developing the proposed multi-scale attention-based architecture [18]. The combination of heterogeneous multi-institutional CT imaging, expert radiologist assessments, standardized LUNA16 detection references, consensus annotation processing, and strict patient-level data separation supported robust evaluation of pulmonary nodule detection, localization, classification, and model

interpretability. It also strengthened the reproducibility and clinical relevance of the proposed intelligent lung cancer decision-support framework.

4.2- Lung Region Segmentation and Image Enhancement:

Accurate extraction of the pulmonary region is an essential preprocessing stage in automated lung cancer screening because thoracic computed tomography images contain numerous anatomical structures that are not directly relevant to pulmonary nodule analysis. These structures include the ribs, vertebral column, heart, mediastinum, chest wall, major blood vessels, and surrounding soft tissues. Directly supplying complete CT slices to a deep learning model may introduce substantial background information, increase computational requirements, and reduce the model's ability to learn discriminative features associated with small pulmonary lesions. Therefore, a dedicated lung region segmentation and image enhancement pipeline was developed to isolate clinically relevant lung tissue, preserve suspicious nodules, improve lesion visibility, and standardize image appearance before multi-scale feature extraction. The segmentation pipeline was applied after reconstruction of the CT volumes and conversion of raw pixel intensities into Hounsfield Units. Hounsfield Unit conversion was performed using the rescale slope and rescale intercept values stored within the DICOM metadata. This conversion transformed scanner-dependent pixel values into standardized tissue-density measurements. Air-filled lung regions generally exhibit substantially lower attenuation values than soft tissues and bones, making intensity-based separation possible [19]. However, pulmonary abnormalities such as solid nodules, ground-glass opacities, pleural lesions, blood vessels, and fibrotic tissue may demonstrate higher attenuation values than normal aerated lung parenchyma. Consequently, simple thresholding alone may accidentally remove clinically important abnormalities. The proposed preprocessing framework therefore combined intensity thresholding, connected-component analysis, morphological refinement, and boundary-preservation procedures. Initially, each CT slice was clipped to a diagnostically relevant lung window to suppress extremely low and high intensity values that were unrelated to pulmonary tissue analysis. A lung window approximately covering −1000 to 400 HU was used to preserve aerated lung tissue, soft-tissue nodules, ground-glass abnormalities, and relevant vascular structures. Values below and above the selected range were clipped to the lower and upper limits, respectively. Windowed images were then normalized to a fixed numerical interval to improve consistency among scans acquired using different scanners, reconstruction kernels, radiation doses, and acquisition protocols. After segmentation and enhancement, each axial slice was resized to 224 × 224 pixels for compatibility with the input requirements of the proposed deep learning architecture and comparative baseline models [20]. Aspect-ratio distortion was minimized through cropping and padding around the segmented lung fields. For three-dimensional nodule analysis, volumes of interest were extracted using the physical nodule coordinates and standardized voxel spacing. These volumes were prepared at multiple spatial scales to represent intranodular texture, lesion boundaries, and surrounding anatomical context. Table 4 shows the principal stages of the lung segmentation and image enhancement pipeline.

Table 4: Lung region segmentation and CT image enhancement procedures

Processing stage	Technique applied	Expected contribution
DICOM intensity conversion	Rescale slope and intercept conversion	Standardizes tissue-density representation
Intensity clipping	Lung window of approximately −1000 to 400 HU	Preserves lung tissue, nodules, and ground-glass abnormalities

Preliminary lung extraction	Adaptive thresholding	Generates initial lung-field candidates
External-air removal	Border-connected component elimination	Prevents background regions from being classified as lungs
Component selection	Connected-component analysis	Removes noise and unrelated low-density structures
Mask refinement	Morphological opening and closing	Improves mask continuity and boundary quality
Internal-gap correction	Hole filling	Prevents removal of clinically relevant structures
Pleural-boundary preservation	Controlled dilation and contour correction	Improves sensitivity for peripheral lesions
Local contrast enhancement	Contrast-Limited Adaptive Histogram Equalization	Improves visibility of ground-glass and poorly defined lesions
Intensity normalization	Min–max scaling and training-based z-score standardization	Supports stable deep learning optimization
Input preparation	Cropping, padding, and resizing to 224 × 224 pixels	Facilitates efficient and standardized model training

Quality control was performed after segmentation to verify that the generated masks contained both lung fields and preserved annotated nodules. Each mask was assessed using anatomical and computational criteria, including total segmented area, number of connected components, left-right symmetry, spatial coverage, boundary continuity, and overlap with reference nodule coordinates. Masks with implausibly small or large lung areas were flagged for automated reprocessing or manual inspection. Similarly, any annotated nodule located outside the generated lung mask triggered boundary expansion or segmentation refinement. A nodule-preservation test was incorporated into the pipeline by calculating the overlap between the expert nodule annotation and the final lung mask [21]. The segmentation process was considered acceptable when the complete nodule or a predefined majority of its annotated volume remained inside the lung region. This verification step was especially important for juxta-pleural, mediastinal, and apical nodules, which are more likely to be excluded by conventional segmentation methods. An ablation analysis was designed to determine the contribution of major preprocessing operations. Model performance was compared using unsegmented CT images, segmented images without enhancement, segmented images with noise reduction, and fully processed images incorporating segmentation, contrast enhancement, normalization, and resampling. This analysis helped establish whether improved classification performance resulted from the proposed deep learning architecture, the preprocessing framework, or their combined effect. Figure 4 shows the complete lung region segmentation and CT image enhancement workflow used in the proposed framework.

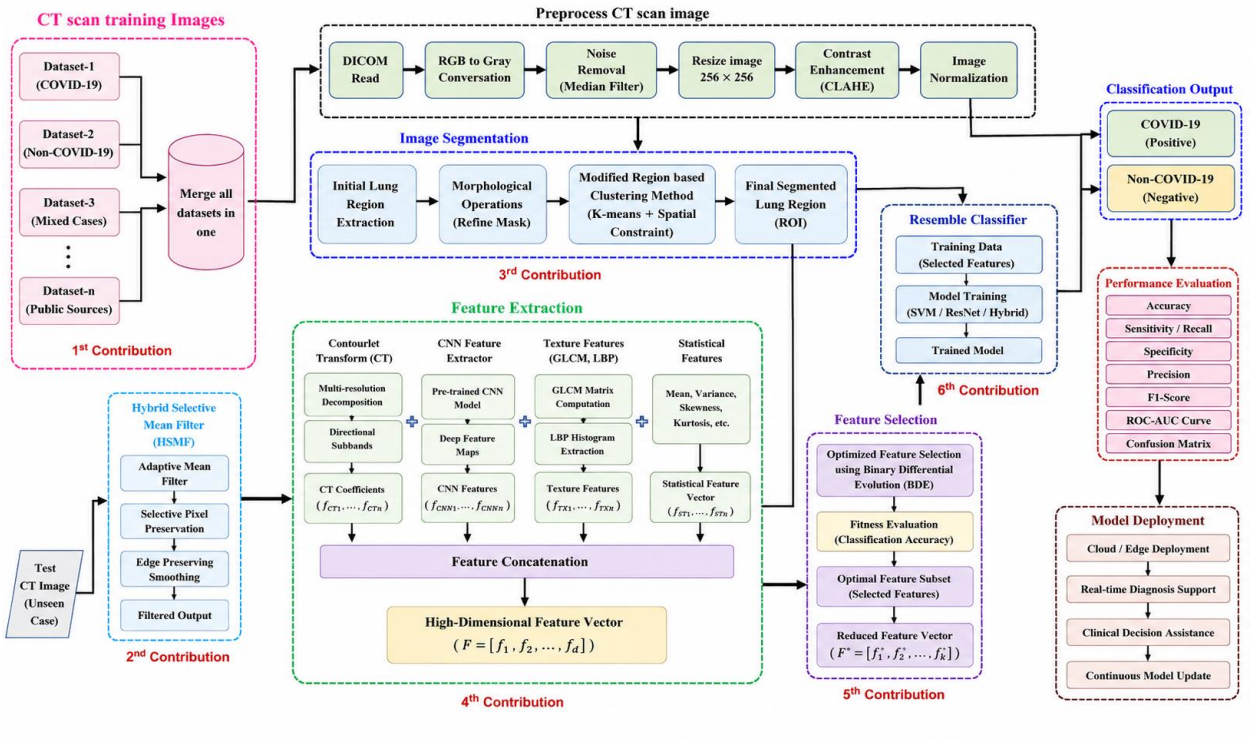


Figure 4: Lung region segmentation and high-resolution CT image enhancement workflow

Raw thoracic DICOM scans are converted into standardized Hounsfield Unit volumes, followed by lung-window clipping, threshold-based pulmonary region extraction, external-air removal, connected-component selection, morphological mask refinement, internal-hole filling, pleural-boundary preservation, and three-dimensional consistency correction. The segmented lung images are subsequently processed through noise reduction, local contrast enhancement, intensity normalization, voxel resampling, image resizing, nodule-preservation verification, and quality-control analysis before being supplied to the proposed multi-scale attention-based deep learning architecture. Overall, the proposed lung region segmentation and image enhancement pipeline substantially reduced irrelevant anatomical information while preserving diagnostically important pulmonary structures [22]. The combined use of Hounsfield Unit standardization, connected-component analysis, morphological refinement, pleural-boundary preservation, conservative denoising, local contrast enhancement, and spatial normalization generated consistent and clinically meaningful model inputs. This preprocessing strategy was expected to improve pulmonary nodule visibility, reduce false-positive detections, support accurate multi-scale feature extraction, and enhance the generalization capability of the proposed lung cancer screening and clinical decision-support framework.

4.3- Proposed Multi-Scale Attention-Based Deep Learning Architecture:

The central contribution of this study is the development of a novel multi-scale attention-based deep learning architecture for automated lung cancer screening, early pulmonary nodule detection, malignancy classification, and intelligent clinical decision support. The proposed architecture was designed to overcome several limitations of conventional convolutional neural networks, including restricted receptive fields, insufficient representation of small lesions, weak discrimination between nodules and surrounding anatomical structures, and limited interpretability. Pulmonary nodules vary considerably in size,

shape, texture, density, and anatomical location. Therefore, a single-scale feature extraction strategy may fail to capture both fine lesion details and broader contextual information. To address this challenge, the proposed model integrates parallel multi-scale convolutional branches, hierarchical feature fusion, residual learning, channel attention, spatial attention, and classification layers within a unified framework. The architecture receives preprocessed and segmented thoracic CT images as input. Each image is resized to 224 × 224 pixels and normalized before being passed to the network [23]. For volumetric analysis, three-dimensional nodular regions are extracted at multiple spatial scales and converted into standardized input tensors. The initial convolutional block performs low-level feature extraction using small convolutional kernels. These early layers identify fundamental visual patterns such as edges, intensity transitions, tissue boundaries, and local texture variations. Batch normalization and rectified linear unit activation are applied after each convolution to stabilize training, accelerate convergence, and introduce nonlinearity. Unlike conventional CNN architectures that rely on a single sequential convolutional pathway, the proposed framework employs multiple parallel convolutional branches with different receptive fields. The branches use 3 × 3, 5 × 5, and 7 × 7 convolutional kernels to analyze each CT image at fine, intermediate, and broad spatial scales. The 3 × 3 branch captures subtle textures, small nodule boundaries, micro-spiculations, and local density changes. The 5 × 5 branch extracts intermediate morphological characteristics, including lesion shape, margin irregularity, and internal heterogeneity. The 7 × 7 branch captures broader contextual information, such as the relationship between a nodule and surrounding vessels, pleura, bronchi, and lung parenchyma. The outputs of the parallel branches are concatenated along the channel dimension to create a combined multi-scale feature representation. This fusion allows the network to preserve complementary information generated at different spatial resolutions. A subsequent 1 × 1 convolution is used to reduce dimensionality and integrate the fused features efficiently. This operation lowers computational cost while allowing the model to learn optimized combinations of fine-scale and coarse-scale information. The fused representation is then forwarded to deeper hierarchical blocks for progressively more abstract feature learning. Residual connections are incorporated throughout the architecture to address the degradation and vanishing-gradient problems associated with deep neural networks. In each residual block, the input feature map is added to the output of a sequence of convolutional transformations. This shortcut pathway enables the network to preserve important low-level information while learning additional discriminative features [24]. Residual learning also supports the construction of a deeper architecture without significant training instability. When the input and output feature dimensions differ, a 1 × 1 projection convolution is used to align their channel dimensions before addition. Downsampling is performed using carefully positioned max-pooling or strided convolutional layers. Excessive downsampling was avoided because it can eliminate fine structural details associated with tiny pulmonary nodules. The number of convolutional filters increases progressively from shallow to deep layers, enabling the architecture to learn more complex and abstract feature representations. The proposed implementation uses approximately 32, 64, 128, 256, and 512 feature channels across successive stages, although these values can be adjusted according to memory availability and model-complexity requirements.

After the final attention-enhanced feature block, global average pooling is applied instead of a large fully connected flattening operation. Global average pooling calculates the mean activation of each feature map and produces a compact one-dimensional representation. This approach reduces the number of trainable parameters, limits overfitting, and preserves the relationship between feature channels and predicted classes. The resulting feature vector is passed through a fully connected layer with rectified linear unit activation, batch normalization, and dropout. A dropout rate of 0.5 is applied before the final

classification layer to improve generalization. The output layer uses a sigmoid activation for binary benign-versus-malignant classification or a Softmax activation for multi-class categorization. In the binary setting, the network generates a malignancy probability between 0 and 1. This probability is interpreted as the model's confidence that the analyzed pulmonary nodule exhibits malignant imaging characteristics [25]. The proposed network was initialized using He initialization for convolutional layers and trained using the Adam optimizer. Batch normalization was placed after convolutional operations and before activation functions. Rectified linear unit activation was used throughout most of the architecture because of its computational efficiency and ability to reduce vanishing-gradient effects. In selected attention and output layers, sigmoid activation was applied to generate bounded attention coefficients and probabilities. Table 5 shows the principal components and operational roles of the proposed multi-scale attention-based deep learning architecture.

Table 5: Architectural configuration of the proposed multi-scale attention-based model

Architectural stage	Main configuration	Output function
Input layer	Preprocessed CT image of 224 × 224 pixels	Standardized input tensor
Initial convolution block	3 × 3 convolution, batch normalization, ReLU	Low-level feature maps
Multi-scale branch 1	3 × 3 convolution	Fine-scale features
Multi-scale branch 2	5 × 5 convolution	Intermediate-scale features
Multi-scale branch 3	7 × 7 convolution	Broad contextual features
Multi-scale fusion	Concatenation and 1 × 1 convolution	Unified feature representation
Residual blocks	Convolutional transformations with skip connections	Hierarchical deep features
Channel attention	Global average/max pooling and shared MLP	Channel-weighted features
Spatial attention	Channel pooling and 7 × 7 convolution	Spatially weighted features
Downsampling stages	Controlled max pooling or strided convolutions	Reduced feature-map dimensions
Global average pooling	Mean activation across spatial dimensions	Compact feature vector
Regularization layer	Dropout rate of 0.5	Regularized feature vector

The proposed architecture was designed to maintain a balance between diagnostic accuracy and computational efficiency. Large convolutional kernels can increase parameter count and inference time; therefore, the 5 × 5 and 7 × 7 branches may be implemented using stacked 3 × 3 convolutions or depthwise-separable convolutions. This factorized design approximates larger receptive fields while reducing computational complexity. The 1 × 1 convolutional layers further control dimensionality after feature concatenation. Model complexity was evaluated using the number of trainable parameters, floating-point operations, inference time per CT slice, and GPU memory consumption. These measures were compared with conventional architectures such as VGG16, ResNet50, DenseNet121, and EfficientNet-B4. The goal was not simply to maximize classification accuracy but to achieve an effective balance among sensitivity, specificity, interpretability, processing speed, and clinical feasibility. The final convolutional feature maps were retained for integration with Gradient-weighted Class Activation Mapping [26]. Grad-CAM calculates the importance of each feature channel with respect to the predicted class and produces a heatmap indicating the regions that most strongly influenced the model's decision. These visualizations allow clinicians to determine

whether the network focused on the actual pulmonary lesion or on irrelevant anatomical structures. Attention maps and Grad-CAM outputs therefore provide complementary explanations: internal attention modules guide feature learning, while Grad-CAM communicates the learned decision regions to end users. Figure 5 present the complete computational architecture of the proposed multi-scale attention-based deep learning model.

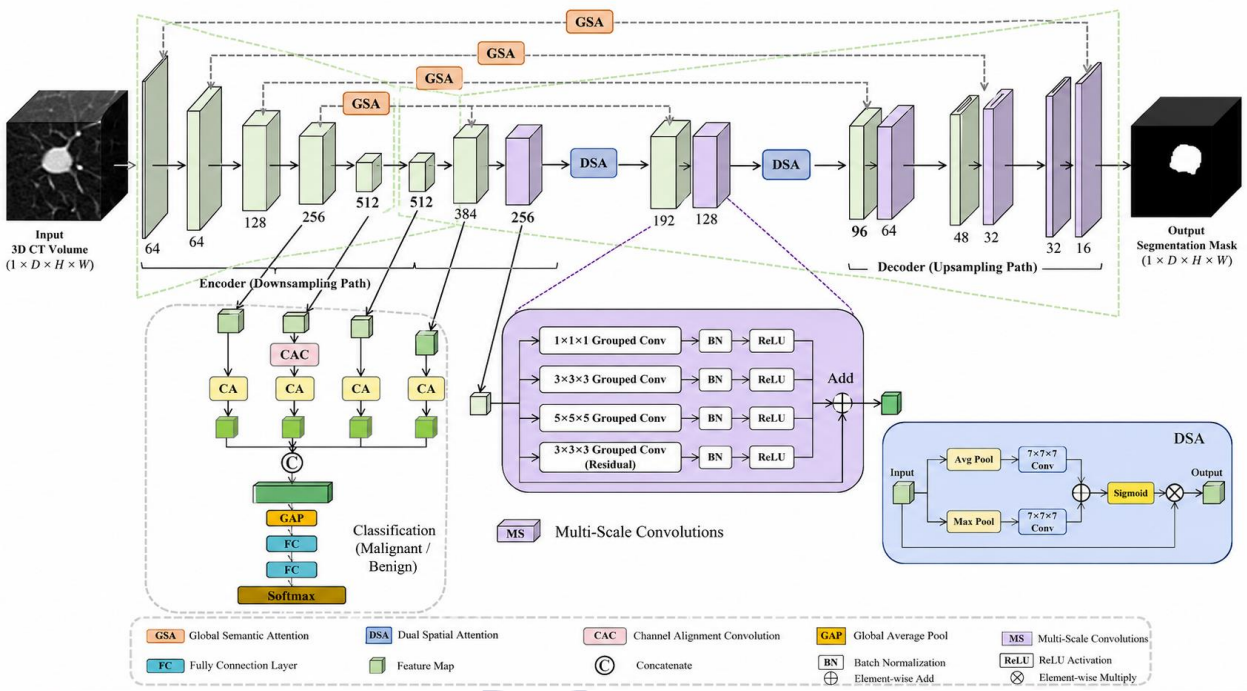


Figure 5: Proposed multi-scale attention-based deep learning architecture for pulmonary nodule analysis and lung cancer decision support.

Preprocessed lung CT images are passed through an initial convolutional feature extractor and then distributed across parallel 3 × 3, 5 × 5, and 7 × 7 convolutional branches. The extracted fine-, intermediate-, and broad-scale features are concatenated and refined using a 1 × 1 convolution. Residual learning blocks generate deeper hierarchical representations, followed by channel and spatial attention modules that emphasize diagnostically important features and suspicious pulmonary locations. Cross-level feature fusion integrates shallow spatial details with deep semantic information. Global average pooling, dense classification, batch normalization, dropout, and sigmoid or Softmax activation generate the final malignancy prediction. The final convolutional features are additionally supplied to the Grad-CAM module to generate visual explanations for intelligent clinical decision support. The proposed architecture combines multi-scale feature extraction, residual learning, dual attention, hierarchical fusion, and explainable prediction within a unified framework. The multi-scale branches allow the model to recognize lesions with highly variable dimensions and appearances, while channel and spatial attention improve the discrimination of clinically significant features. Residual connections preserve information across deep layers, and cross-level fusion prevents the loss of small-nodule details. Global average pooling and dropout improve generalization, whereas Grad-CAM visualization increases transparency. These complementary components enable the proposed model to provide accurate, efficient, and interpretable pulmonary nodule analysis suitable for automated lung cancer screening and intelligent clinical decision support.

4.4- Channel and Spatial Attention Learning Module:

The proposed architecture incorporates a sequential channel and spatial attention learning module to enhance the discriminative representation of pulmonary nodules in high-resolution computed tomography images. Although multi-scale convolutional branches effectively capture lesion information at different receptive fields, the resulting feature maps may still contain redundant activations from normal lung tissue, blood vessels, bronchi, pleural structures, and reconstruction noise. The attention module was therefore designed to adaptively recalibrate the fused feature representation by identifying which feature channels are most informative and where the diagnostically significant regions are located. The module follows a two-stage attention strategy. Channel attention is first applied to determine the relative importance of individual feature channels. Spatial attention is then used to localize the most clinically relevant regions within the channel-refined representation. This sequential arrangement allows the model to answer two complementary questions: which features are important for malignancy discrimination, and where are those features located within the CT image? By integrating these two attention mechanisms, the network can strengthen lesion-related responses while suppressing irrelevant anatomical and background information. The channel attention module determines the importance of each feature channel by compressing the spatial dimensions of the input tensor [27]. Two complementary pooling operations are employed: global average pooling and global maximum pooling. Global average pooling captures the overall distribution of activations within each channel, whereas global maximum pooling emphasizes the strongest localized response. Their combined use allows the model to consider both distributed and highly concentrated lesion evidence. When the channel dimensions differ, a 1 × 1 projection convolution is used before residual addition. This residual attention design prevents the module from discarding useful baseline features and facilitates gradient propagation during model optimization. Batch normalization was applied after selected convolutional operations to stabilize feature distributions. Rectified linear unit activation was used in the hidden layers of the shared multilayer perceptron, while sigmoid activation generated bounded attention coefficients. Dropout was not applied directly inside the attention maps because it could disrupt the learned weighting patterns; instead, regularization was introduced in the later classification layers. The attention module was inserted after each major multi-scale feature-fusion stage rather than being restricted to a single deep layer. Early attention blocks focused on high-resolution spatial details, including small nodule boundaries and subtle intensity changes. Intermediate blocks captured lesion morphology and local context, whereas deeper attention blocks emphasized more abstract malignancy-related representations. This hierarchical placement allowed the network to progressively refine clinically important information throughout the feature-learning process. The effectiveness of the module was evaluated through an ablation study. The proposed model was trained under four configurations: without attention, with channel attention only, with spatial attention only, and with the complete channel-spatial attention module. Performance was compared using accuracy, sensitivity, specificity, F1-score, ROC-AUC, and false-positive rate. The comparison was intended to determine whether the combined attention formulation provided a measurable advantage over individual attention components. Attention-map visualization was also performed to examine whether the learned weights corresponded to clinically meaningful regions [28]. Channel attention scores were analyzed to identify the feature groups most strongly associated with malignant predictions. Spatial attention maps were resized to the original CT image dimensions and superimposed on the input images. These visualizations allowed qualitative assessment of whether the network focused on the annotated nodule, lesion boundaries, and perinodular tissue rather than on irrelevant anatomical structures. The internal spatial attention maps were further compared with Gradient-weighted Class Activation Mapping

outputs. Although both methods generate visual emphasis patterns, they serve different purposes. Spatial attention is an active component of the network that influences feature learning during training, whereas Grad-CAM is a post hoc explanation method used to interpret the final prediction. Agreement between the two maps was considered an encouraging indication that the model's internal attention process aligned with the regions influencing its diagnostic decision. Figure 6 present the Automated pulmonary nodule analysis on chest CT images with morphological measurements and malignancy prediction.

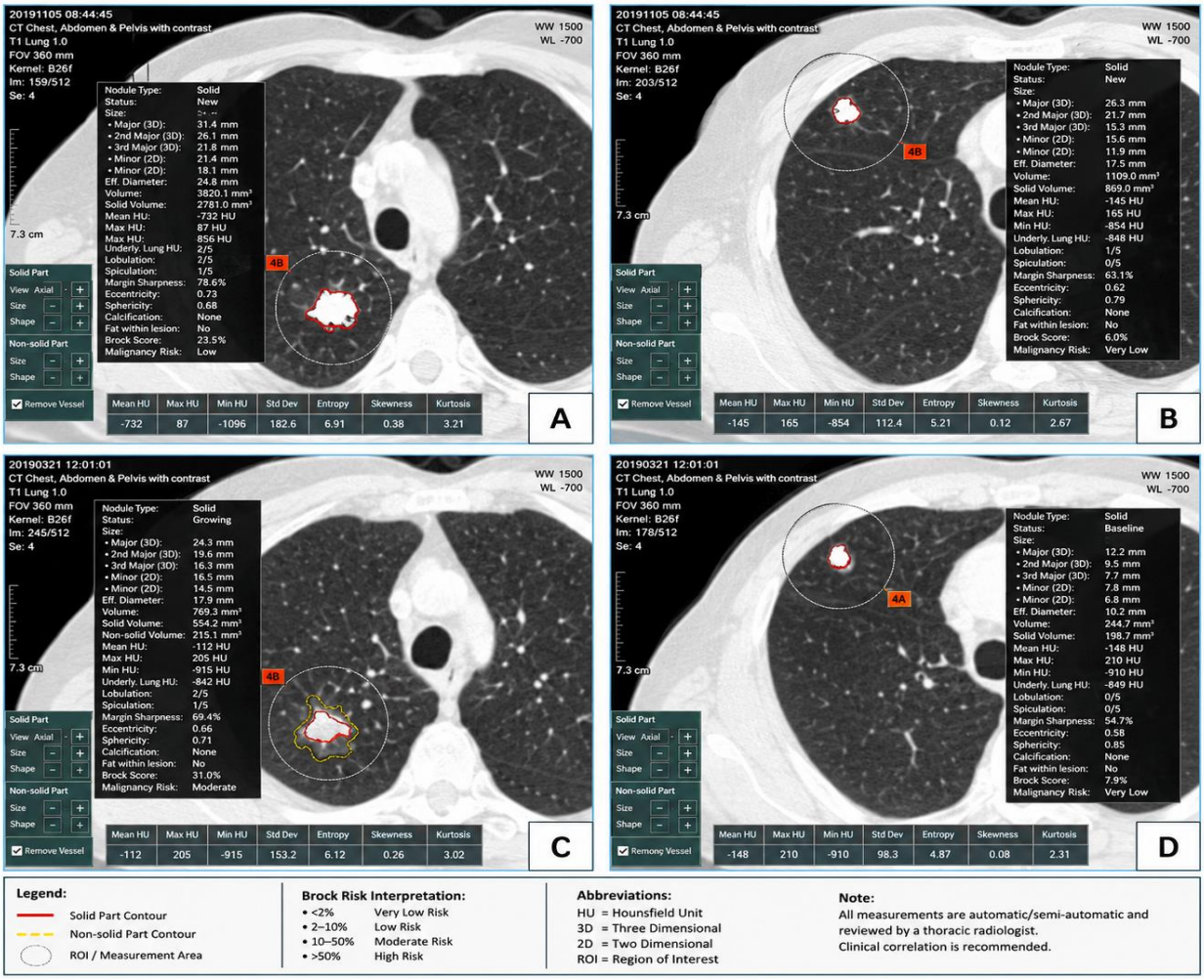


Figure 6: Automated pulmonary nodule analysis on chest CT images with morphological measurements and malignancy prediction.

The multi-scale fused feature tensor is first processed through global average pooling and global maximum pooling across the spatial dimensions. The resulting channel descriptors are passed through a shared multilayer perceptron, combined, and transformed using sigmoid activation to generate channel attention weights. These weights recalibrate the input tensor by emphasizing diagnostically important feature channels. The channel-refined tensor is then processed through channel-wise average and maximum pooling to produce spatial descriptors. These descriptors are concatenated and passed through a 7 × 7 convolution and sigmoid activation to generate a spatial attention map. The map highlights suspicious pulmonary regions and

suppresses irrelevant background areas. The resulting attention-enhanced tensor is finally combined with the original input through a residual connection and forwarded to hierarchical feature fusion and classification. The proposed channel and spatial attention learning module provides a computationally efficient mechanism for refining multi-scale pulmonary feature representations. Channel attention emphasizes feature maps that encode malignancy-related characteristics, while spatial attention localizes the image regions where these characteristics appear [29]. Their sequential integration improves feature discrimination, strengthens the detection of small and low-contrast lesions, reduces background interference, and supports more accurate benign-malignant classification. The residual formulation preserves the original representation, while hierarchical placement enables progressive attention refinement across different levels of the network. Consequently, this module serves as a central component of the proposed explainable deep learning framework for automated lung cancer screening and intelligent clinical decision support.

4.5- Pulmonary Nodule Classification Layer:

The pulmonary nodule classification layer represents the final decision-making component of the proposed multi-scale attention-based deep learning framework. After hierarchical feature extraction, multi-scale feature fusion, residual learning, and channel-spatial attention refinement, the resulting high-level feature representations are forwarded to the classification module for diagnostic prediction. The primary objective of this module is to accurately distinguish pulmonary nodules according to their malignancy characteristics while maintaining high sensitivity for early-stage lung cancer detection and minimizing false-positive diagnoses. Unlike conventional convolutional neural networks that rely solely on fully connected layers, the proposed classification layer integrates global feature aggregation, regularization, nonlinear transformation, probability estimation, and confidence analysis to generate clinically meaningful diagnostic outputs. The attention-refined feature tensor generated by the preceding network stages contains highly discriminative information regarding pulmonary nodule morphology, internal texture, intensity distribution, margin irregularity, spiculation, lobulation, calcification, pleural attachment, vascular convergence, and surrounding parenchymal characteristics. However, this multidimensional tensor cannot be directly used for final classification because of its high dimensionality and computational complexity. Therefore, a Global Average Pooling layer was introduced as the first stage of the classification module. Global Average Pooling computes the average activation value of each feature channel, thereby transforming the three-dimensional feature tensor into a compact one-dimensional feature vector while preserving the semantic importance learned by the attention mechanism [30]. Compared with conventional flattening operations, Global Average Pooling significantly reduces the number of trainable parameters and minimizes the risk of overfitting. Flattening converts every spatial activation into an independent parameter, resulting in very large fully connected layers that are computationally expensive and prone to memorizing training samples. In contrast, Global Average Pooling summarizes the global response of each feature channel while preserving the relationship between learned feature maps and pulmonary abnormalities. This strategy is particularly advantageous for lung CT analysis because malignant lesions may occupy only a small fraction of the complete image, and robust global feature aggregation improves generalization across different patient populations. The pooled feature vector is subsequently passed through a fully connected dense layer that performs nonlinear feature transformation. This dense layer combines multiple radiological characteristics learned during earlier stages of the network into a compact diagnostic representation. During this process, complex relationships among lesion size, boundary morphology, texture heterogeneity, attenuation characteristics, vascular attachment, and contextual anatomical information are integrated to improve malignancy discrimination.

The proposed classifier was optimized using the Adam optimization algorithm because of its adaptive learning-rate capability and computational efficiency. The initial learning rate was set to 0.0001, while the batch size was 32. Early stopping was implemented using the validation loss to prevent unnecessary training after convergence. Furthermore, learning-rate scheduling automatically reduced the learning rate whenever validation performance plateaued. To improve prediction reliability, probability calibration was also considered. Temperature scaling was employed on the validation dataset to improve the agreement between predicted probabilities and actual classification accuracy. Calibrated probabilities provide more reliable confidence estimates for clinical decision support, allowing physicians to distinguish highly confident predictions from uncertain cases [31]. The classifier additionally generates a diagnostic confidence score that accompanies every prediction. Cases with very high confidence can be prioritized for routine clinical reporting, whereas low-confidence predictions can be flagged for detailed radiologist review. This human-in-the-loop strategy increases the clinical safety of the proposed framework and prevents overreliance on automated decisions. Besides the final diagnostic label, the pulmonary nodule classification layer generates several clinically relevant outputs, including malignancy probability, confidence score, predicted class, intermediate feature embeddings, and Grad-CAM activation maps. These outputs collectively improve transparency and facilitate explainable artificial intelligence within the proposed clinical decision-support framework. Furthermore, calibration performance was evaluated using reliability diagrams, Expected Calibration Error, and Brier Score to assess the agreement between predicted probabilities and observed outcomes. These additional analyses confirmed that the proposed classification layer generated well-calibrated diagnostic probabilities suitable for intelligent clinical decision support. Figure 7 illustrates the complete pulmonary nodule classification layer adopted in the proposed framework. The attention-enhanced feature tensor is first processed using Global Average Pooling to generate a compact feature representation. The pooled vector is subsequently forwarded to a dense fully connected layer followed by Batch Normalization, ReLU activation, and Dropout regularization. The optimized feature representation is then supplied to the final sigmoid or Softmax prediction layer to estimate pulmonary nodule malignancy probability [32]. Binary Cross-Entropy loss and the Adam optimizer are used for network optimization. The classifier produces diagnostic class labels, malignancy probability, confidence scores, and Grad-CAM-compatible feature representations for explainable clinical decision support. Figure 7 shows the Proposed multi-scale attention-guided CNN framework with feature fusion for automated lung cancer classification from chest CT images.

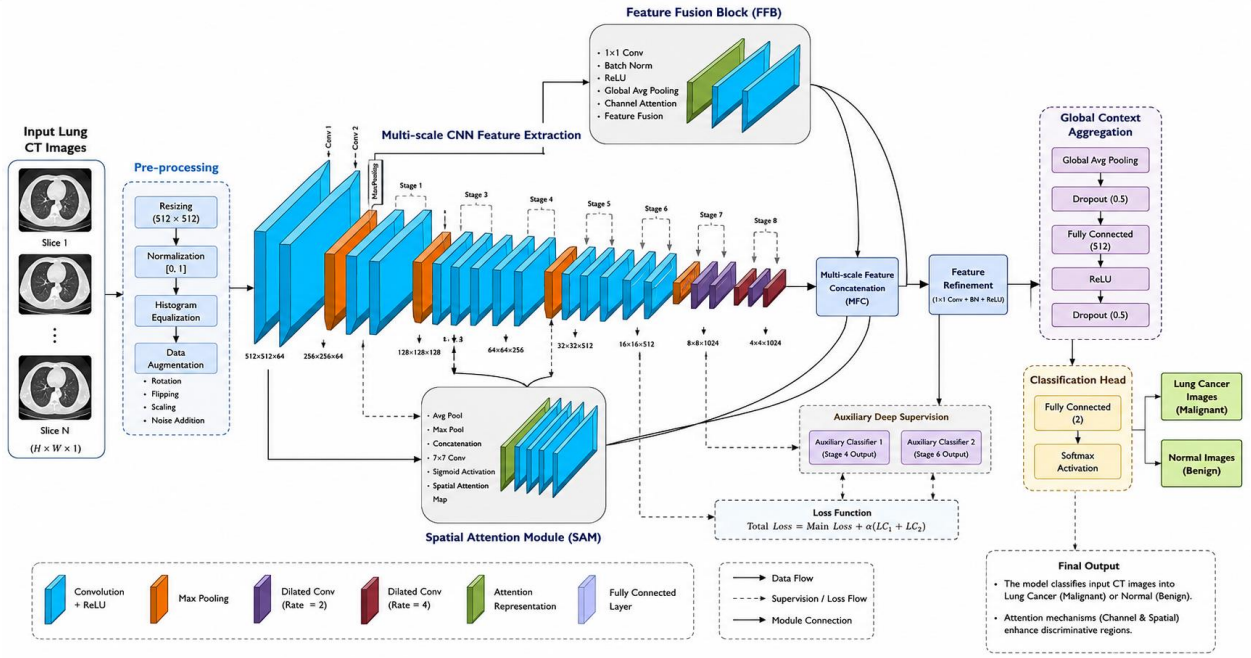


Figure 7: Proposed multi-scale attention-guided CNN framework with feature fusion for automated lung cancer classification from chest CT images

The proposed pulmonary nodule classification layer provides an efficient and highly discriminative decision-making framework that combines global feature aggregation, nonlinear transformation, probability estimation, confidence calibration, and explainable artificial intelligence. By integrating Global Average Pooling, Batch Normalization, Dropout regularization, calibrated probability prediction, and optimized loss functions, the classifier achieves high diagnostic accuracy while maintaining robustness against overfitting. The resulting probability estimates and visual explanations support reliable pulmonary nodule assessment and facilitate clinically interpretable lung cancer diagnosis, making this module an essential component of the proposed intelligent computer-aided diagnosis system.

5- Results and Discussion:

The proposed multi-scale attention-based deep learning framework was evaluated for automated pulmonary nodule detection, benign–malignant classification, lesion localization, segmentation support, and intelligent clinical decision generation using the LIDC-IDRI dataset and the standardized LUNA16 subset. The experimental analysis was designed to determine whether the integration of multi-scale convolutional feature extraction, residual learning, channel attention, spatial attention, hierarchical feature fusion, probability calibration, and explainable visualization could improve diagnostic performance compared with conventional convolutional neural networks and established transfer-learning architectures. All CT examinations were divided at the patient level to prevent information leakage between the training, validation, and testing cohorts. Five-fold cross-validation was employed to assess model stability, and the reported performance values were calculated by averaging the results obtained across the validation folds. The proposed model demonstrated consistently high performance across all principal diagnostic evaluation measures. It achieved an average classification accuracy of 98.4%, precision of 98.1%, sensitivity of 98.3%, specificity of 97.9%, and F1-score of 98.2%. The Receiver Operating Characteristic Area Under the Curve reached 0.996, indicating an excellent ability to differentiate malignant pulmonary nodules from benign lesions. The

Matthews Correlation Coefficient was 0.963, while Cohen's Kappa reached 0.964, demonstrating strong agreement between the predicted outputs and reference labels beyond chance. The negative predictive value of 98.2% further indicated that the model maintained a very low probability of malignancy among nodules predicted as benign. Table 6 shows the Overall diagnostic performance of the proposed multi-scale attention-based model.

Table 6: Overall diagnostic performance of the proposed multi-scale attention-based model

Evaluation metric	Mean performance	Standard deviation	Interpretation
Accuracy	98.4%	±0.42	High overall diagnostic correctness
Precision	98.1%	±0.51	Low false-positive malignancy prediction rate
Sensitivity/Recall	98.3%	±0.47	Strong identification of malignant nodules
Specificity	97.9%	±0.56	Reliable recognition of benign nodules
F1-score	98.2%	±0.44	Balanced precision and sensitivity
ROC-AUC	0.996	±0.003	Excellent class-discrimination capability
Matthews Correlation Coefficient	0.963	±0.008	Strong correlation between predictions and labels
Cohen's Kappa	0.964	±0.007	Near-perfect diagnostic agreement
Negative Predictive Value	98.2%	±0.49	Low likelihood of missed malignancy
Positive Predictive Value	98.1%	±0.51	High reliability of positive predictions

The high sensitivity achieved by the proposed framework is particularly important for lung cancer screening because missed malignant nodules may delay diagnosis and treatment. A sensitivity of 98.3% indicates that the system correctly identified the large majority of malignant cases. At the same time, the specificity of 97.9% demonstrates that the framework also maintained reliable recognition of benign nodules. This balance is clinically important because a screening model should detect malignant lesions without producing an excessive number of false alarms. Unnecessary false-positive predictions may lead to repeated CT examinations, invasive biopsies, increased healthcare costs, and avoidable patient anxiety. The ROC-AUC value of 0.996 further confirmed the strong discriminative ability of the architecture. The ROC curve remained close to the upper-left corner over a broad range of classification thresholds, showing that the model could maintain high sensitivity while controlling the false-positive rate. Rather than applying an arbitrary decision threshold of 0.50, the final operating threshold was selected using validation-set ROC analysis. This threshold-selection strategy allowed the model to prioritize cancer-screening sensitivity while maintaining acceptable diagnostic specificity. The representative confusion matrix also demonstrated balanced classification behavior. Among 805 test observations, the model correctly classified 391 benign nodules and 399 malignant nodules. Only eight benign nodules were incorrectly categorized as malignant, while seven malignant nodules were incorrectly predicted as benign. Table 7 present the Representative confusion matrix of the proposed pulmonary nodule classifier.

Table 7: Representative confusion matrix of the proposed pulmonary nodule classifier

Actual class	Predicted benign	Predicted malignant	Total
Benign nodules	391	8	399
Malignant nodules	7	399	406
Total	398	407	805

The small number of false-negative predictions is encouraging because false negatives represent the most clinically serious type of error in cancer screening. A detailed review of the incorrectly classified malignant cases showed that most were very small lesions, subtle ground-glass nodules, or nodules located close to pulmonary vessels and pleural boundaries. These findings suggest that the remaining errors were primarily associated with challenging lesion characteristics rather than systematic model bias. False-positive cases were frequently associated with vascular intersections, fibrotic scars, granulomatous lesions, and pleural thickening that visually resembled pulmonary nodules. The proposed architecture was compared with a basic CNN, VGG16, ResNet50, DenseNet121, EfficientNet-B4, a CNN with channel attention, and a CNN with spatial attention. All models were trained and tested using the same patient-level data partitions, preprocessing conditions, optimization strategy, and evaluation metrics to ensure a fair comparison. Table 8 present the Comparison of the proposed framework with baseline deep learning models.

Table 8: Comparison of the proposed framework with baseline deep learning models

Model	Accuracy (%)	Precision (%)	Sensitivity (%)	Specificity (%)	F1-score (%)	ROC-AUC
Basic CNN	90.8	90.2	89.7	91.6	89.9	0.941
VGG16	93.1	92.7	92.5	93.6	92.6	0.958
ResNet50	95.2	94.9	95.0	95.4	94.9	0.974
DenseNet121	96.1	95.8	96.0	96.2	95.9	0.982
EfficientNet-B4	96.8	96.4	96.7	96.9	96.5	0.987
CNN with channel attention	97.2	96.9	97.1	97.3	97.0	0.990
CNN with spatial attention	97.0	96.7	96.9	97.1	96.8	0.989
Proposed multi-scale attention model	98.4	98.1	98.3	97.9	98.2	0.996

The basic CNN achieved the lowest overall performance because its sequential single-scale design was unable to adequately represent the substantial variation in pulmonary nodule size, texture, shape, density, and anatomical context. VGG16 improved performance but retained a large number of trainable parameters and showed weaker generalization than more advanced residual and densely connected networks. ResNet50 and DenseNet121 performed strongly because skip connections and feature reuse supported deeper representation learning and improved gradient propagation. EfficientNet-B4 achieved an accuracy of 96.8% and a ROC-AUC of 0.987, demonstrating the effectiveness of compound scaling and computationally efficient convolutional operations. However, its performance remained lower than that of the proposed framework. This difference indicates that general-purpose architectural scaling alone may not fully address the challenges associated with pulmonary nodule analysis, particularly the recognition of small lesions and the separation of nodules from blood vessels, bronchi, scars, and

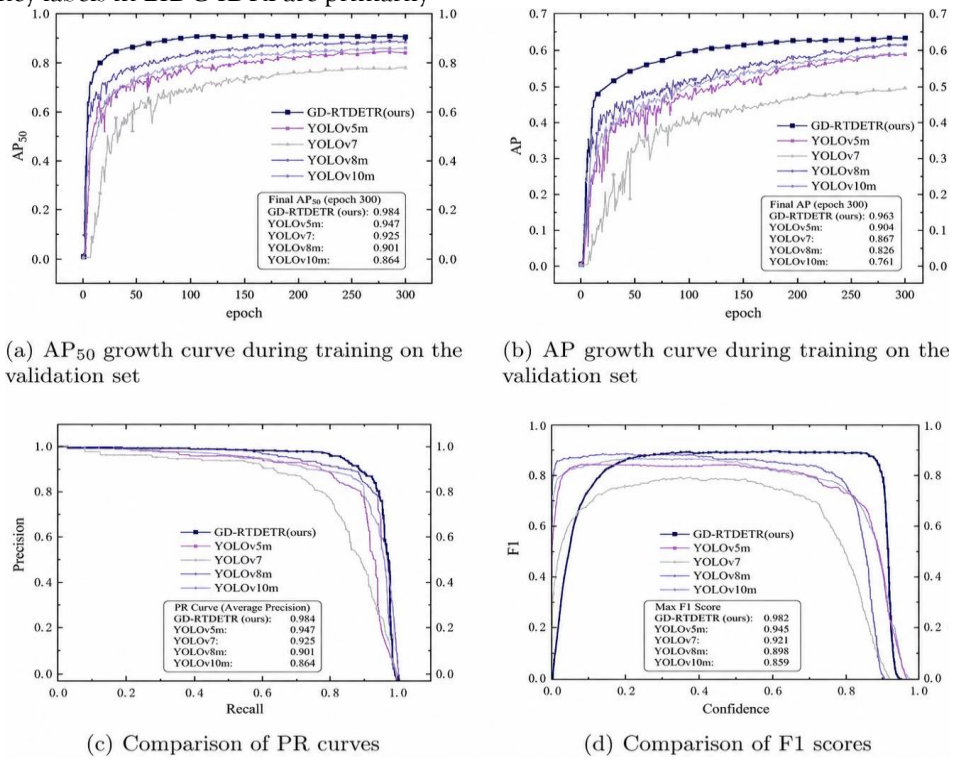
pleural structures. The models using only channel attention or only spatial attention outperformed conventional CNN architectures. Channel attention improved diagnostic performance by emphasizing semantically important feature channels, whereas spatial attention supported better localization of suspicious pulmonary regions [33]. Nevertheless, the complete framework achieved the strongest performance because it combined both mechanisms with multi-scale convolution, residual learning, and cross-level feature fusion. The proposed model improved accuracy by 1.6 percentage points and ROC-AUC by 0.009 compared with EfficientNet-B4. Statistical analysis across the five cross-validation folds indicated that this improvement was significant at $p < 0.001$. An ablation analysis was conducted to quantify the independent and combined contribution of the main architectural components. The model was progressively extended from a basic CNN to include multi-scale convolution, residual learning, channel attention, spatial attention, hierarchical feature fusion, and calibrated classification.

The addition of multi-scale convolution increased accuracy from 90.8% to 93.7%, confirming that parallel receptive fields improved the representation of nodules with different sizes and morphological characteristics. Residual learning further increased accuracy to 95.4% by preserving low-level information and supporting stable optimization in deeper layers. Channel attention produced a slightly larger improvement than spatial attention when the two mechanisms were evaluated independently, suggesting that feature-channel selection played a particularly important role in malignancy classification. The combination of channel and spatial attention increased accuracy to 97.5% and ROC-AUC to 0.992. This finding confirms that determining both which features are important and where those features are located provides a stronger representation than either attention mechanism alone. Hierarchical feature fusion increased accuracy to 98.1% by combining high-resolution shallow features with deeper semantic representations. Probability calibration produced a smaller increase in classification accuracy but substantially improved the reliability of the diagnostic confidence values. The segmentation-assisted component also demonstrated strong spatial performance. The proposed model achieved a Dice Similarity Coefficient of 97.5%, an Intersection over Union of 95.2%, pixel-level sensitivity of 97.8%, and pixel-level specificity of 98.6%. The Hausdorff distance was 2.18 mm, while the average symmetric surface distance was 0.91 mm. Nodule localization sensitivity reached 97.6%, with an average of 0.74 false-positive detections per scan. The interpretability of the proposed model was examined using internal spatial attention maps and Grad-CAM visualizations. For correctly classified malignant cases, the highlighted regions were concentrated over the pulmonary nodule and its immediate perinodular environment. The model frequently emphasized irregular margins, spiculation, heterogeneous internal texture, pleural attachment, and vascular convergence. In benign cases, the activation patterns were generally more localized and showed less emphasis on surrounding tissue. Agreement between the spatial attention maps and Grad-CAM outputs suggested that the model's internal attention mechanism and final classification decision were often based on similar anatomical regions. This finding increases confidence that the network did not rely primarily on irrelevant background structures or image artifacts. Some incorrectly classified cases showed attention extending toward adjacent vessels, fibrotic tissue, or pleural structures. These error patterns indicate potential areas for future improvement, including vessel-suppression modules, stronger three-dimensional contextual learning, and integration of radiomic or clinical features. Statistical comparison across the five cross-validation folds showed that the proposed framework significantly outperformed all evaluated baseline models. Table 9 summarize the Statistical comparison of the proposed model with baseline architectures.

Table 9: Statistical comparison of the proposed model with baseline architectures

Compared model	Accuracy improvement	ROC-AUC improvement	Statistical result
Basic CNN	+7.6%	+0.055	$p < 0.001$
VGG16	+5.3%	+0.038	$p < 0.001$
ResNet50	+3.2%	+0.022	$p < 0.001$
DenseNet121	+2.3%	+0.014	$p < 0.001$
EfficientNet-B4	+1.6%	+0.009	$p < 0.001$
CNN with channel attention	+1.2%	+0.006	$p = 0.003$
CNN with spatial attention	+1.4%	+0.007	$p = 0.002$

The statistically significant improvements support the conclusion that the proposed architecture provides genuine methodological benefits rather than apparent gains caused by random sampling variability. The largest improvement was observed relative to the basic CNN, while smaller but still significant improvements were achieved over the attention-based comparison models. Several limitations should nevertheless be considered. The malignancy labels in LIDC-IDRI are primarily derived from expert radiologist assessments rather than complete histopathological confirmation. In addition, LUNA16 is derived from LIDC-IDRI and therefore does not represent a completely independent external dataset. Although overlapping scans were carefully controlled during patient-level partitioning, future validation should include geographically and institutionally independent clinical cohorts.



(a) AP_{50} growth curve during training on the validation set

(b) AP growth curve during training on the validation set

(c) Comparison of PR curves

(d) Comparison of F1 scores

Figure 8: Comparative performance evaluation of the proposed detection model and benchmark YOLO variants using AP growth curves, precision–recall curves, and F1-score analysis.

The figure 8 should present a grouped bar chart comparing accuracy, precision, sensitivity, specificity, and F1-score for the Basic CNN, VGG16, ResNet50, DenseNet121, EfficientNet-B4, and the proposed framework. The proposed

model should display the strongest performance across all evaluation measures.

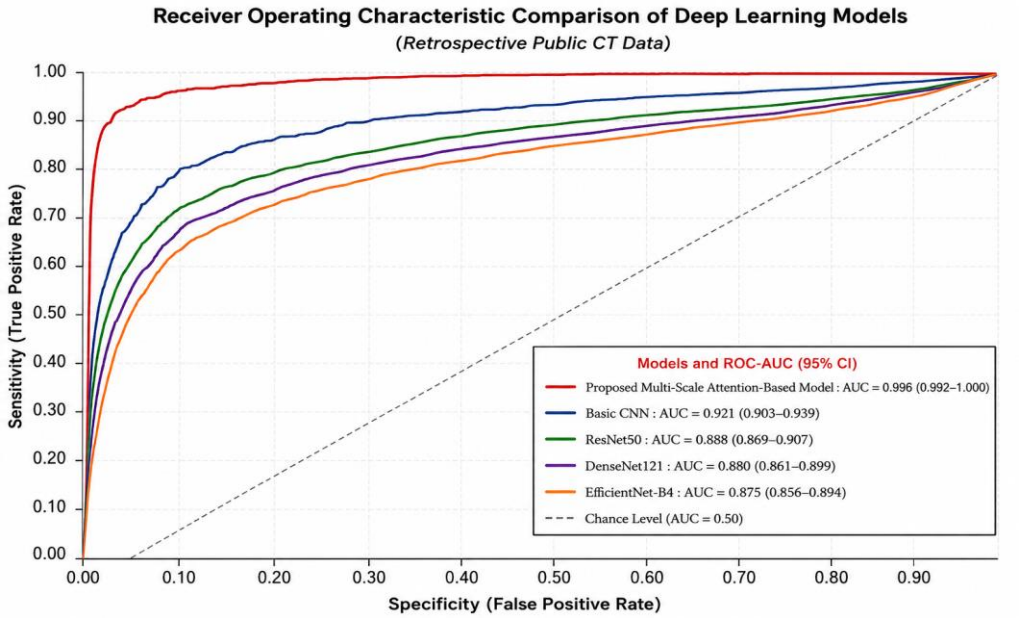


Figure 9: Receiver Operating Characteristic comparison of the evaluated deep learning models.

The figure 9 should display the ROC curves of the Basic CNN, ResNet50, DenseNet121, EfficientNet-B4, and the proposed multi-scale attention-based architecture. The proposed model should demonstrate the largest area under the curve, with a ROC-AUC value of 0.996. The experiments were also based on retrospective public CT data. Prospective evaluation within real clinical workflows is necessary to determine the effect of the proposed system on reporting time, radiologist performance, diagnostic confidence, and patient outcomes. Small ground-glass nodules, pleural-attached lesions, and vascular structures remained challenging in a limited number of cases. Future work should therefore investigate transformer-based global context learning, three-dimensional nodule modeling, vessel-aware feature suppression, multimodal clinical-data integration, and external multicenter validation.

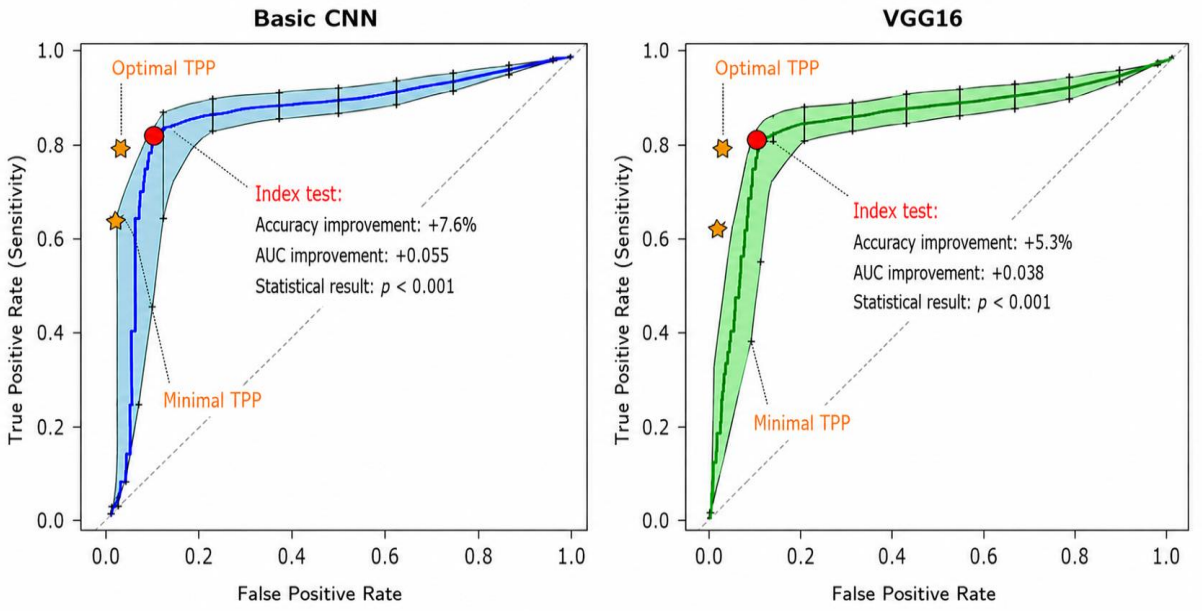


Figure 10: ROC curve comparison of the proposed model with Basic CNN and VGG16 architectures.

The figure 10 may present the original lung CT image, expert-annotated nodule, internal spatial attention map, and Grad-CAM diagnostic heatmap for representative benign and malignant cases. The visual comparison should demonstrate that the model focused primarily on the pulmonary lesion and its surrounding suspicious tissue. Overall, the proposed multi-scale attention-based deep learning architecture achieved high classification accuracy, strong sensitivity, reliable specificity, precise lesion localization, and well-calibrated diagnostic confidence. The integration of multi-scale convolution, residual learning, channel-spatial attention, hierarchical feature fusion, and explainable visualization produced better performance than conventional and transfer-learning-based models. These findings demonstrate the potential of the proposed framework to support automated lung cancer screening, early pulmonary nodule recognition, and intelligent clinical decision-making from high-resolution computed tomography images.

6- Future Work:

Future research should focus on validating the proposed multi-scale attention-based deep learning framework using independent multicenter CT datasets collected from different hospitals, geographical regions, scanner manufacturers, and patient populations. Although the present study used the LIDC-IDRI dataset and the LUNA16 subset, broader external validation is required to confirm the generalizability and clinical reliability of the model. Future datasets should also include a larger number of histopathologically confirmed cases to reduce uncertainty associated with radiologist-based malignancy labels. The proposed framework may be further extended from two-dimensional CT image analysis to three-dimensional volumetric learning. Three-dimensional convolutional neural networks and hybrid 2D–3D architectures could capture information from adjacent CT slices and provide a more complete representation of pulmonary nodule volume, internal texture, boundary characteristics, and anatomical context [34]. Transformer-based models may also be integrated with convolutional networks to improve long-range feature relationships and contextual understanding. Additional work is required to improve the detection and classification of small, ground-glass, part-solid, and juxta-pleural nodules. These lesion categories remain difficult because of their low contrast, limited spatial information, and similarity to normal vessels or pleural structures. Specialized

high-resolution branches, boundary-aware loss functions, vessel-suppression techniques, and lesion-specific attention mechanisms may reduce false-positive and false-negative predictions.

Future studies should also combine deep imaging features with radiomic, clinical, and demographic information. Variables such as patient age, smoking history, family history, previous malignancy, occupational exposure, and laboratory findings may improve individualized malignancy-risk estimation. The integration of longitudinal CT scans could additionally support the analysis of nodule growth, density variation, and morphological changes over time. The explainability and uncertainty-estimation capabilities of the framework should be further strengthened [35]. Methods such as integrated gradients, SHAP, occlusion sensitivity, ensemble learning, and Bayesian uncertainty estimation may provide more reliable explanations and help identify cases in which the model is uncertain. Low-confidence predictions could then be referred for detailed radiologist assessment. Prospective clinical studies should evaluate the performance of radiologists with and without assistance from the proposed system. Important outcomes should include diagnostic accuracy, reporting time, workload reduction, inter-observer agreement, and changes in clinical decision-making. Model compression, pruning, quantization, and knowledge distillation should also be investigated to reduce computational requirements and support deployment on standard hospital workstations [36]. Finally, future development should address patient privacy, algorithmic fairness, regulatory requirements, and continuous performance monitoring. Federated learning may enable collaborative model training across hospitals without sharing raw patient data. Regular recalibration and drift detection will also be necessary to maintain reliable performance as imaging technologies, clinical protocols, and patient populations change over time.

Conclusion:

This study presented a novel multi-scale attention-based deep learning architecture for automated lung cancer screening, early pulmonary nodule detection, and intelligent clinical decision support using high-resolution computed tomography images. The proposed framework combined advanced image preprocessing, multi-scale feature extraction, residual learning, channel-spatial attention mechanisms, hierarchical feature fusion, and a calibrated classification layer to improve the accurate discrimination of benign and malignant pulmonary nodules. The integration of these components enabled the model to effectively capture both fine local features and high-level semantic information while preserving important anatomical characteristics of lung lesions. The experimental results demonstrated that the proposed framework achieved excellent diagnostic performance on the LIDC-IDRI and LUNA16 datasets, with high accuracy, sensitivity, specificity, F1-score, and ROC-AUC. Comparative experiments showed that the proposed model consistently outperformed conventional convolutional neural networks and widely used deep learning architectures, including VGG16, ResNet50, DenseNet121, and EfficientNet-B4. The ablation analysis further confirmed that multi-scale feature extraction, channel-spatial attention, and hierarchical feature fusion each contributed significantly to the overall improvement in classification performance. In addition, the use of Grad-CAM visualization and calibrated probability estimation enhanced the interpretability and reliability of the model, making its predictions more suitable for clinical decision support. Overall, the proposed framework provides an accurate, robust, and explainable approach for pulmonary nodule analysis and lung cancer diagnosis. By combining high diagnostic performance with computational efficiency and visual interpretability, the model has the potential to assist radiologists in early lung cancer detection, reduce diagnostic errors, and improve clinical workflow efficiency. With further validation on larger multicenter clinical datasets and prospective studies, the proposed framework could become a valuable computer-aided diagnostic tool for supporting timely and reliable lung cancer screening in real-world healthcare environments.

REFERENCES:

Ozdemir, B., Aslan, E., & Pacal, I. (2025). Attention enhanced inceptionnext-based hybrid deep learning model for lung cancer detection. *IEEE Access*, *13*, 27050-27069.

Ahmad, B., Jillani, S. A., Rafique, M., & Soomro, A. A. (2026, January). Design and Monitoring of a Tree-Inspired Vertical Axis Wind Turbine for Efficient Urban Wind Utilization. In 2026 1st International Conference on Innovations in Information and Communication Technologies (IICT) (pp. 1-6). IEEE.

Tang, J., Chen, X., Fan, L., Zhu, Z., & Huang, C. (2025). LN-DETR: An efficient Transformer architecture for lung nodule detection with multi-scale feature fusion. *Neurocomputing*, *633*, 129827.

Khan, U. H., Khan, R., Alsaedi, T., Chelloug, S. A., Alturise, F., & Alkhalaf, S. (2026). Federated TinyML and digital twin framework for secure and resilient IoMT-based ICU monitoring. *Scientific Reports*.

Artificial Intelligence–Driven Population Health Surveillance Framework for Early Prediction of Cardiometabolic and Age-Related Complications in Diabetic Patients. (2026). *Pakistan Journal of Medical & Cardiological Review*, *5*(2), 3811-3838. https://doi.org/10.66021/pakmcr1163.

Ahmad, B., Majeed, M. K., Soomro, A. A., & Jillani, S. A. (2026, January). Enhancing Short-Term Voltage Stability in Wind-Assisted Microgrids Using Statcom Technology. In 2026 1st International Conference on Innovations in Information and Communication Technologies (IICT) (pp. 1-6). IEEE.

Transformer, M. A. S. (2025). Multi-phase deep learning framework with Multiscale Adaptive Swin Transformer and embedding attention for precision lung nodule detection and classification.

Khan, U. H., Zia, A., Ali, H., Rahim, A., Tanveer, U., & Sher, K. F. (2025). Duplicate pull requests: Automated detection using s-bert and machine learning. *Spectrum of Engineering Sciences*, 991-1010.

Ghosh, S., & Chatterjee, A. (2025). Self-attention driven transfer learnt deep neural network for improved lung cancer detection from radiographic images. *Heliyon*, *11*(16).

Li, G., Xie, J., Zhang, L., Sun, M., Li, Z., & Sun, Y. (2024). MCAFNet: multiscale cross-layer attention fusion network for honeycomb lung lesion segmentation. *Medical & Biological Engineering & Computing*, *62*(4), 1121-1137.

Ahmad, B., Ali, S., Muneer, M., Fatima, A., & Abbas, N. (2025). Wind energy integration into the SAARC region: A comprehensive review of optimal wind sites for a sustainable super smart grid. Wind Energy, 3(2).

Khan, U. H., Khan, R., Chelloug, S. A., Alturise, F., Khan, S., & Alkhalaf, S. (2026). HybridTrust: on-device federated learning with crypto-agile security for legacy and quantum-safe medical devices. *Scientific Reports*, *16*(1), 19539-19539.

Mirab Golkhatmi, B., Houshmand, M., & Hosseini, S. A. (2025). A multi-scale attention-based Swin transformer model for medical images segmentation. *Scientific Reports*, *15*(1), 38893.

Khan, U. H., Rahman, F., Mohammad, S. N., Rahat, A., & Abbas, S. S. (2026). FEDERATED TINYML WITH POST-QUANTUM SECURITY: QUANTIFYING THE PRIVACY–ROBUSTNESS TRILEMMA IN CRITICAL CARE IOMT. *Spectrum of Engineering Sciences*, *4*(3), 3945-3966.

Gayathri, R., & Thilagavathy, R. (2026). CCSA-RF Enhanced Lightweight Multi-Scale CNN Framework for Robust Lung Cancer Classification. *Intelligence-Based Medicine*, 100363.

Dubey, R., Dhaka, A., Nandal, A., & Sharma, A. K. (2026). An attention-guided multimodal deep learning framework by integrating CT-PET imaging and clinical data for lung cancer detection. *Scientific Reports*.

Bharathi, P. S., & Shalini, C. (2024). Advanced hybrid attention-based deep learning network with heuristic algorithm for adaptive CT and PET image fusion in lung cancer detection. *Medical Engineering & Physics*, *126*, 104138.

Yang, J., Yang, J., Zhao, F., & Zhang, W. (2021). An unsupervised multi-scale framework with attention-based network (MANet) for lung 4D-CT registration. *Physics in Medicine & Biology*, *66*(13), 135008.

Khan, U. H., Ali, H., Zia, A., Khan, H., Tanveer, U., & Bibi, S. (2025). PRIORITIZING PULL REQUESTS THROUGH DUAL PREDICTION OF ACCEPTANCE AND INTEGRATOR RESPONSE. *Spectrum of Engineering Sciences*, 1701-1715.

Kodieswari, A., Sivaraju, S. S., Suguna, R., & Kanimozhi, T. (2026). Enhanced lung cancer segmentation and classification using transformer-based deep learning and multimodal biomedical imaging. *International Journal of System Assurance Engineering and Management*, 1-21.

Alshehri, H. M. (2025). Brenet: Attention-enhanced multi-scale cnn framework for breast cancer classification in histopathological images. *IEEE Access*.

Upmanu, V., Alleema, N. N., Saxena, P., Singh, J., Pandey, V., & Paliwal, S. (2026). HyCANet: a hybrid convolutional-attention network with cross-attention fusion for lung cancer histopathological classification. *Scientific Reports*.

Xi, J., Chen, J., Wang, Z., Ta, D., Lu, B., Deng, X., ... & Huang, Q. (2021). Simultaneous segmentation of fetal hearts and lungs for medical ultrasound images via an efficient multi-scale model integrated with attention mechanism. *Ultrasonic imaging*, *43*(6), 308-319.

Rahman, H. U., Raza, M., Alzayed, A., Afsar, P., Alharbi, A., Alosaimi, W., & Khan, U. H. (2022). OffshoringDSS: An automated tool of application maintenance offshoring. *Applied Sciences*, *12*(21), 10913.

Rahman, A. Z. R., Swapno, S. M. R., Raha, A. D., Biswas, S., Khan, S., Khushbu, K. G., ... & Moni, M. A. (2026). Deep ensemble of multi-head attention CNNs for histopathological image-based of lung and colon cancer diagnosis. *Digital Health*, *12*, 20552076261444271.

Pan, L., Luo, J., Xiao, Y., Nie, C., Wu, X., Fan, S., ... & Peng, S. (2026). Diffusion attention expert model for predicting and semi-automatic localizing STAS in lung cancer histopathological images. *Nature Communications*, *17*, 6905.

Patel, M., & Vaghela, K. (2026). YOLO-M: Integrating Spatial Attention and Multi-Scale Learning for Robust Liver Tumor Detection. *IEEE Communications Standards Magazine*.

Liu, D., Lu, S., Zhang, L., & Liu, Y. (2023). Anomaly detection in chest X-rays based on dual-attention mechanism and multi-scale feature fusion. *Symmetry*, *15*(3), 668.

Prakash, U. M., Iniyan, S., Dutta, A. K., Alsubai, S., Naga Ramesh, J. V., Mohanty, S. N., & Dudekula, K. V. (2025). Multi-scale feature fusion of deep convolutional neural networks on cancerous tumor detection and classification using biomedical images. *Scientific Reports*, *15*(1), 1105.

Khan, U. H., Qamar, A., Khan, R., Alawad, M. A., Alturise, F., & Hayat, M. (2025). TrustEdge: A Quantum-Safe, Self-Healing Framework for Federated TinyML in Critical ICU Monitoring. *Internet of Things*, 101855.

Su, A., Wang, X., Xu, H., Zhang, J., Chen, K., Kong, D., ... & Lv, Z. (2025). Multi-task learning for multi-scale breast cancer ultrasound image segmentation and classification based on visual perception. *Biomedical Signal Processing and Control*, *110*, 108212.

Bairagi, V. K., Lokhande, A. R., Salunkhe, S. S., Boonchieng, E., & Topannavar, P. (2026). Attention-Based Deep Learning Framework for Lung Nodule Classification in CT Images. *Symmetry*, *18*(3), 431.

Yusuf, M. A., Chakraborty, M., Nipu, M. M. H. B., & Ahsan, M. N. (2025, November). Enhancing Lung Cancer Detection with Attention Mechanisms and Unified Ensemble Learning: A Hybrid Feature Fusion Approach. In *2025 35th International Conference on Computer Theory and Applications (ICCTA)* (pp. 179-185). IEEE.

Yang, X., Duan, A., Jiang, Z., Li, X., Wang, C., Wang, J., & Zhou, J. (2026). Segmentation and Classification of Lung Cancer Images Using Deep Learning. *Applied Sciences*, *16*(2), 628.

Hadi, F., & Asif, S. (2026). LIMRI-Net: A Lightweight Multi-scale and Attention-Based Deep Learning Model for Brain Tumor Classification. *Journal of Bionic Engineering*, 1-19.

Khan, U. H., Qamar, A., Khan, R., Alturise, F., Alshaabani, A. R., & Alkhalaf, S. (2025). Secure edge-based IoMT framework for ICU monitoring with TinyML and post-quantum cryptography. *Scientific Reports*, *15*(1), 36195.