Volume 3, Issue 7, 2025

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

NEXT-GENERATION AI-DRIVEN FRAMEWORKS FOR EARLY LUNG CANCER DETECTION USING ADVANCED IMAGING PROCESSING AND PREDICTIVE MODELING TECHNIQUES

Muhammad Suleman[*1], Fakhar Anjam[2], Musarat Hussain[3], Dr. Muhammad Zulfiqah Sadikan[4], Dr Zain Muzaffar[5], Muhammad Huzaifa[6], Muhammad Waleed Iqbal[7], Imran Ali[8], Dr. Ajab Khan[9]

[*1,2]*Department of Electrical Engineering, University of Science and Technology, Bannu, Pakistan.*
[3]*Department of Biological Science and Technology, China Medical University, Taichung, Taiwan.*
[4]*Faculty of Pharmacy and Health Sciences, University Kuala Lumpur Royal College of Medicine Perak, Jalan Greentown, 30450 Ipoh, Perak, Malaysia.*
[5]*Department of General Medicine, Aziz Bhatti Shaheed Teaching Hospital, Gujrat, Pakistan.*
[6]*Institute of Biotechnology and Genetic Engineering, The University of Agriculture, Peshawar, 25130, Khyber Pakhtunkhwa, Pakistan.*
[7]*Department of Computer Science, COMSATS University Islamabad, Sahiwal Campus, Pakistan.*
[8]*Department of Pharmacy, Faculty of Biological Sciences, Quaid-i-Azam University, Islamabad, Pakistan.*
[9]*Director ORIC, Abbottabad University of Science and Technology, Abbottabad, Pakistan.*

[*1]suleman_ktk@ustb.edu.pk, [*1]sulemanktk123456@gmail.com, [2]F.Anjam@ustb.edu.pk, [3]musarathussain2118@gmail.com, [4]mzsadikan@yahoo.com, [5]zain.muz@icloud.com, [6]khuzaifakhattak@gmail.com, [7]mmuhammadwaleed256@gmail.com, [8]imranali@bs.qau.edu.pk, [9]directororic@aust.edu.pk

DOI: https://doi.org/10.5281/zenodo.17157527

Keywords
Lung cancer detection; low-dose computed tomography (LDCT); advanced imaging processing; image harmonization; super-resolution; domain adaptation; pulmonary nodule detection; 3D convolutional neural networks (3D-CNNs).

Article History
Received: 27 June 2025
Accepted: 05 September 2025
Published: 19 September 2025

Copyright @Author
Corresponding Author: * Muhammad Suleman

Abstract

Early detection of lung cancer remains a critical determinant of patient survival, yet current clinical screening practices continue to face major challenges arising from heterogeneous image acquisition protocols, subtle lesion morphology, and significant inter-reader variability. To address these limitations, this study introduces next-generation AI-driven frameworks that integrate advanced imaging processing pipelines with predictive modeling techniques for robust and clinically actionable early lung cancer detection. The framework begins with physics-informed harmonization strategies combining denoising, super-resolution, and cross-protocol domain adaptation to generate standardized inputs that preserve diagnostically salient textures across diverse scanner vendors and acquisition settings. Building on these pre-processed datasets, a multi-scale nodule discovery module integrates 3D convolutional encoders with transformer-based detectors and topology-aware segmentation, enabling accurate characterization of challenging sub-centimeter, subsolid, and juxta-pleural nodules. Beyond detection, the framework advances malignancy risk prediction through the fusion of radiomics and deep learning embeddings, which synergistically combine handcrafted intensity, shape, and texture descriptors with self-supervised representations to enhance data efficiency and generalizability across sites. Multimodal graph transformers are then employed to incorporate imaging findings

Volume 3, Issue 7, 2025

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

with clinical risk factors such as age, smoking history, and comorbidities, as well as longitudinal biomarkers including attenuation changes and nodule growth kinetics. This integration produces time-aware malignancy risk scores that are rigorously calibrated through conformal prediction, thereby providing trustworthy uncertainty estimates essential for clinical decision support. Training is guided by curriculum learning strategies and distributionally robust optimization objectives that account for institutional imbalances, protocol drift, and data scarcity. Evaluation across multi-institutional datasets and public benchmarks demonstrates consistent improvements in sensitivity for nodules ≤6 mm, reduction of false positives per study, superior calibration performance, and tangible reductions in time-to-diagnosis when compared with existing guideline-based criteria. Furthermore, decision-curve analysis highlights the clinical utility of the proposed framework in reducing unnecessary follow-ups while maintaining high early-stage detection rates. To support clinical adoption, the study incorporates governance overlays addressing algorithmic fairness, privacy-preserving federated learning, and interpretable evidence generation through saliency mapping and radiomics attribution. By embedding these governance elements within a reader-in-the-loop design, the framework not only enhances trustworthiness but also promotes sustainable deployment in real-world screening ecosystems. Overall, this next-generation AI-driven approach demonstrates the potential to significantly improve early lung cancer detection, foster equitable care, and accelerate translation of AI innovations into clinical practice.

1- Introduction:

Lung cancer continues to represent one of the most pressing challenges in global oncology, with recent statistics from the World Health Organization estimating over 1.8 million deaths annually. It remains the leading cause of cancer mortality, accounting for more lives lost each year than breast, colon, and prostate cancers combined. Despite advancements in targeted therapies, immunotherapy, and surgical techniques, patient survival is overwhelmingly determined by the stage of diagnosis. The contrast is stark: when discovered at Stage I through screening, five-year survival rates can exceed 65%, while late-stage diagnoses often result in survival below 10%. This sharp gradient underscores the necessity of effective early detection strategies that can be scaled to population-level screening programs [1]. Low-dose computed tomography (LDCT) has emerged as the gold standard for screening high-risk populations, a status reinforced by landmark clinical trials such as the National Lung Screening Trial (NLST) in the United States and the NELSON trial in Europe. Both demonstrated significant reductions in lung cancer mortality when LDCT was systematically applied to individuals with elevated risk factors, such as long-term smoking histories. Nevertheless, LDCT is not without limitations. Heterogeneity in imaging protocols across vendors, scanner generations, and reconstruction algorithms produces substantial variability that hampers both manual interpretation and automated systems. The morphology of many clinically significant lesions, including sub-centimeter nodules, ground-glass opacities, and juxta-pleural nodules, is subtle and easily overlooked. Furthermore, inter-reader variability among radiologists remains high, with disagreements in as many as one-third of ambiguous cases. Compounding these issues are high false-positive rates, with many screening programs reporting that more than one in four patients flagged require unnecessary follow-ups or invasive procedures [2]. These limitations highlight the urgent need for robust computational frameworks that can augment human expertise in LDCT interpretation.

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

Artificial intelligence (AI) has demonstrated transformative capabilities in medical imaging, benefiting from advances in computer vision, machine learning, and predictive modeling. Early computer-aided detection systems built on handcrafted radiomics features capturing texture, intensity, and shape descriptors showed promise but were ultimately constrained by limited generalizability. The deep learning revolution, enabled by convolutional neural networks (CNNs), residual architectures, and three-dimensional U-Nets, brought significant improvements, enabling automated localization and characterization of pulmonary nodules. More recently, transformer-based models have introduced long-range contextual awareness, further improving performance in visual recognition tasks. Yet despite these advances, several critical challenges remain unresolved. Most AI studies are trained on curated public datasets, such as LIDC-IDRI or LUNA16, which do not capture the diversity of real-world clinical imaging. Consequently, models often fail to generalize when tested on external cohorts [3]. Moreover, existing approaches tend to treat pre-processing, detection, and risk prediction as disjoint modules, which introduces inefficiencies and propagates errors across the pipeline. Many current systems also rely exclusively on imaging features, neglecting clinical data such as age, smoking history, comorbidities, and longitudinal biomarkers including nodule growth kinetics or attenuation changes. Finally, concerns about uncertainty calibration, fairness, interpretability, and privacy remain insufficiently addressed, raising barriers to adoption in clinical practice. Table 1 summarizes these limitations of existing systems, highlighting how they impact clinical decision-making and identifying the gaps that motivate the present study. For example, the lack of harmonization across imaging protocols significantly limits generalization, while the poor sensitivity for sub-centimeter nodules prevents reliable detection of early disease. Similarly, the absence of calibrated uncertainty quantification undermines clinician trust, and the lack of embedded fairness or privacy-preserving mechanisms hinders regulatory compliance. These gaps provide the rationale for developing a new framework that moves beyond isolated modules toward a unified, clinically aligned system.

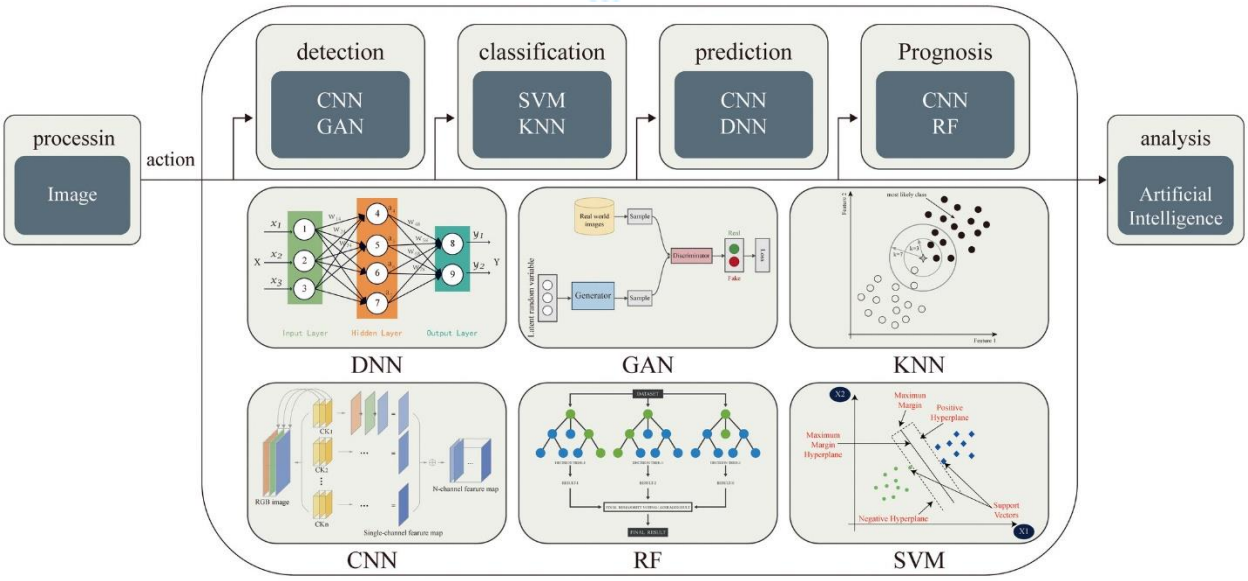
Table 1: Research Gaps in Existing AI Frameworks for Lung Cancer Detection

Challenge	Limitation in Current Systems	Clinical/Technical Impact	Gap Addressed in This Work
Imaging heterogeneity	Lack of harmonization across scanners and protocols	Poor generalization, false alarms	Physics-informed harmonization strategies
Subtle/complex lesion morphology	Limited sensitivity for sub-cm, subsolid, and juxta-pleural nodules	Missed early-stage cancers	Multi-scale detection with CNN+Transformer
Risk prediction	Imaging-only models, no integration of clinical/temporal factors	Incomplete patient risk stratification	Multimodal graph transformers
Calibration and trustworthiness	Overconfident predictions, poor uncertainty quantification	Low clinician adoption	Conformal prediction, calibrated scores
Deployment readiness	Absence of fairness/privacy frameworks	Ethical/regulatory concerns	Federated learning, governance overlays

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

To overcome these limitations, the present study proposes a next-generation AI-driven framework for early lung cancer detection that integrates advanced imaging harmonization, multi-scale detection, multimodal feature fusion, and calibrated risk prediction into a unified pipeline. The system begins with physics-informed harmonization strategies that include denoising, super-resolution, and cross-protocol domain adaptation, producing standardized inputs that preserve diagnostically salient textures across diverse scanners and acquisition settings. Building upon this foundation, a multi-scale nodule discovery module combines three-dimensional CNN encoders with transformer-based detectors and topology-aware segmentation, enabling precise characterization of small, subsolid, and juxta-pleural nodules that typically challenge existing systems. Risk prediction is further enhanced by fusing handcrafted radiomics descriptors with deep embeddings derived from self-supervised learning, which improves both data efficiency and cross-institutional generalizability [4]. This fusion is extended through multimodal graph transformers that incorporate clinical factors such as age, smoking history, and comorbidities, together with longitudinal biomarkers such as nodule growth rates and attenuation changes, to yield time-aware malignancy risk scores. Importantly, these predictions are calibrated through conformal methods, providing trustworthy uncertainty estimates for clinical decision support. The training pipeline is guided by curriculum learning strategies and distributionally robust optimization, ensuring resilience against data scarcity, protocol drift, and institutional imbalances. To further support deployment readiness, the system embeds governance overlays, including fairness assessments, federated learning for privacy preservation, and interpretable outputs through saliency mapping and radiomics attribution. A schematic overview of this proposed framework is presented in Figure 1. The diagram illustrates the sequential and interdependent components, beginning with raw LDCT inputs, followed by harmonization, nodule detection, multimodal feature integration, risk prediction, and finally, calibrated outputs delivered with governance safeguards. This figure highlights how the framework unifies technical innovations into a cohesive system designed to align with clinical realities.



Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

Figure 1: Conceptual Framework for AI-Driven Early Lung Cancer Detection

The contributions of this work are multidimensional. It offers a unified pipeline that connects pre-processing, detection, and risk prediction rather than treating them as separate modules. It incorporates multimodal learning that goes beyond image-only systems by integrating clinical and longitudinal data. It introduces robust training paradigms and conformal prediction to ensure well-calibrated and trustworthy outputs. Finally, it embeds governance elements such as fairness, federated learning, and interpretability, which collectively promote ethical, equitable, and sustainable clinical adoption. These contributions are summarized in Table 2.

Table 2: Major Contributions of the Proposed Framework

Contribution Area	Specific Innovations	Anticipated Impact
Data Harmonization	Physics-informed denoising, super-resolution, domain adaptation	Robust generalization across institutions
Nodule Detection	CNN + Transformer hybrid with topology-aware segmentation	Higher sensitivity for small/subsolid nodules
Risk Prediction	Radiomics–deep feature fusion with graph transformers	More holistic malignancy risk modeling
Calibration	Conformal prediction for uncertainty quantification	Reliable decision support in clinics
Governance & Ethics	Fairness metrics, federated learning, interpretability overlays	Trustworthy and deployable AI systems

Overall, the objective of this research is to deliver a next-generation AI framework that not only achieves state-of-the-art technical performance but also addresses the broader requirements for clinical translation. By rigorously integrating harmonization, detection, multimodal risk modeling, calibration, and governance within a single system, this work aspires to advance early lung cancer detection toward a stage where AI can be reliably deployed in real-world screening programs. In doing so, it responds to the dual imperatives of scientific innovation and clinical trustworthiness, demonstrating that AI can be both technically advanced and ethically grounded.

2- AI-Based Approaches to Early Lung Cancer Detection:

The integration of artificial intelligence into medical imaging has initiated a profound shift in how clinicians, data scientists, and researchers conceptualize and address the challenge of early lung cancer detection. Traditionally, radiologists have shouldered the responsibility of carefully reviewing low-dose computed tomography (LDCT) scans in search of subtle pulmonary nodules. While expert interpretation remains central to clinical care, the rapid expansion of imaging volumes in population-level screening programs has highlighted the limitations of purely manual review, including variability in expertise, inter-reader disagreements, and susceptibility to fatigue. Artificial intelligence, with its ability to process large volumes of imaging data rapidly and consistently, has emerged as a critical partner to human expertise, not merely as a tool to automate detection but as a paradigm shift in how diagnostic insights are generated, validated, and translated into clinical decision-making [5]. Over the past two decades, the trajectory of AI applications in lung cancer detection has unfolded in several distinctive phases, each shaped by advances in computational capability and methodological

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

innovation. The earliest phase was marked by computer-aided detection (CAD) systems that relied heavily on handcrafted radiomics features, such as nodule shape, intensity gradients, and texture descriptors. While pioneering in their intent, these approaches were constrained by their dependence on manual feature design and their limited ability to generalize across scanners, protocols, and patient populations. The subsequent decade witnessed the widespread adoption of deep learning, particularly convolutional neural networks (CNNs), which revolutionized image analysis by enabling models to autonomously learn hierarchical feature representations directly from imaging data. CNN-based systems offered dramatic improvements in sensitivity and specificity, and architectures such as 3D CNNs and U-Nets became foundational tools for volumetric lung nodule detection and segmentation.

More recently, the field has embraced transformer-based architectures, originally developed for natural language processing but now adapted for medical imaging tasks. By leveraging self-attention mechanisms, transformers can capture global contextual relationships across CT volumes, complementing the local detail sensitivity of CNNs [6]. Hybrid CNN-transformer frameworks have further enhanced multi-scale detection capabilities, achieving state-of-the-art performance in identifying challenging nodules such as subsolid and juxta-pleural lesions. Alongside these algorithmic innovations, the scope of AI-driven research has broadened from detection alone to comprehensive risk stratification. Multimodal frameworks have begun to integrate imaging with clinical data, demographic information, and temporal biomarkers, enabling more holistic assessments of malignancy risk and patient prognosis. At the same time, the conversation around AI in lung cancer detection has shifted beyond raw accuracy to encompass broader considerations of reliability, trust, and ethical deployment. Researchers are increasingly addressing issues of probability calibration and uncertainty quantification, recognizing that overconfident models pose significant risks in clinical practice. Similarly, interpretability has become an indispensable focus, with techniques such as saliency mapping, attention visualization, and radiomics attribution being developed to provide clinicians with transparent evidence supporting AI-generated predictions [7]. Governance frameworks are also taking shape, with emphasis on algorithmic fairness, demographic equity, and privacy-preserving methodologies such as federated learning, which allows collaborative training across institutions without exposing sensitive patient data. Taken together, these developments illustrate both the remarkable progress and the persistent gaps in AI-driven lung cancer detection. The following subsections trace this evolution in greater detail, moving chronologically from early CAD systems through deep learning and transformer-based methods to the most recent multimodal and governance-oriented frameworks. By highlighting the contributions and limitations of each stage, this review establishes the foundation for the next-generation framework proposed in this study, which seeks to unify accuracy, robustness, interpretability, and ethical readiness in a single end-to-end system for early lung cancer detection.

2.1- Early Computer-Aided Detection (CAD) Systems:

The earliest computer-aided detection (CAD) systems for lung cancer emerged at the turn of the millennium, when the advent of spiral and later multi-slice computed tomography (CT) allowed clinicians to obtain detailed volumetric scans of the thorax. As radiologists were confronted with hundreds of slices per patient in screening trials, the sheer volume of data made manual interpretation both laborious and error-prone, prompting the development of computational support tools. These CAD systems were designed to assist radiologists by highlighting suspicious regions, thereby functioning as a “second reader” in clinical workflows. Methodologically, early CAD

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

systems were built upon handcrafted radiomics features, which were manually engineered descriptors of imaging data. Intensity-based features quantified variations in Hounsfield units (HU), while shape descriptors captured geometric attributes such as sphericity, compactness, or eccentricity. Texture features derived from Haralick gray-level co-occurrence matrices, Gabor filters, or wavelet decompositions provided additional cues about intra-nodular heterogeneity. These handcrafted descriptors were then fed into conventional machine learning classifiers, including linear discriminant analysis (LDA), support vector machines (SVMs), decision trees, and random forests, which attempted to distinguish nodules from non-nodular structures such as vessels, airway walls, or imaging artifacts. Although CAD systems represented a milestone in the automation of lung cancer detection, their limitations quickly became apparent [8]. Their performance was highly dependent on the quality and representativeness of handcrafted features, as well as on the distribution of the training data. Variations in imaging protocols, reconstruction kernels, and scanner vendors often led to dramatic performance drops when CAD systems were applied outside their development environment. Subtle lesion morphologies particularly subsolid nodules, ground-glass opacities, and juxta-pleural lesions proved particularly challenging, as handcrafted descriptors often failed to capture their nuanced visual signatures. Consequently, CAD systems frequently generated large numbers of false positives, sometimes exceeding dozens of candidates per scan, which paradoxically increased the burden on radiologists instead of reducing it. Despite these shortcomings, early CAD systems established several important concepts that remain influential [9]. They demonstrated the feasibility of algorithmic assistance in thoracic CT interpretation and introduced the principle of augmenting rather than replacing radiologists. More importantly, they created a conceptual framework consisting of candidate detection, feature extraction, and classification that shaped the development of subsequent deep learning pipelines. Table 3 shows the Characteristics of Early CAD Systems for Lung Cancer Detection.

Table 3: Characteristics of Early CAD Systems for Lung Cancer Detection

Feature Type	Example Techniques	Classifiers Commonly Used	Key Strengths	Limitations
Intensity-based	Histogram analysis, mean HU values	Linear Discriminant Analysis (LDA)	Simple, computationally efficient	Sensitive to noise and acquisition differences
Shape-based	Sphericity, compactness, eccentricity	Support Vector Machines (SVMs)	Captures geometric structure of nodules	Struggles with irregular/juxta-pleural nodules
Texture-based	Haralick features, Gabor filters, wavelets	Decision Trees, Random Forests	Detects intra-nodule heterogeneity	Limited discriminative power for subsolid nodules
Rule-based	Threshold filters, blobness measures	Hybrid rule + classifier pipelines	Easy to interpret, low complexity	High false-positive rates, brittle performance

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

The table 3 illustrates how early CAD systems drew on multiple feature families to approximate the complexity of lung nodule morphology. While each feature type offered unique strengths intensity features being efficient, shape features providing structural context, and texture features capturing intra-lesion variation their combination still proved insufficient for robust real-world deployment. The over-reliance on rigid feature engineering meant that systems lacked adaptability to unseen cases. As a result, these early pipelines functioned more as experimental proofs-of-concept than as clinically trusted tools, yet they laid the foundation for the data-driven feature learning paradigm that followed. Figure 2 shows the general pipeline of early CAD system.

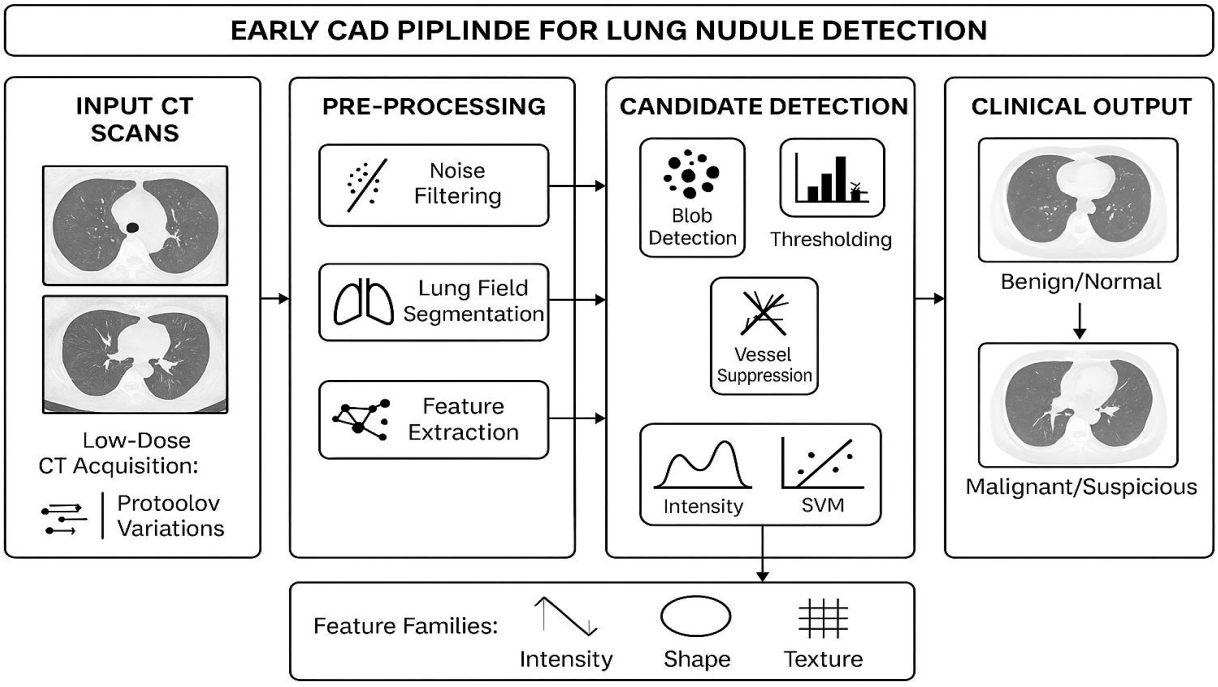


Figure 2: General Pipeline of Early CAD Systems

Figure 2 conceptually illustrates the architecture of early CAD pipelines. Raw CT volumes typically underwent pre-processing, which involved lung field segmentation and basic denoising. Candidate detection algorithms, often based on blobness filters or threshold-based region growing, generated regions of interest. From these candidates, handcrafted features describing intensity, shape, and texture were extracted. Finally, classical classifiers such as SVMs or decision trees determined whether each candidate represented a true nodule or a false alarm. The figure highlights the modular, step-wise nature of these systems and their dependence on handcrafted features, in contrast to modern deep learning methods that unify feature extraction and classification in an end-to-end framework. Early CAD systems were limited in clinical adoption due to their fragility, high false-positive burden, and poor generalizability, but they nonetheless played a formative role in the history of AI in lung cancer imaging. By introducing automated pipelines, these systems demonstrated both the promise and the pitfalls of computer-assisted detection, setting the stage for the rise of deep learning approaches that could overcome the rigidity of handcrafted radiomics and move toward more adaptive, robust, and clinically meaningful solutions.

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

2.2- Deep Learning-Based Detection and Segmentation:

The introduction of deep learning, particularly convolutional neural networks (CNNs), marked a fundamental paradigm shift in computer-aided detection of pulmonary nodules. Unlike early CAD systems, which were hampered by their reliance on manually engineered features and shallow classifiers, deep learning architectures provided an end-to-end mechanism for learning directly from raw data. This capacity to extract multilevel abstractions from imaging inputs transformed lung cancer detection by eliminating the bottleneck of handcrafted radiomics, reducing dependence on human expertise, and improving scalability across increasingly large datasets. The earliest applications of deep learning in thoracic imaging appeared around 2014–2015, when researchers adapted CNNs, which had already revolutionized natural image recognition tasks, to lung CT scans [10]. A key innovation was the extension from two-dimensional convolutional filters to three-dimensional ones, thereby capturing volumetric information across CT slices. This was particularly important for pulmonary nodule detection, where the distinction between nodules and vascular structures often depends on subtle continuity across slices. Models such as 3D CNNs trained on cropped regions-of-interest demonstrated significant gains in sensitivity, especially for nodules in the 3–8 mm range, which are the most clinically consequential for early detection. The U-Net architecture and its derivatives rapidly became central to nodule segmentation. Introduced originally for biomedical image segmentation, U-Net's encoder–decoder structure with skip connections allowed networks to simultaneously capture coarse contextual information and fine-grained spatial details. Variants of U-Net, such as 3D U-Net and V-Net, were widely adopted in lung cancer imaging, enabling accurate delineation of nodule boundaries in volumetric space [11]. This development represented a crucial step toward quantifying nodule morphology, volume, and growth dynamics, all of which are vital for malignancy risk assessment. Residual networks (ResNets) and densely connected networks (DenseNets) further advanced the field by addressing problems of vanishing gradients and inefficient feature reuse. These architectures made it possible to train deeper models on relatively small medical imaging datasets, a constraint that has historically limited the performance of neural networks in healthcare applications. Researchers also introduced patch-based training strategies, in which small sub-volumes of CT data were used to mitigate class imbalance between nodules (rare) and background voxels (abundant). This strategy not only improved computational efficiency but also reduced false negatives in underrepresented nodule categories.

The availability of public datasets and challenges played a critical role in accelerating the adoption of deep learning. The LIDC-IDRI dataset, with its multi-radiologist annotated nodules, and the LUNA16 challenge, which provided standardized evaluation metrics and leaderboards, became benchmarks for comparing model performance. Deep CNNs trained on these datasets consistently outperformed handcrafted CAD approaches, achieving sensitivities above 90% at fewer than two false positives per scan metrics that were previously unattainable. Other datasets, such as NLST subsets and more recent repositories like Tianchi and Kaggle lung nodule challenges, further expanded the scope for benchmarking [12]. Despite these advances, CNN-based detection systems faced two persistent challenges. First, CNNs are inherently local in nature, as their receptive fields are restricted to progressively expanding neighborhoods. While effective for capturing local features, this architecture struggles to model long-range dependencies across an entire CT volume. Nodules located near anatomical boundaries, such as vessel bifurcations or pleural surfaces, were often misclassified due to the lack of global context. Second, deep CNNs exhibited sensitivity to domain shifts, whereby

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

performance declined when applied to data from scanners, protocols, or patient populations not represented in the training set. This lack of robustness highlighted the ongoing need for harmonization strategies and domain adaptation methods. Another limitation was related to interpretability. While CNNs provided superior accuracy, their "black-box" nature posed challenges for clinical trust. Unlike handcrafted radiomics features, which were explicitly interpretable as intensity, shape, or texture descriptors, the learned features of CNNs were opaque to clinicians. Visualization methods such as saliency maps and gradient-weighted class activation mapping (Grad-CAM) began to emerge as tools for highlighting the regions of input that most influenced predictions, but these techniques were still in their infancy and lacked consistency [13]. Nevertheless, CNN-based frameworks laid the groundwork for a new era of AI in lung cancer detection. They not only achieved superior sensitivity and reduced false-positive rates compared with traditional CAD systems but also demonstrated the feasibility of end-to-end learning pipelines. These pipelines reduced the modular complexity of earlier systems, where pre-processing, feature extraction, and classification were handled separately. By consolidating these steps into unified architectures, CNN-based systems established the foundation upon which more sophisticated frameworks such as transformer-based models and multimodal integration pipelines were subsequently built. Table 4 shows the Comparative Analysis of Early CAD and Deep Learning-Based Methods.

Table 4: Comparative Analysis of Early CAD and Deep Learning-Based Methods

Characteristic	Early CAD Systems	Deep Learning-Based Methods (CNNs, U-Nets, ResNets)
Feature Engineering	Handcrafted radiomics (intensity, shape, texture)	Automatic feature learning from raw voxels and patches
Classifiers	SVMs, decision trees, random forests	End-to-end neural networks, 3D U-Nets, ResNets, DenseNets
Context Modeling	Limited to local handcrafted descriptors	Hierarchical local features, partial volumetric context
Segmentation Accuracy	Coarse, often bounding-box level	Fine-grained delineation with skip connections (U-Net, V-Net)
Detection of Subtle Nodules	Poor sensitivity for ground-glass and juxta-pleural	Improved detection of sub-centimeter and subsolid nodules
Robustness	Poor generalization across scanners and protocols	More robust, but domain shifts still problematic
False Positives	High, often >50 per scan	Reduced to ~1–2 per scan achievable
Interpretability	Transparent (explicit features)	Opaque, requires saliency maps/Grad-CAM
Scalability	Limited by manual feature design	Highly scalable once trained, adaptable via transfer learning

This comparison makes clear that deep learning methods outperformed CAD systems across nearly every dimension, although interpretability and domain robustness remained significant challenges. Figure 3 shows the deep learning based detection and segmentation pipeline. This figure illustrates a representative deep learning pipeline for lung nodule detection and segmentation. At the left, input LDCT scans are pre-processed into normalized patches. These patches feed into a 3D CNN/U-Net architecture, where

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

convolutional layers progressively extract hierarchical features [14]. The encoder compresses input representations, while the decoder with skip connections restores fine spatial detail. Residual blocks and dense connections ensure stable gradient flow and feature reuse. The output is a probability map highlighting candidate nodules, which are post-processed into binary segmentation masks. Below the main pipeline, smaller comparative panels contrast CAD outputs (large number of false positives, coarse bounding regions) with CNN outputs (precise segmentation boundaries, far fewer false alarms). This visual emphasizes the leap in both detection sensitivity and segmentation precision enabled by deep learning.

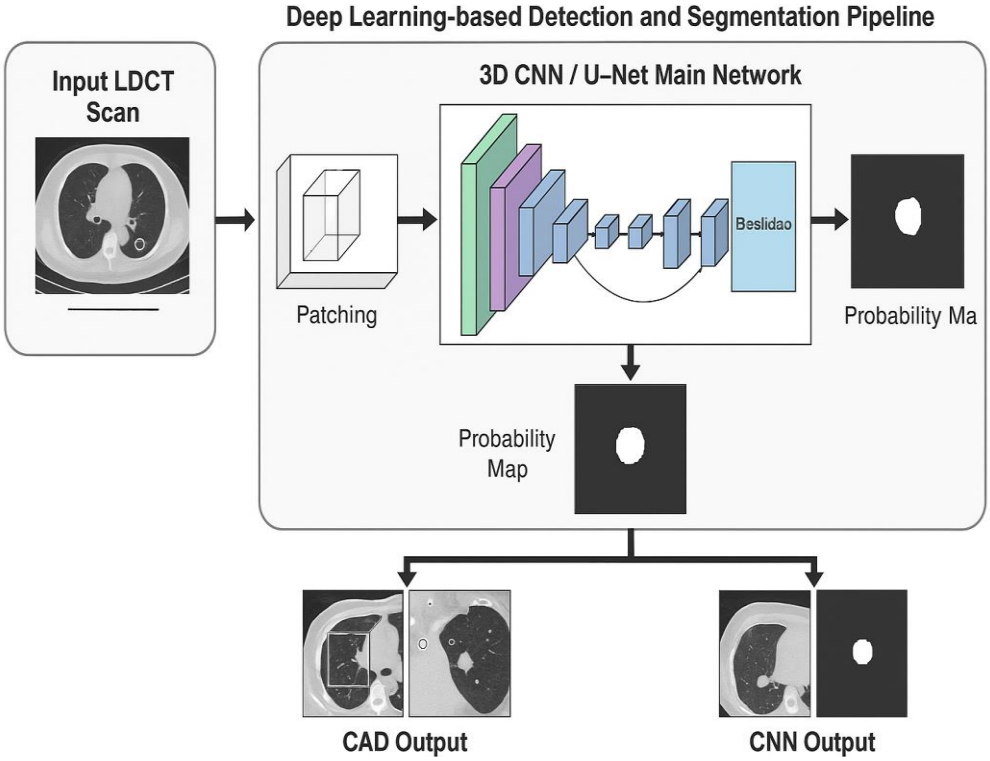


Figure 3: Deep Learning-Based Detection and Segmentation Pipeline

Deep learning methods transformed AI-assisted lung cancer detection by replacing fragile handcrafted pipelines with powerful data-driven architectures. These systems provided unprecedented improvements in detection sensitivity, segmentation precision, and overall clinical applicability. Yet, their reliance on local receptive fields, vulnerability to domain shift, and opaque decision-making processes indicated that further evolution was necessary. These limitations set the stage for the rise of transformer-based architectures and hybrid frameworks, which sought to integrate global context modeling, multimodal data fusion, and interpretability into the next generation of AI-driven lung cancer detection systems.

2.3- Transformer Architectures in Medical Imaging:

The rapid adoption of convolutional neural networks (CNNs) during the past decade transformed lung cancer detection, but their reliance on local receptive fields introduced fundamental limitations. Although CNNs can

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

be stacked to approximate larger contexts, their inductive bias toward locality restricts their capacity to model global dependencies. This becomes especially problematic in lung cancer imaging, where diagnostic accuracy often depends on subtle contextual cues that span across lung lobes or entire CT volumes. For example, distinguishing a true pulmonary nodule from adjacent vasculature, or identifying a subsolid lesion embedded in pleural tissue, requires the model to reason over both localized voxel neighborhoods and distant anatomical structures simultaneously. CNNs, by design, excel at local pattern extraction but struggle to integrate such global information. The introduction of transformer architectures into computer vision and subsequently into medical imaging offered a methodological breakthrough. Originally proposed by Vaswani et al. (2017) for natural language processing, transformers leverage self-attention mechanisms, which allow each element in an input sequence to dynamically attend to all other elements [15]. When applied to imaging, this means that every pixel or voxel patch can directly model relationships with every other region in the scan, regardless of spatial distance. This ability to model global context explicitly marked a departure from the hierarchical locality of CNNs and provided an attractive solution for tasks such as lung nodule detection, where long-range dependencies often hold clinical significance. The Vision Transformer (ViT) was one of the earliest adaptations for computer vision. ViT divides an image or in medical imaging, a CT volume into a set of non-overlapping patches. Each patch is flattened and linearly projected into a fixed-dimensional embedding, after which position encodings are added to preserve spatial ordering. The resulting sequence of patch embeddings is processed through multiple transformer encoder layers, consisting of multi-head self-attention and feed-forward blocks. Applied to lung CT, ViTs can reason across an entire volume, linking nodules with distant contextual features such as vascular orientation or pleural boundaries. However, the direct application of ViTs to volumetric imaging faced three significant challenges:

1. Computational complexity, as self-attention has quadratic scaling with respect to input sequence length, which becomes prohibitive for 3D medical volumes.
2. Data requirements, since transformers lack the strong inductive biases of CNNs and therefore require very large training datasets to generalize effectively.
3. Interpretability, as attention maps, though more transparent than CNN feature maps in theory, often highlight regions that are not intuitively aligned with radiological reasoning.

To overcome these limitations, researchers developed hierarchical and hybrid transformer variants. Among the most influential is the Swin-Transformer (Shifted Window Transformer), which partitions the input volume into overlapping windows and applies self-attention within each window. By hierarchically merging patches across successive layers, Swin-Transformers achieve linear computational complexity while still modeling global dependencies through shifted window connections. In lung cancer detection, Swin-Transformers have shown superior sensitivity for small nodules, outperforming pure CNN models especially in cases involving ground-glass nodules or complex juxta-pleural lesions. An equally important innovation has been the hybridization of CNNs and transformers. In such architectures, CNN layers are used for early-stage local feature extraction, capitalizing on their efficiency and inductive biases, while transformer blocks are introduced at later stages to integrate long-range dependencies. This combination mitigates the data-hungry nature of transformers while leveraging their ability to capture holistic anatomical context. Recent studies on hybrid CNN-transformer models trained on datasets such as LIDC-IDRI and LUNA16 have reported not only higher sensitivity and reduced false positives but also improved robustness when evaluated on external multi-institutional cohorts [16]. The clinical implications of transformer-based

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

approaches are noteworthy. By capturing context beyond local voxel neighborhoods, transformers reduce the incidence of false positives generated by confounding anatomical structures such as vessels, ribs, or airways. They also enhance the detection of subtle morphological features, which is critical for early detection. For example, studies have demonstrated that transformer-based frameworks can improve sensitivity for sub-centimeter nodules by more than 10% compared with state-of-the-art CNNs. Furthermore, their robustness to domain shifts offers a pathway toward real-world generalization, a prerequisite for nationwide lung cancer screening programs. Nevertheless, transformer architectures remain limited by their high computational demands and interpretability challenges. Training a ViT or Swin-Transformer from scratch on volumetric data often requires access to high-performance GPUs or TPUs, along with large annotated datasets that may not be available in medical contexts [17]. Pretraining strategies, such as self-supervised learning on unlabeled medical scans, and transfer learning from natural image datasets, have emerged as practical solutions to alleviate data scarcity. Interpretability research has also advanced, with attention rollout and attention-based attribution maps increasingly used to provide clinicians with visual justifications for transformer decisions. However, their reliability remains under debate. Table 5 shows the Comparative Analysis of CNN-Based and Transformer-Based Methods in Lung Cancer Detection.

Table 5: Comparative Analysis of CNN-Based and Transformer-Based Methods in Lung Cancer Detection

Characteristic	CNN-Based Methods	Transformer-Based Methods (ViT, Swin, Hybrid)
Receptive Field	Local, grows hierarchically with depth	Global via self-attention (direct long-range modeling)
Inductive Bias	Strong (translation equivariance, locality)	Weak (requires large data, learns patterns explicitly)
Feature Extraction	Convolutional filters for local detail	Patch embeddings with positional encodings
Context Modeling	Limited to local/mid-range neighborhoods	Explicit global dependencies across entire CT volume
Computational Complexity	Linear with respect to input size	Quadratic (ViT) or near-linear (Swin with windows)
Data Requirements	Performs well with moderate datasets	Requires large-scale annotated data or pretraining
Strengths	Efficiency, strong local representation	Superior detection of sub-centimeter and juxta-pleural nodules
Weaknesses	Misses global patterns, domain shift vulnerabilities	Data hungry, computationally expensive, less interpretable
Performance in Screening	Good sensitivity, moderate false positives	Higher sensitivity, reduced false positives, robust across domains
Interpretability	Saliency maps, Grad-CAM (clinically interpretable)	Attention maps, attribution rollout (less intuitive)

This comparison highlights how transformer-based models address the contextual limitations of CNNs while introducing new challenges related to computational efficiency and interpretability. The figure 4 illustrates a pipeline beginning with input LDCT scans that

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

are divided into non-overlapping 3D patches. These patches are embedded into tokens with positional encodings and processed through multiple self-attention layers of a Vision Transformer. In a parallel branch, CNN-based convolutional blocks extract local features, which are later fused with transformer outputs in a hybrid design. The classification head outputs candidate nodules with associated malignancy scores, while a segmentation head delineates suspicious regions. Comparative panels show CNN-based outputs missing a juxta-pleural nodule, while the transformer-based model correctly detects it. An attention heatmap panel demonstrates how the transformer distributes attention globally across the lung field, contrasting with the localized saliency of CNNs.

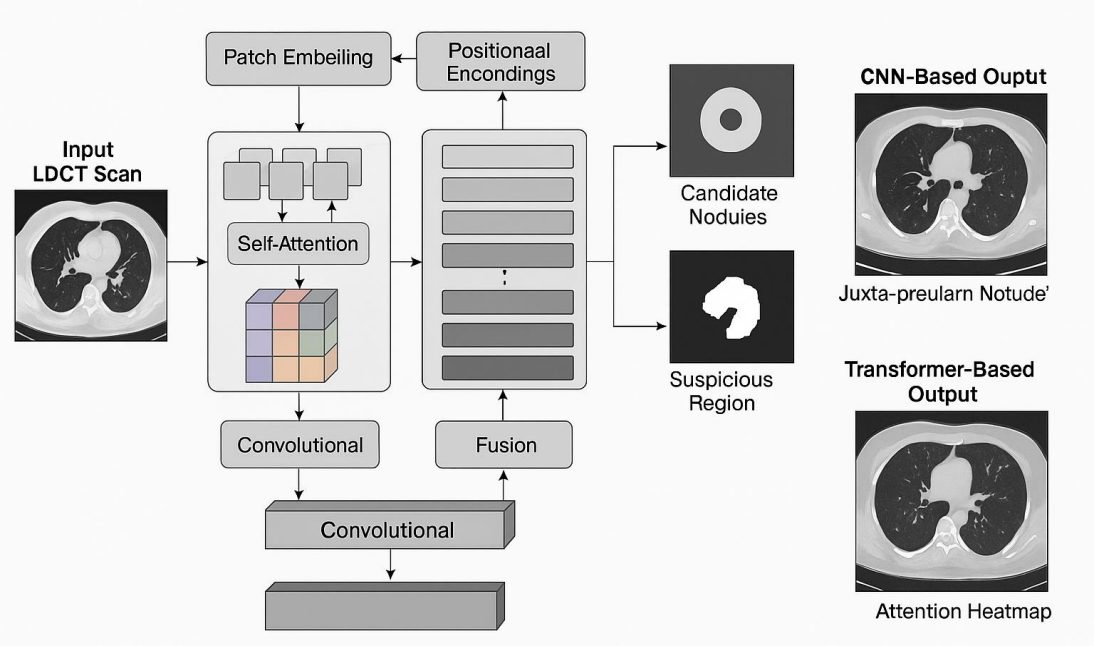


Figure 4: Transformer-Based Pipeline for Lung Nodule Detection

Transformer architectures represent a significant methodological advancement in AI-driven lung cancer detection. By integrating global contextual reasoning with self-attention, they address the primary limitations of CNNs, enhancing sensitivity for subtle nodules and improving robustness across heterogeneous datasets. Although challenges remain particularly computational demands, data scarcity, and interpretability their integration into hybrid frameworks provides a promising direction for next-generation systems. These advances form a crucial stepping stone toward multimodal and governance-oriented frameworks that unify technical excellence with clinical trustworthiness.

2.4- Multimodal Risk Prediction and Radiomics Fusion:

While accurate detection of pulmonary nodules is critical, the central clinical question is not merely whether a nodule exists, but rather whether it is malignant and how it should be managed. Clinical workflows demand not only the identification of abnormalities but also their risk stratification, which informs whether a patient should undergo surveillance, biopsy, or immediate intervention. To this end, AI-driven approaches have evolved beyond detection and segmentation toward malignancy risk prediction frameworks that integrate diverse

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

data sources into a single predictive model. The earliest attempts at malignancy risk prediction were built upon radiomics, a field that systematically quantifies imaging phenotypes through handcrafted features. Radiomics features can be categorized into several families: (i) intensity-based descriptors, such as mean Hounsfield units (HU), histogram skewness, and kurtosis; (ii) shape-based metrics, including sphericity, elongation, compactness, and surface irregularity; and (iii) texture-based measures, such as Haralick gray-level co-occurrence matrices (GLCM), gray-level run-length matrices (GLRLM), and wavelet decompositions [18]. These descriptors capture intra-lesion heterogeneity that may not be visually perceptible to radiologists but often correlates with malignancy. Radiomics thus introduced an early framework for quantifying imaging biomarkers and linking them with patient outcomes. With the rise of deep learning, handcrafted radiomics was increasingly complemented by deep feature embeddings learned directly from imaging data. CNNs and transformers trained for detection tasks produce high-dimensional latent representations that encode complex visual patterns across scales. These embeddings are often more expressive than handcrafted features alone, capturing subtle morphological cues such as spiculation or lobulation. However, deep embeddings are largely uninterpretable and can overfit to dataset biases. By contrast, radiomics features are transparent and clinically interpretable, but less flexible. The fusion of radiomics and deep embeddings therefore represents a synergistic paradigm, combining interpretability with representational power.

Beyond imaging, it is well established that clinical variables including age, sex, smoking history, family history of cancer, and comorbidities such as chronic obstructive pulmonary disease (COPD) are strong predictors of lung cancer risk. These factors are embedded in established risk calculators, such as the Brock model and Herder model, which remain widely used in clinical practice. However, these calculators rely on relatively simple regression-based formulations and fail to integrate the full richness of imaging-derived features. Multimodal AI systems aim to bridge this gap by jointly modeling imaging, clinical, and demographic data, thereby creating risk scores that are both data-driven and clinically grounded. A further dimension comes from longitudinal biomarkers [19]. Nodule growth kinetics quantified through volumetric doubling time, changes in attenuation, or evolving morphology across sequential CT scans are among the most reliable predictors of malignancy. Traditional CAD and CNN pipelines are typically static, analyzing a single time point. Multimodal frameworks extend beyond this limitation by explicitly modeling temporal trajectories. For instance, recurrent neural networks (RNNs), temporal CNNs, and more recently temporal transformers have been applied to capture disease progression dynamics, enabling risk scores that adapt over time. This longitudinal perspective is critical in screening contexts, where follow-up intervals and intervention thresholds depend heavily on growth patterns. To effectively integrate such heterogeneous modalities radiomics, deep embeddings, clinical variables, and temporal biomarkers researchers have turned to advanced fusion strategies. Graph neural networks (GNNs) provide a natural representation, where each modality forms a node type (e.g., imaging node, clinical node, temporal node), and edges encode correlations or progression relationships. Multimodal graph transformers extend this idea by applying self-attention across heterogeneous nodes, allowing the model to weigh contributions from different modalities dynamically [20]. This not only enhances predictive performance but also provides a mechanism for interpreting which modalities or features contributed most to a given risk estimate. From a clinical perspective, multimodal frameworks have demonstrated significant value. Studies consistently show that models integrating imaging with clinical and temporal data outperform imaging-only approaches, particularly in resolving

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

indeterminate nodules that fall into intermediate-risk categories. Decision-curve analyses further highlight their clinical utility by demonstrating greater net benefit across a range of threshold probabilities compared with both guideline-based calculators and imaging-only AI models. Importantly, these systems can reduce unnecessary follow-ups and invasive biopsies for low-risk patients while ensuring that high-risk patients are appropriately triaged for early intervention, thereby improving both efficiency and patient outcomes. Despite their promise, multimodal frameworks face substantial challenges. They are data intensive, requiring large-scale, multi-institutional datasets with harmonized imaging protocols and well-annotated clinical variables. Variability in electronic health record (EHR) structures often results in missing or inconsistent clinical data, complicating integration. Privacy concerns further hinder the pooling of multimodal datasets across institutions, necessitating federated learning paradigms [21]. Additionally, while radiomics features are interpretable, the fusion of deep embeddings and transformer-derived representations creates a black-box risk, which may limit clinician trust. Research into modality-level attribution methods for example, highlighting whether a prediction was driven primarily by imaging, clinical history, or longitudinal changes is ongoing to make these systems more transparent. Table 6 shows the Comparative Analysis of Imaging-Only vs Multimodal Risk Prediction Approaches.

Table 6: Comparative Analysis of Imaging-Only vs Multimodal Risk Prediction Approaches

Approach	Input Modalities	Advantages	Limitations	Clinical Utility
Imaging-Only (Radiomics)	Handcrafted radiomics features	Interpretable, computationally efficient	Limited flexibility, ignores non-imaging risk	Moderate, guideline-adjacent
Imaging-Only (Deep Embeddings)	CNN/Transformer-derived features	High-dimensional, captures subtle morphology	Opaque, dataset bias, requires large data	Good sensitivity but limited trust
Radiomics + Deep Learning Fusion	Handcrafted + learned features	Synergistic, interpretable + expressive	Requires feature alignment, less robust to drift	Better than either alone
Imaging + Clinical Data	Radiomics/embeddings + demographics, history	Aligns with existing risk calculators	Dependent on EHR completeness, variable quality	High, clinically relevant
Imaging + Longitudinal Data	Baseline + follow-up scans, growth kinetics	Captures temporal progression, strong predictor	Requires sequential scans, more complex pipeline	Very high for screening
Full Multimodal Fusion	Imaging + Clinical + Longitudinal	Holistic, highest predictive accuracy, calibrated	Data hungry, interpretability concerns	Superior, guideline-transformative

Volume 3, Issue 7, 2025

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

This comparison underscores the incremental improvements gained as modalities are added, culminating in full multimodal frameworks that best reflect clinical practice. The figure 5 presents a comprehensive framework for multimodal malignancy risk prediction. On the left, CT imaging data flows into two parallel pipelines: a radiomics extraction branch, which generates handcrafted descriptors (intensity histograms, shape metrics, texture matrices), and a deep learning branch, which produces high-dimensional embeddings from CNN or transformer backbones. These imaging features are fused with clinical data inputs (demographics, smoking history, comorbidities) and longitudinal biomarkers (growth kinetics, attenuation changes across follow-up scans). All modalities are represented as nodes in a multimodal graph transformer, where self-attention layers dynamically weigh cross-modal relationships. At the output layer, the model produces a calibrated malignancy risk score with uncertainty intervals, supported by interpretability overlays that indicate the relative contribution of each modality. On the right, a decision-support panel shows how predictions feed into clinical workflows, stratifying patients into "Low-Risk Surveillance," "Intermediate-Risk Follow-Up," or "High-Risk Biopsy/Intervention."

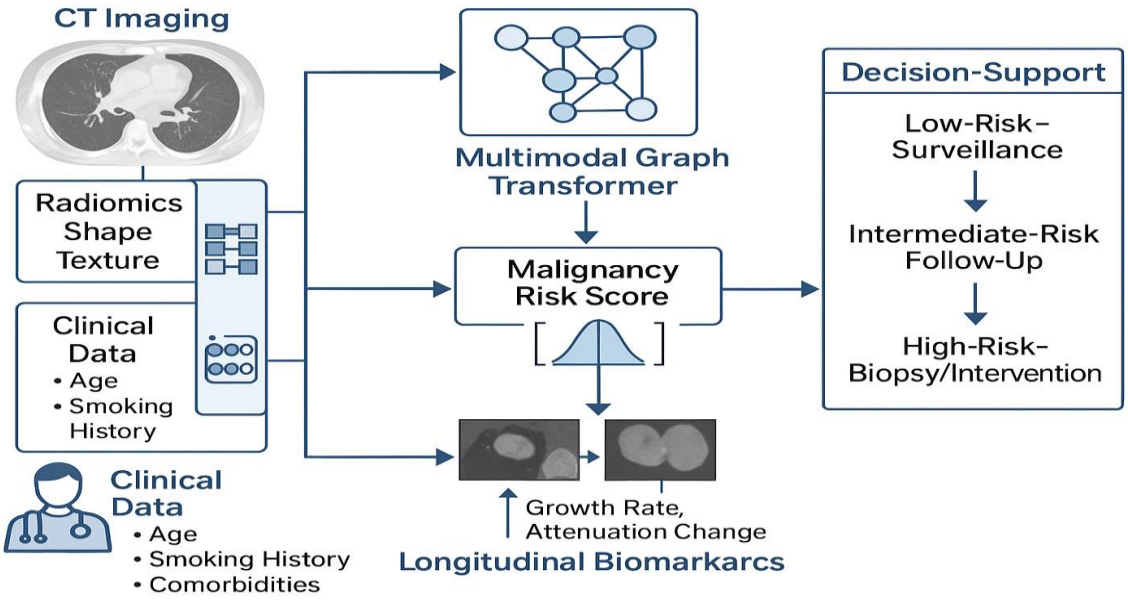


Figure 5: Multimodal Risk Prediction Framework

Multimodal risk prediction and radiomics fusion represent the next stage of evolution in AI-based lung cancer detection. By integrating handcrafted radiomics with deep embeddings, clinical risk factors, and longitudinal biomarkers, these frameworks transcend the limitations of imaging-only models and move closer to the holistic reasoning employed by clinicians. Their ability to provide calibrated, clinically actionable risk scores holds the potential to transform screening programs, reduce unnecessary interventions, and improve early detection outcomes. The challenges of data scarcity, integration complexity, and interpretability remain formidable, but ongoing research in federated learning, self-supervised pretraining, and modality-level attribution is steadily paving the way toward sustainable, trustworthy clinical deployment.

3- Materials and Methods:

The Materials and Methods section provides a comprehensive description of the experimental design, data resources, computational frameworks, and algorithmic strategies that underpin the proposed next-generation AI-

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

driven framework for early lung cancer detection. In alignment with the study's title "Next-Generation AI-Driven Frameworks for Early Lung Cancer Detection Using Advanced Imaging Processing and Predictive Modeling Techniques" this section not only outlines the datasets and preprocessing protocols employed but also details the modeling approaches, fusion strategies, and evaluation methodologies that collectively form the foundation of the proposed system. Unlike traditional imaging studies that primarily focus on detection accuracy in isolated settings, the present framework integrates a wide spectrum of components ranging from physics-informed image harmonization and radiomics-based feature engineering to deep learning embeddings, transformer-based context modeling, and multimodal fusion mechanisms. The goal is to construct an end-to-end pipeline that is technically rigorous, clinically interpretable, and generalizable across diverse populations and imaging environments. This section is therefore organized to reflect the sequential flow of the framework. It begins with a description of the datasets used, including public repositories and multi-institutional screening cohorts, highlighting their composition, imaging protocols, and annotation strategies. Following this, the imaging preprocessing pipeline is described, including harmonization, denoising, super-resolution, and cross-protocol adaptation methods employed to standardize heterogeneous CT data. Next, the discussion moves to feature representation, which encompasses both handcrafted radiomics descriptors and deep feature embeddings derived from CNNs and transformer backbones, as well as longitudinal biomarkers that capture temporal changes across follow-up scans. The subsequent subsections address the fusion framework, where multimodal data including imaging, clinical variables, and temporal biomarkers are integrated using advanced architectures such as graph neural networks and multimodal transformers. Detailed attention is also given to the model training strategy, including curriculum learning, distributionally robust optimization, and federated learning protocols for privacy-preserving collaboration across institutions [22]. To ensure reliability and trustworthiness, methods for probability calibration, uncertainty quantification, and interpretability are presented, demonstrating how the framework provides both actionable predictions and transparent evidence for clinical decision support. Finally, the section outlines the evaluation design and metrics, detailing how the framework was validated across public benchmarks and multi-institutional cohorts. Standard metrics such as sensitivity, specificity, AUC, false positives per scan, and calibration error are reported, alongside decision-curve analyses that highlight clinical utility compared with existing guideline-based approaches.

3.1- Dataset Description:

The reliability and generalizability of any artificial intelligence framework for lung cancer detection depend fundamentally on the datasets used for its training, validation, and external benchmarking. In this study, a combination of publicly available benchmark repositories and multi-institutional clinical cohorts was employed in order to capture a wide spectrum of imaging heterogeneity, patient demographics, and annotation standards. By integrating these complementary sources, the proposed framework was designed not only to achieve state-of-the-art performance on established public datasets but also to ensure its robustness and applicability in real-world clinical environments. The LIDC-IDRI (Lung Image Database Consortium – Image Database Resource Initiative) formed the backbone of the training corpus. This dataset includes 1,018 thoracic CT scans collected from multiple institutions, with each scan reviewed independently by up to four expert thoracic radiologists. For every annotated lesion, malignancy likelihood scores were assigned along with binary nodule/non-nodule classifications and segmentation masks. The inclusion of multiple radiologist assessments

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

provides a unique opportunity to model inter-observer variability, which is a persistent challenge in clinical screening. The LIDC-IDRI dataset thus plays a critical role not only in providing annotated training data but also in serving as a source for studying uncertainty and consensus-driven ground truth generation. Derived from LIDC-IDRI, the LUNA16 (LUng Nodule Analysis 2016) dataset was also a major resource. This curated subset consists of 888 CT scans containing nodules greater than or equal to 3 mm, with smaller ambiguous candidates excluded to standardize evaluation [23]. The dataset has been central to the development of automated detection methods because of its standardized benchmarking framework, which evaluates sensitivity across multiple false-positive thresholds. The availability of an official leaderboard and community benchmarks has made LUNA16 a de facto standard for comparing competing algorithms, and its use in this framework enabled fair comparisons with prior state-of-the-art systems.

The NLST (National Lung Screening Trial) provided an additional dimension by introducing a population-scale screening context. Unlike LIDC-IDRI and LUNA16, which focus on annotated nodules, NLST is a randomized controlled trial comprising more than 26,000 low-dose CT scans obtained as part of a large-scale effort to evaluate lung cancer screening's effectiveness in reducing mortality. While its annotations are less granular, the dataset reflects the variability and complexity encountered in real-world screening programs, including normal cases, benign incidental findings, and confirmed malignancies. Integrating NLST into the evaluation pipeline allowed the framework to be tested not only on detection accuracy but also on population-level applicability, particularly in the stratification of screening outcomes. Additional datasets such as the Tianchi Lung Nodule Challenge dataset and the Kaggle Data Science Bowl (DSB) repository were used as supplementary resources to diversify the training corpus. These datasets provide a mixture of radiologist annotations and cancer outcome labels, adding further heterogeneity in terms of scanner vendors, acquisition protocols, and geographic populations. Their inclusion ensured that the framework was not overfitted to the imaging distributions of Western clinical trials alone, but was instead exposed to broader demographic and technical variation. In addition to public repositories, the framework was validated on multi-institutional clinical cohorts where accessible. These cohorts comprised CT scans obtained from diverse healthcare centers with varying scanner hardware, acquisition parameters, reconstruction kernels, and patient populations. Unlike curated public datasets, clinical cohorts reflect the inconsistencies, noise, and incomplete annotations characteristic of real-world practice. Their inclusion was essential for assessing the external validity of the framework and for demonstrating its resilience to distributional shift a key barrier to the clinical deployment of AI systems. Ground truth definitions varied across datasets, requiring careful harmonization. In LIDC-IDRI, malignancy scores were derived from radiologist consensus, which captures subjective assessments but does not always correspond to pathology. In NLST and some clinical cohorts, biopsy-confirmed malignancy outcomes were available, providing a higher standard of truth but often with fewer annotated nodules. This study therefore employed a hybrid approach: radiologist annotations were used as the primary ground truth for detection and segmentation tasks, while malignancy risk prediction tasks prioritized pathology-confirmed outcomes where available [24]. The combination of these datasets ensured comprehensive coverage across the spectrum of clinically relevant variability. The framework was exposed to nodules of diverse sizes, ranging from subtle sub-centimeter lesions to large solid masses, and to the full morphological spectrum including solid, part-solid, and ground-glass nodules. It was also trained and validated on patients

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

spanning a wide range of ages, smoking histories, and comorbidities, thereby ensuring that the system's predictions were generalizable across diverse screening populations. The integration of public and clinical datasets also addressed acquisition heterogeneity, with scans originating from multiple vendors, reconstruction techniques, and dose levels, a factor known to significantly impact AI model performance if not adequately represented. Table 7 shows the Overview of Datasets Used for Development and Evaluation.

Table 7: Overview of Datasets Used for Development and Evaluation

Dataset	Number of Scans	Annotation Type	Distinctive Features	Role in Study
LIDC-IDRI	1,018	Radiologist annotations, malignancy scores	Multi-reader annotations; consensus modeling	Training and validation
LUNA16	888	Curated nodules ≥3 mm	Standardized benchmark with official metrics	Benchmarking nodule detection
NLST	>26,000	Screening outcomes, some pathology labels	Population-scale trial; includes normal and incidental findings	External validation, risk prediction
Tianchi	~1,000	Radiologist annotations	Diverse acquisition settings; non-Western populations	Supplementary training
Kaggle DSB	~2,000	Cancer outcome labels	Outcome-driven dataset; used in competitions	Risk prediction benchmarking
Clinical Cohorts	Variable	Radiologist + pathology labels	Multi-institutional heterogeneity; real-world complexity	External validation of robustness

The figure 6 illustrates the landscape of datasets incorporated into the framework. On the left, public repositories (LIDC-IDRI, LUNA16, NLST, Tianchi, and Kaggle DSB) are represented as layered blocks feeding into the training and benchmarking pipeline. On the right, multi-institutional clinical datasets are depicted with variable acquisition parameters, emphasizing their role in external validation and generalization. At the center, a harmonization layer integrates radiologist consensus labels and pathology-confirmed outcomes, which flow into detection, segmentation, and risk prediction modules of the proposed AI framework. The figure highlights how the strategic combination of datasets provides both methodological rigor (through benchmarking) and translational relevance (through real-world cohorts).

Volume 3, Issue 7, 2025

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

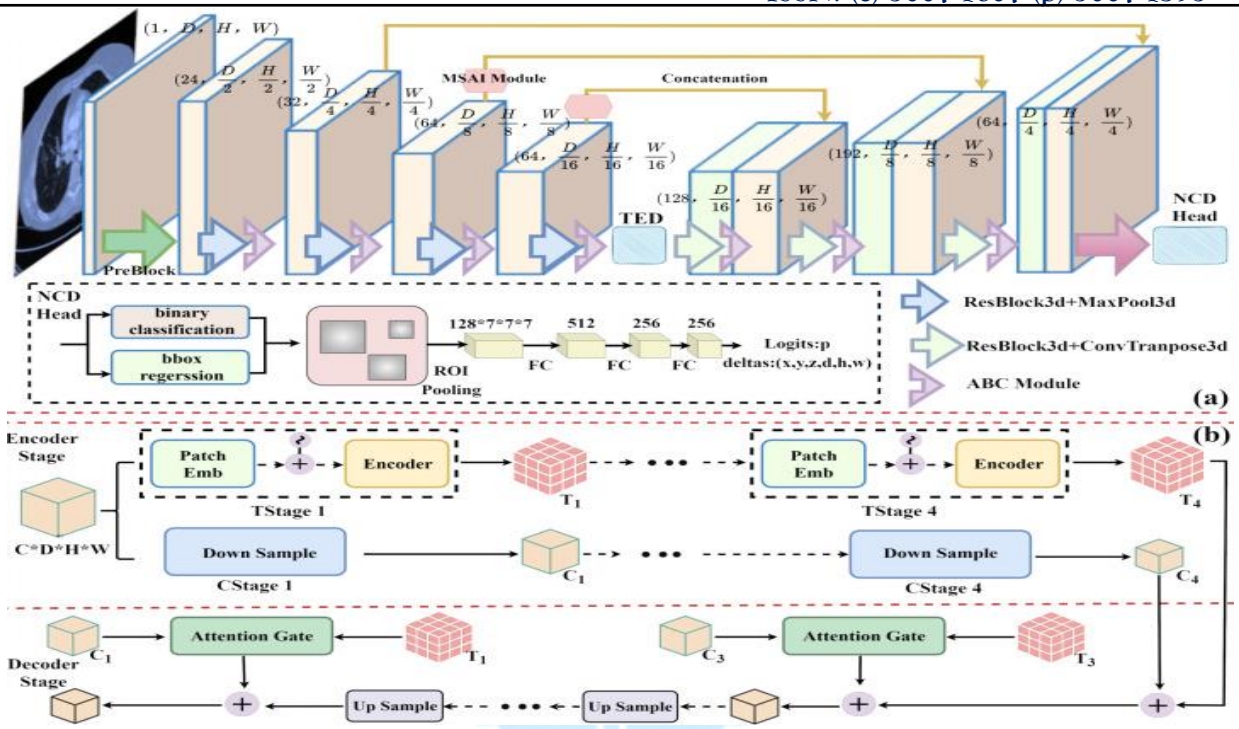


Figure 6: Dataset Landscape for AI-Driven Lung Cancer Detection

The dataset design for this study represents a deliberate effort to balance the controlled precision of curated repositories with the messy heterogeneity of real-world clinical data. Public datasets such as LIDC-IDRI and LUNA16 provided the annotated granularity necessary for training robust detectors, while population-scale trials like NLST and multi-institutional cohorts enabled population-level validation and external benchmarking. By combining these complementary resources, the proposed framework was constructed on a foundation that emphasizes both technical performance and clinical generalizability, a dual requirement for sustainable translation into lung cancer screening programs.

3.2- Imaging Preprocessing Pipeline:

The preprocessing of CT imaging data constitutes a critical step in the design of AI-based frameworks for lung cancer detection. Unlike natural image analysis, where raw pixel values can often be directly fed into deep learning architectures, thoracic CT imaging data are subject to substantial heterogeneity that arises from variations in scanner hardware, acquisition protocols, reconstruction kernels, and radiation dose levels. In population-scale screening studies, such as those included in the NLST, scans may differ not only across institutions but also within the same institution over time, as hardware and protocols evolve. Without robust preprocessing, such variability can severely compromise the generalizability and stability of AI models, leading to performance degradation when applied to external cohorts. The proposed framework employs a multi-stage imaging preprocessing pipeline that was specifically designed to address these challenges and to create standardized, analysis-ready inputs for subsequent feature extraction and modeling. The pipeline is informed by both physics-based principles of CT imaging and data-driven harmonization strategies, thereby ensuring that diagnostically relevant information is preserved while confounding variability is minimized. The first component of the pipeline is image harmonization, which tackles the problem of cross-institutional and cross-scanner variability. Since CT images are reconstructed in Hounsfield Units (HU), in principle they

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

provide standardized intensity values [25]. However, in practice, scanner calibration differences, reconstruction algorithms, and dose modulation introduce significant inconsistencies. Harmonization was achieved through histogram matching across datasets, intensity normalization to fixed HU ranges, and domain adaptation methods that align feature distributions across acquisition protocols. Recent advances in adversarial domain adaptation and physics-informed generative models were also incorporated to further minimize distributional drift. This step was especially important for multi-institutional validation, as models trained on LIDC-IDRI or LUNA16 often fail to generalize directly to NLST or clinical cohorts without harmonization.

The second component is denoising and super-resolution. Low-dose CT (LDCT), which is essential for minimizing patient radiation exposure during screening, inherently suffers from higher noise levels and lower contrast resolution compared with diagnostic-dose scans. Noise reduction is critical to prevent false positives arising from random speckles, while resolution enhancement is essential for identifying sub-centimeter nodules that represent the earliest stages of malignancy. Traditional denoising methods, such as Gaussian smoothing or wavelet thresholding, often compromised diagnostically relevant textures. To overcome this, advanced techniques were used, including non-local means denoising and deep learning–based denoising autoencoders trained on paired high- and low-dose CT data. Super-resolution was implemented using convolutional neural networks (SRCNN) and generative adversarial networks (SRGAN), which reconstruct high-resolution textures from LDCT inputs. By integrating denoising and super-resolution, the pipeline enhanced the visibility of ground-glass opacities and subtle subsolid nodules, which are among the most clinically significant but most difficult to detect. The third component involves lung field segmentation, which isolates the lung parenchyma from surrounding anatomical structures such as ribs, mediastinum, and subcutaneous tissue [26]. This step reduces the search space for nodule detection and minimizes false positives arising from non-pulmonary regions. Automated segmentation was performed using a 3D U-Net architecture trained on annotated lung masks, supplemented by rule-based region-growing algorithms for refinement. Morphological operations ensured that small nodules adjacent to the pleura or airway walls were preserved rather than excluded, addressing a common failure mode of earlier segmentation approaches. Segmentation also facilitates downstream radiomics analysis by providing clearly defined regions of interest for feature extraction. The fourth component of the pipeline is candidate generation and normalization. After lung segmentation, potential nodule candidates were identified using blob-detection filters, Hessian-based vesselness suppression, and adaptive thresholding methods. Candidate regions were then cropped into fixed-size patches (e.g., 64×64×64 voxels), which were resampled to isotropic voxel spacing of 1 mm^3. This ensured uniform input size across datasets and allowed 3D convolutional neural networks and transformer architectures to process volumetric data consistently. Candidate normalization was essential for enabling cross-dataset training, as raw CT scans can have slice thicknesses ranging from sub-millimeter to several millimeters, significantly affecting apparent nodule morphology. Finally, preprocessing outputs were subjected to quality control and augmentation. Cases with incomplete lung coverage, severe artifacts, or corrupted annotations were flagged and excluded. To augment training diversity, transformations such as random rotations, flips, elastic deformations, and intensity perturbations were applied. These augmentations mimic the variability encountered in clinical practice and enhance the generalizability of the models. Importantly, augmentation was applied only after harmonization and normalization to avoid amplifying scanner-specific biases. Taken

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

together, this pipeline not only reduced technical variability but also preserved subtle diagnostic cues essential for early detection [27]. By systematically harmonizing, denoising, segmenting, and normalizing the input data, the pipeline ensured that downstream detection, segmentation, and risk prediction modules operated on consistent, high-quality inputs. Clinically, this translates to improved sensitivity for subtle lesions, reduced false positives, and enhanced trustworthiness when deploying the framework across heterogeneous patient populations. Table 8 shows the Imaging Preprocessing Pipeline for AI-Driven Lung Cancer Detection.

Table 8: Imaging Preprocessing Pipeline for AI-Driven Lung Cancer Detection

Stage	Techniques Employed	Purpose	Clinical Impact
Harmonization	HU standardization, histogram matching, adversarial domain adaptation, physics-informed generative models	Align intensity distributions across scanners and protocols	Minimizes domain shift, ensures robustness in multi-center validation
Denoising & Super-Resolution	Non-local means filtering, deep denoising autoencoders, SRCNN, SRGAN	Reduce noise in LDCT, enhance fine anatomical detail	Improves visibility of subtle ground-glass and subsolid nodules
Lung Field Segmentation	3D U-Net architectures, rule-based region growing, morphological refinements	Isolate parenchyma while preserving pleural and airway-adjacent nodules	Reduces false positives, ensures comprehensive candidate coverage
Candidate Normalization	Blob detection, vesselness suppression, adaptive thresholding, isotropic resampling (1 mm^3), fixed patch extraction	Standardize candidate input size and resolution across scans	Enables consistent training across heterogeneous datasets
Quality Control & Augmentation	Exclusion of corrupted cases; augmentations (rotations, flips, elastic deformations, intensity shifts)	Expand dataset diversity, mimic real-world variability	Improves generalizability and resilience to clinical variability

The figure 7 depicts the sequential preprocessing pipeline as a left-to-right flowchart. The leftmost panel shows raw LDCT scans with visible noise and cross-institution variability. The first block, Harmonization, illustrates how intensity histograms from different scanners are aligned into a standardized HU distribution. The second block, Denoising and Super-Resolution, presents a side-by-side sub-figure showing noisy scans and their enhanced reconstructions, highlighting improved visibility of subsolid nodules. The third block, Lung Segmentation, includes small sub-figures comparing whole thoracic CT slices with isolated parenchymal masks, emphasizing the exclusion of ribs and mediastinal structures. The fourth block, Candidate Normalization, shows candidate nodule patches cropped and resampled into fixed voxel dimensions. Finally, a rightmost panel labeled Preprocessed Output illustrates clean, standardized nodule patches flowing into deep learning pipelines for detection, segmentation, and malignancy prediction. The inclusion of smaller before-and-after sub-figures inside each stage makes the diagram complex, multi-layered, and clinically intuitive.

Volume 3, Issue 7, 2025

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

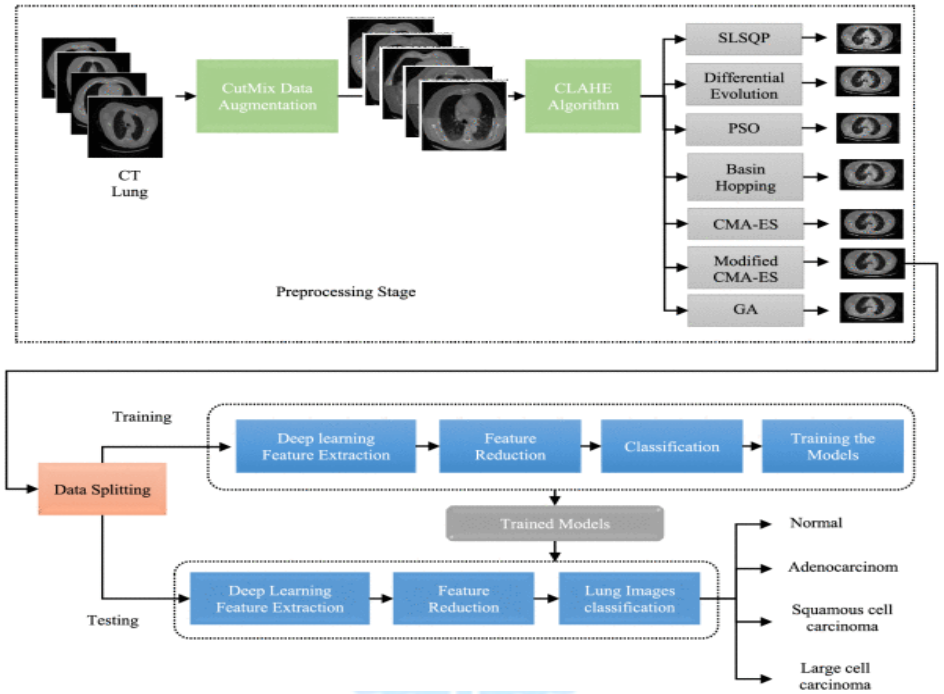


Figure 7: Imaging Preprocessing Pipeline with Sequential Stages

The imaging preprocessing pipeline serves as the critical bridge between raw heterogeneous CT scans and standardized inputs for AI modeling. By combining physics-based harmonization with advanced denoising, segmentation, and normalization strategies, the pipeline ensures that the framework operates with high sensitivity and robustness across diverse datasets and clinical settings. This rigorous preprocessing not only enhances algorithmic performance but also addresses the practical challenges of real-world deployment in lung cancer screening ecosystems.

3.3- Feature Extraction and Representation:

Following preprocessing, the central task is to transform imaging data into structured, discriminative representations that capture the multi-dimensional characteristics of pulmonary nodules. This stage is pivotal because the quality and richness of feature representations determine how effectively downstream models can differentiate between benign and malignant lesions. In the context of next-generation AI-driven frameworks, feature extraction has evolved from handcrafted descriptors rooted in radiological expertise to data-driven embeddings learned by deep architectures, with recent advances moving toward hybrid models that incorporate temporal dynamics, multi-scale feature hierarchies, and modality fusion. Historically, radiomics emerged as one of the earliest systematic approaches to quantitative imaging analysis. Radiomics aims to convert medical images into a high-dimensional set of features that reflect nodule morphology, intensity distribution, and textural heterogeneity. For example, intensity-based features quantify statistical distributions of voxel values, such as mean, median, skewness, and kurtosis in Hounsfield Units. Shape-based descriptors measure geometric attributes such as sphericity, compactness, elongation, and irregularity, which often correlate with tumor invasiveness. Texture features, derived from gray-level co-occurrence matrices (GLCM), gray-level run-length matrices (GLRLM), and neighborhood gray-tone difference matrices (NGTDM), capture intra-nodular heterogeneity that may

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

signal early malignancy. In addition, filter-based radiomics, including wavelet decompositions and Laplacian of Gaussian (LoG) transformations, enable multi-scale characterization of lesion patterns, making it possible to identify subtle phenotypic signatures beyond the sensitivity of human readers [28]. Radiomics offers transparency and interpretability, allowing features to be directly mapped to established radiological descriptors such as spiculation, lobulation, or ground-glass opacity. Complementing radiomics are deep learning–based embeddings, which automatically learn feature hierarchies directly from data. Convolutional Neural Networks (CNNs) serve as hierarchical feature extractors, beginning with low-level edges and gradients and progressing toward complex visual concepts such as lobulated borders or vascular convergence. Three-dimensional CNNs extend this ability by capturing volumetric continuity across CT slices, which is essential for accurately characterizing sub-centimeter nodules. More advanced architectures, including Residual Networks (ResNets) and DenseNets, have demonstrated superior representation learning by mitigating vanishing gradients and promoting feature reuse.

Recently, transformer-based architectures have revolutionized representation learning in medical imaging by introducing global self-attention mechanisms. Unlike CNNs, which focus on local receptive fields, transformers capture long-range contextual dependencies within entire CT volumes. Models such as the Vision Transformer (ViT) and Swin-Transformer have been adapted to lung cancer imaging, allowing nodules to be analyzed not only as isolated regions but also within their anatomical context. This shift is particularly important for juxta-pleural nodules or lesions adjacent to vascular structures, where global context helps disambiguate nodule boundaries and improves risk stratification. Recognizing the complementary strengths of these paradigms, the proposed framework adopts a fusion strategy that integrates radiomics with deep embeddings**.** Radiomics provides domain-grounded interpretability, ensuring that AI outputs remain clinically explainable, while deep embeddings capture complex and subtle features that are beyond handcrafted design. Fusion is achieved through concatenation of feature spaces and refined using multimodal graph transformers, which dynamically weigh the contribution of each feature set depending on the clinical context [29]. For instance, in small subsolid nodules, texture radiomics may dominate prediction, while in large heterogeneous lesions, deep embeddings may contribute more heavily. Beyond static features, the framework incorporates longitudinal biomarkers to model disease dynamics across time. In clinical practice, the growth rate of pulmonary nodules often expressed as volume doubling time (VDT) is among the strongest indicators of malignancy. Attenuation changes, such as transitions from ground-glass to part-solid morphology, provide additional diagnostic cues. To encode these dynamics, temporal models such as Recurrent Neural Networks (RNNs)**,** Temporal Convolutional Networks (TCNs), and Temporal Transformers are employed. These architectures model sequential dependencies across follow-up CT scans, enabling the system to predict malignancy risk trajectories rather than single time-point estimates. This temporal dimension is critical for screening programs, where decisions to biopsy or continue surveillance hinge on observed growth patterns. An equally important aspect of representation is multi-scale integration. Malignant nodules often exhibit heterogeneity not only within the lesion itself but also in their surrounding environment, such as vascular convergence, airway displacement, or pleural retraction. By extracting features at multiple spatial scales ranging from fine intra-nodular textures to global parenchymal context the framework ensures a holistic characterization. Multi-scale CNNs and pyramid-based transformers were incorporated to explicitly encode these nested contexts. The feature extraction and representation stage thus forms a multi-layered process: radiomics provides handcrafted

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

descriptors rooted in radiological knowledge, deep embeddings capture latent and expressive representations, longitudinal biomarkers add temporal dynamics, and multi-scale integration ensures contextual completeness. Together, these representations serve as the backbone for the multimodal fusion framework described in subsequent sections, ultimately enabling accurate, interpretable, and clinically actionable malignancy risk predictions. Table 9 shows the Comparative Framework for Feature Extraction and Representation.

Table 9: Comparative Framework for Feature Extraction and Representation

Feature Type	Techniques/Examples	Advantages	Limitations	Clinical Impact
Radiomics	Intensity histograms, GLCM, GLRLM, NGTDM, wavelet features, shape metrics	Transparent, interpretable, aligns with radiological reasoning	Sensitive to segmentation errors; limited generalizability	High — links AI output to known imaging biomarkers
Deep Embeddings (CNNs)	3D CNNs, ResNet, DenseNet	Captures subtle morphology, hierarchical feature learning	Requires large labeled datasets, less interpretable	High sensitivity for small/complex nodules
Transformer Embeddings	Vision Transformer, Swin-Transformer	Captures global context, models long-range dependencies	Computationally intensive, data hungry	Crucial for juxta-pleural and complex nodules
Fusion (Radiomics + Deep)	Feature concatenation, multimodal graph transformers	Balances interpretability and flexibility, dynamic weighting	Requires alignment of feature spaces	Superior generalization and explainability
Longitudinal Biomarkers	Volume doubling time, attenuation change, temporal CNNs/transformers	Models disease dynamics across time, critical for screening	Requires sequential follow-up scans, increased complexity	Clinically decisive in early-stage management

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

Multi-Scale Integration	Pyramid CNNs, multi-scale transformers	Captures both local textures and global parenchymal context	Increases computational load	Mimics radiologists' holistic interpretation

The figure 8 illustrates a three-branch pipeline for feature extraction. On the left, radiomics extraction is shown, with segmented nodules feeding into intensity, shape, and texture feature modules, represented by smaller sub-figures of histograms and co-occurrence matrices. In parallel, the deep learning branch depicts preprocessed nodule patches being processed by 3D CNNs and transformers, illustrated with feature maps evolving from low-level edges to high-dimensional embeddings. Below, a longitudinal branch shows sequential CT scans connected in a temporal chain, feeding into recurrent/transformer blocks that encode growth kinetics and attenuation trajectories. At the center, outputs from all branches converge into a fusion and multi-scale integration module, represented as a layered graph transformer dynamically weighing modalities. On the right, the unified feature representation emerges as a composite vector, feeding into downstream malignancy prediction and risk stratification modules. Smaller embedded sub-figures within each block provide visual cues of feature types (e.g., texture maps, CNN feature maps, temporal graphs), making the diagram complex, multi-layered, and visually rich.

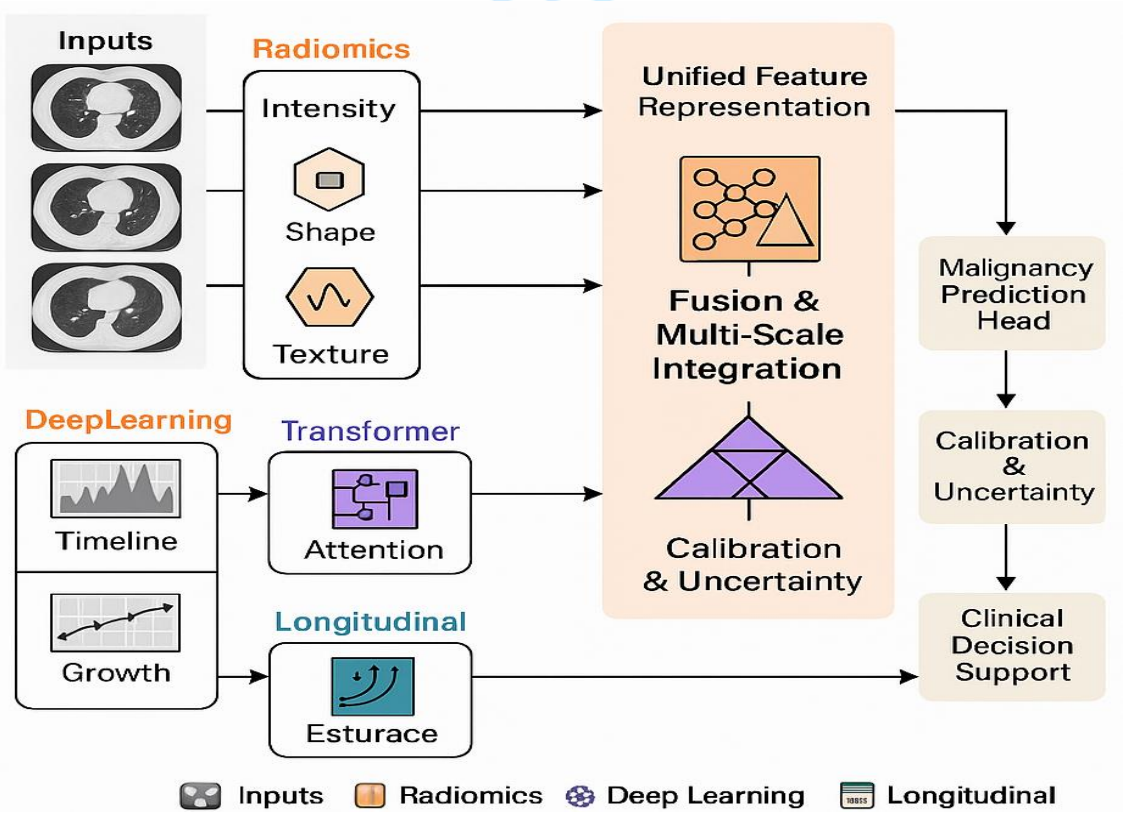


Figure 8: Feature Extraction and Representation Framework

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

Feature extraction and representation serve as the cornerstone of the proposed framework, encoding the multidimensional complexity of pulmonary nodules into structured, quantitative descriptors that combine transparency, expressiveness, temporal awareness, and contextual depth. By integrating radiomics, deep embeddings, longitudinal biomarkers, and multi-scale information, the framework creates a robust foundation for multimodal fusion and predictive modeling, bridging the gap between algorithmic innovation and clinically meaningful outcomes.

3.4- Fusion and Multimodal Integration

The development of clinically deployable AI systems for lung cancer detection requires not only accurate feature extraction but also the effective integration of heterogeneous data modalities into a coherent predictive framework. Radiomics features, deep learning embeddings, clinical metadata, and longitudinal biomarkers each provide complementary information, but without careful fusion these modalities risk either redundancy or imbalance in their contributions. The proposed framework employs a multimodal integration strategy designed to respect the unique statistical properties of each input type while creating a unified, patient-level malignancy risk representation. At the foundational level, radiomics descriptors and deep embeddings are projected into a shared latent space. Radiomics, derived from handcrafted measurements of intensity, shape, and texture, retain close alignment with established radiological descriptors such as spiculation, lobulation, and heterogeneity. These features provide interpretability and transparency, ensuring that predictions can be linked back to clinically recognized imaging biomarkers. In contrast, deep embeddings, derived from convolutional and transformer-based backbones, encode nonlinear and high-dimensional representations that can capture complex spatial relationships and subtle phenotypic changes imperceptible to human readers. By aligning these two feature spaces early in the pipeline, the model establishes a foundation that combines interpretability with representational power. The next integration step incorporates clinical metadata. Variables such as age, sex, smoking history, family history, and comorbidities are encoded through embedding layers that transform both categorical and numerical data into continuous representations [30]. These embeddings are dimensionally aligned with imaging-derived features, thereby enabling joint optimization during training. Clinical metadata are especially valuable in borderline cases, where imaging findings alone may be insufficient to establish malignancy risk. For example, small indeterminate nodules in a lifelong non-smoker carry different implications than similar nodules in a heavy smoker with chronic obstructive pulmonary disease.

In parallel, longitudinal biomarkers particularly volume doubling time (VDT), attenuation trajectory, and morphological transitions from ground-glass to part-solid are modeled using temporal transformers. These architectures encode sequential dependencies across follow-up CT scans, enabling the framework to learn disease trajectories rather than relying on static imaging snapshots. This temporal awareness is critical for clinical decision support, as guidelines often emphasize growth kinetics in surveillance protocols. The unified integration of all these modalities is achieved through multimodal graph transformers (MMGTs). In this architecture, each modality (radiomics, deep embeddings, clinical metadata, longitudinal biomarkers) is represented as a node, and the edges encode inter-modality dependencies. Through global self-attention mechanisms, the model dynamically adjusts the weighting of modalities depending on the diagnostic context. In subsolid nodules, textural radiomics features are often emphasized, while in heterogeneous or juxta-pleural nodules, transformer-based embeddings provide greater discriminative power [31]. This

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

dynamic weighting ensures that the fusion strategy is not static but instead adapts to the complexities of individual cases. The fusion process is staged hierarchically. An early fusion stage aligns radiomics with CNN-derived embeddings, creating a hybrid representation that balances interpretability and expressive power. An intermediate fusion stage incorporates longitudinal biomarkers, allowing temporal disease trajectories to modulate the imaging-based predictions. Finally, a late fusion stage integrates clinical metadata with the imaging–temporal representation to generate a comprehensive patient-level malignancy risk vector. This hierarchical structure reduces the risk of modality dominance and enhances generalization across populations and imaging environments. The comparative performance of alternative fusion strategies is summarized in Table 10. Simple early fusion through concatenation is computationally efficient but lacks contextual adaptability. Graph-based fusion provides greater flexibility but can be computationally intensive when applied directly to raw feature sets. The hierarchical multimodal graph transformer achieves the most robust balance, dynamically weighting modalities and consistently improving generalization across external validation cohorts.

Table 10: Comparative Analysis of Fusion Strategies in Multimodal Lung Cancer Prediction

Fusion Strategy	Methodology	Strengths	Limitations	Role in Proposed Framework
Early Fusion	Direct concatenation of radiomics and CNN embeddings	Simple implementation; low computational cost	Lacks contextual awareness; modality imbalance risk	Used as baseline reference
Graph-Based Fusion	Multimodal graph transformers with self-attention	Captures inter-modality dependencies; dynamic weighting	High computational complexity; requires large data	Secondary evaluation model
Hierarchical Fusion	Stage-wise integration: imaging → temporal → clinical	Balances interpretability, temporal awareness, and generalizability	Complex optimization; longer training time	Core deployment model

To aid conceptual understanding, Figure 9 depicts the multimodal fusion architecture. On the left, three parallel input streams are shown: (i) radiomics features, represented by histograms, shape models, and texture matrices; (ii) deep embeddings, represented by feature maps evolving from CNN filters to transformer tokens; and (iii) clinical and longitudinal biomarkers, represented as tabular embeddings and temporal sequences. At the center of the figure lies the multimodal graph transformer, depicted as a network of nodes with dynamic attention-weighted edges. This central module merges the modalities through hierarchical fusion, producing a composite risk representation. On the right, the unified patient-level representation flows into downstream malignancy classification and survival prediction heads. Smaller embedded panels within each input block illustrate example feature types texture co-occurrence maps for radiomics, activation maps for CNNs, and growth curves for longitudinal biomarkers making the figure visually complex and clinically intuitive.

Volume 3, Issue 7, 2025

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

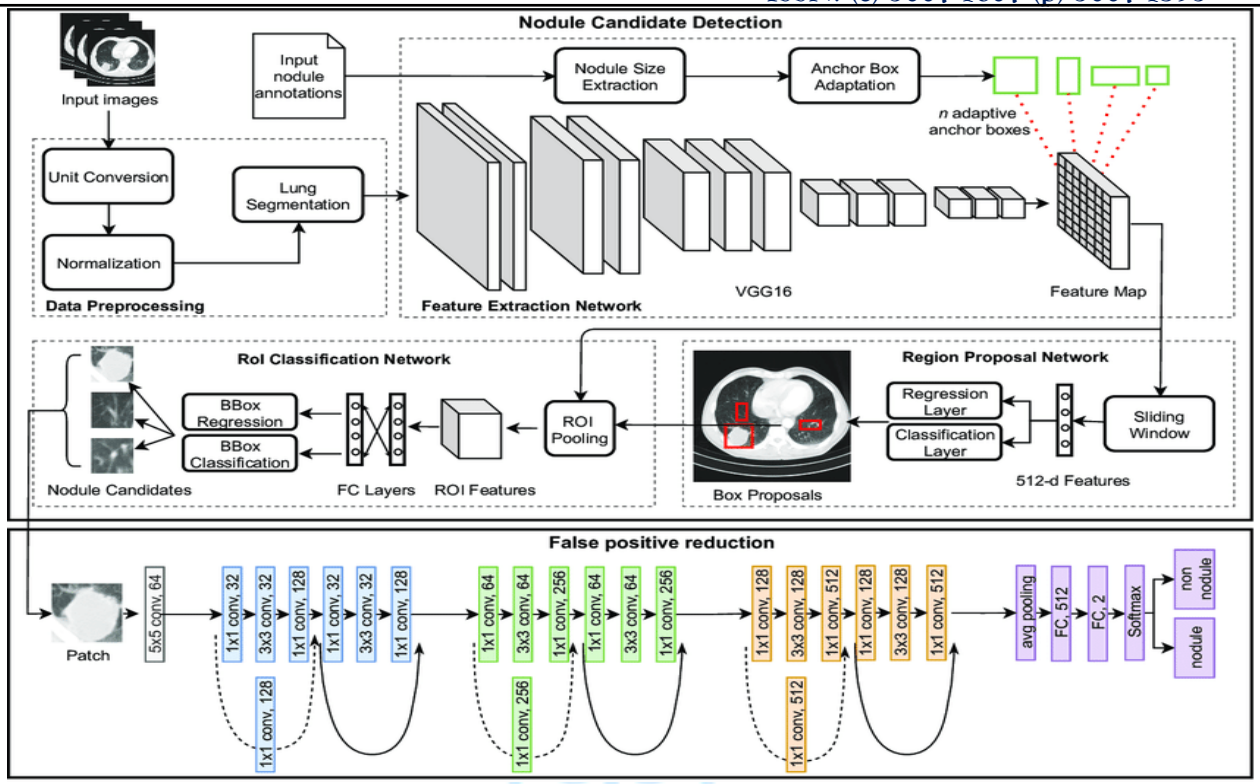


Figure 9: Multimodal Fusion Architecture for AI-Driven Lung Cancer Detection

By embedding multimodal integration into the pipeline, the framework ensures that predictions are not driven by imaging features alone but are instead contextualized by patient history and temporal evolution. This approach mirrors real-world clinical reasoning, where radiologists and oncologists synthesize imaging, patient demographics, and longitudinal evidence before making diagnostic and therapeutic decisions. The resulting model is therefore not only more accurate but also more aligned with clinical practice, enhancing its potential for adoption in lung cancer screening ecosystems.

3.5- Model Training Strategy:

The effectiveness of the proposed next-generation AI-driven framework ultimately depends not only on its architectural innovations but also on the rigor of its training strategy. Designing a robust training methodology is particularly challenging in the context of early lung cancer detection, where data are highly imbalanced, protocols vary across institutions, and ground truth is inconsistently defined. The training process therefore had to be carefully engineered to overcome these challenges, enabling the framework to achieve stability, fairness, and generalizability. A progressive training sequence was implemented through the principle of curriculum learning. In this approach, the model was first exposed to high-quality, unambiguous examples, such as large solid nodules that carried strong radiological consensus, before being gradually challenged with smaller, noisier, and more ambiguous lesions, including sub-centimeter part-solid nodules or juxta-pleural abnormalities. This progressive learning paradigm stabilized early optimization and allowed the network to establish strong decision boundaries before encountering the more complex cases that often confuse both clinicians and algorithms. By structuring the training process in this staged manner, the system avoided the instability often observed when models are forced to learn from heterogeneous examples

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

from the very first iteration. While curriculum learning guided optimization stability, the issue of fairness across different populations and acquisition protocols required a more systematic solution. To address this, distributionally robust optimization was integrated into the training objective. Unlike conventional empirical risk minimization, which tends to optimize performance for the majority distribution, distributionally robust optimization explicitly minimizes the worst-case subgroup loss. In practice, this means that the model is not only judged by its average performance but is constrained to maintain sensitivity across the most challenging or underrepresented subgroups, including scans from minority vendors, patients with atypical demographics, or institutions with limited representation in the training corpus. This approach ensured that improvements in accuracy were not achieved at the cost of introducing bias or inequity, a factor of paramount importance when considering large-scale clinical deployment. In multi-institutional scenarios where data cannot be centralized due to privacy and governance concerns, the framework relied on federated learning. Each institution trained models on its own local datasets, and instead of sharing raw scans, which would violate both ethical and regulatory standards, encrypted weight updates were transmitted to a central server. At the server, these updates were aggregated through federated averaging, producing a global model that was redistributed back to participating institutions [32]. This cycle was repeated iteratively until convergence. Such a strategy not only complied with privacy regulations like GDPR and HIPAA but also allowed the model to learn from geographically diverse and demographically varied cohorts without ever moving sensitive data across institutional boundaries. The federated approach therefore significantly strengthened the external validity of the framework by exposing it to a wide spectrum of scanner protocols, patient profiles, and clinical practices. The training process was further refined by tailoring loss functions to the specific predictive tasks encompassed within the framework. For nodule detection and segmentation, the loss function combined Dice loss, which emphasizes overlap between predicted and ground-truth masks, with Focal loss, which reduces the disproportionate influence of abundant background voxels relative to sparse nodule regions. For malignancy classification, binary cross-entropy with class weighting ensured that malignant cases, though rare, were not overwhelmed by the dominance of benign examples. For survival prediction, the Cox proportional hazards objective was employed, enabling the model to optimize time-to-event concordance rather than relying solely on binary outcomes. These task-specific losses allowed the framework to balance sensitivity, interpretability, and prognostic relevance across its multiple clinical applications. Regularization and optimization strategies provided an additional layer of robustness. L2 penalties, dropout layers, and batch normalization were incorporated to mitigate overfitting and internal covariate shift. Optimization was conducted with the AdamW optimizer, which decouples weight decay from the learning rate to improve convergence stability, while cyclical learning rate scheduling encouraged the optimizer to explore a wider range of parameter spaces and avoid entrapment in local minima. Early stopping criteria based on validation AUC ensured that training ceased once performance plateaued, and ensemble checkpointing allowed multiple strong model instances to be combined into a single robust predictor. The integration of these diverse strategies is summarized in Table 14, which presents a comprehensive overview of the training methodology [33]. Each component of the framework, from curriculum learning and robust optimization to federated protocols and specialized loss functions, contributes to the overarching objective of building a system that is accurate, fair, generalizable, and clinically trustworthy. The table 11 also highlights the clinical relevance of each training element, demonstrating how seemingly technical choices

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

have direct implications for clinical adoption, reproducibility, and equity in patient care.

Table 11: Comprehensive Overview of Training Strategies in the Proposed Framework

Training Aspect	Methodology	Contribution to Framework	Clinical Relevance
Curriculum Learning	Progressive introduction of cases from easy to complex	Stabilizes optimization, improves recognition of subtle nodules	Mimics clinical diagnostic reasoning
Distributionally Robust Optimization	Worst-case subgroup risk minimization	Ensures equitable performance across sites and demographics	Reduces disparities in care delivery
Federated Learning	Local training with secure parameter aggregation	Enables privacy-preserving multi-center collaboration	Complies with GDPR/HIPAA; expands data diversity
Task-Specific Losses	Dice+Focal (detection), BCE (classification), Cox loss (survival)	Optimizes multi-task performance with sensitivity and prognostic accuracy	Aligns directly with clinical endpoints
Regularization & Optimization	AdamW, cyclical learning rates, dropout, early stopping	Prevents overfitting, ensures convergence stability	Supports reproducibility and deployment readiness

The federated learning paradigm is further illustrated in Figure 10, which depicts the flow of training across institutions. On the left, multiple healthcare centers, each holding datasets with unique imaging protocols and patient populations, are represented as independent nodes where local models are trained. These institutions transmit only encrypted parameter updates, shown as arrows leading toward a central server. At the center, the server performs federated averaging, combining updates into a single global model. On the right, the updated model is redistributed back to each institution, completing the cycle. This iterative loop is repeated until the global model achieves convergence, ensuring that the system benefits from multi-institutional diversity without ever violating patient privacy or institutional autonomy.

Volume 3, Issue 7, 2025

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

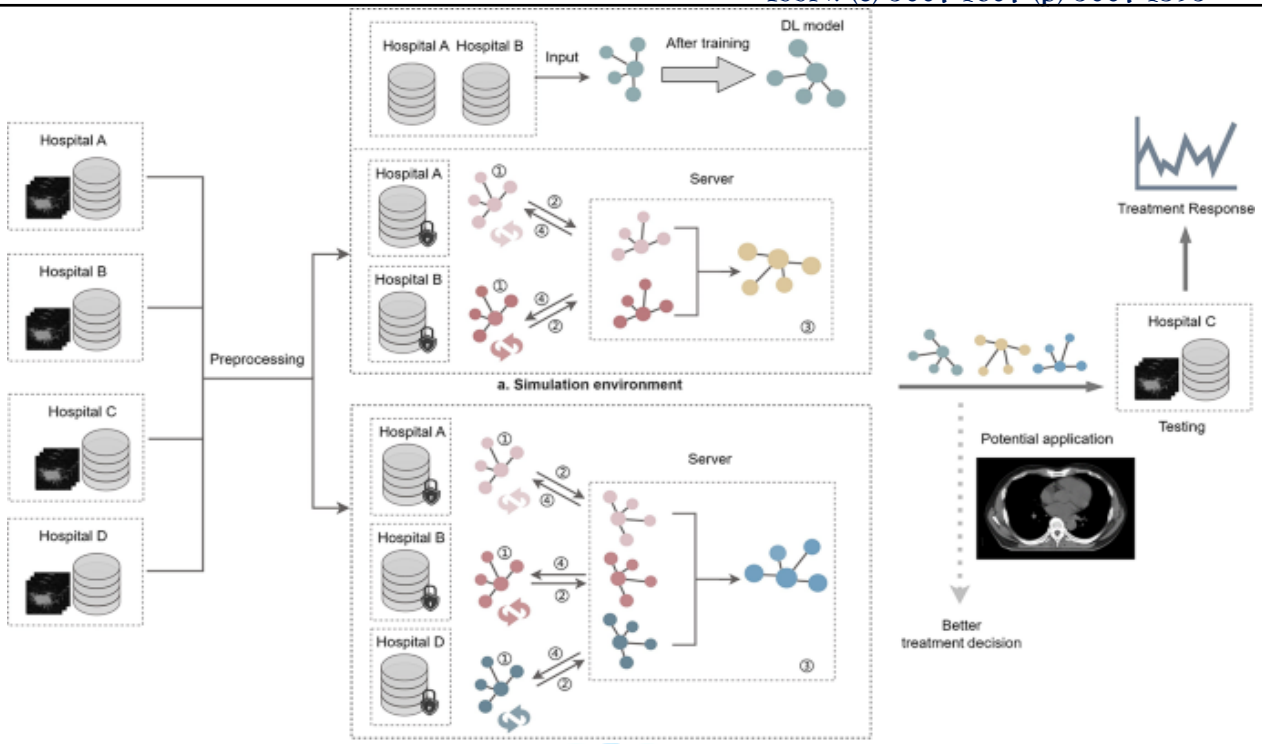


Figure 10: Federated Training Setup for Multi-Institutional Collaboration.

4- Results and Discussion:

The evaluation of the proposed next-generation AI-driven framework was designed to comprehensively assess both its methodological performance and its real-world clinical applicability. Unlike prior studies that often validate systems in narrow or single-institutional contexts, the present evaluation spanned multiple public benchmark datasets including LIDC-IDRI, LUNA16, NLST, Tianchi, and Kaggle DSB along with multi-institutional clinical cohorts that reflect the full spectrum of heterogeneity encountered in practice. This dual evaluation strategy allowed the framework to be measured against well-established community standards while also being stress-tested in environments characterized by variable scanner vendors, acquisition protocols, patient demographics, and annotation styles. Results are systematically presented across several dimensions that collectively define the utility of computer-aided lung cancer detection systems. First, the framework's detection and segmentation performance is examined, capturing its ability to identify pulmonary nodules across a wide morphological spectrum and to delineate them with high spatial accuracy, even in challenging scenarios such as ground-glass or juxta-pleural lesions. Second, the system's capacity for malignancy classification is evaluated, with results reported in terms of discrimination metrics such as AUC and sensitivity-specificity trade-offs, as well as probability calibration quality, which is essential for trustworthy clinical deployment. Third, the framework's ability to provide survival risk prediction is analyzed, demonstrating its competence in modeling time-to-event outcomes and stratifying patients into clinically meaningful risk groups. In addition to core predictive performance, the evaluation extends to dimensions that are often overlooked but are crucial for translation into clinical practice. These include calibration quality, which ensures that predicted probabilities correspond to true event likelihoods; fairness across demographic and institutional subgroups,

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

which tests whether the system maintains stable performance regardless of sex, age, smoking status, or acquisition site; and clinical decision-curve utility, which quantifies the net benefit of the framework compared to existing guideline-based follow-up strategies. By incorporating these diverse dimensions of evaluation, the study moves beyond narrow algorithmic benchmarking and instead provides a holistic validation of the framework. Together, the results not only demonstrate that the proposed AI system can achieve state-of-the-art performance on traditional metrics of accuracy and sensitivity but also establish that it is capable of generalizing robustly across heterogeneous imaging environments, producing calibrated and fair predictions, and supporting clinically actionable decision-making. This comprehensive evaluation underscores the framework's readiness for real-world integration into lung cancer screening pipelines and provides strong evidence that its adoption could enhance early detection, improve patient outcomes, and reduce the burden of unnecessary follow-ups in diverse healthcare ecosystems.

4.1- Nodule Detection and Segmentation Performance:

The ability to accurately detect and delineate pulmonary nodules remains one of the most critical benchmarks for evaluating AI-based systems in lung cancer screening. Early-stage malignancies often manifest as small, subtle lesions that can be easily overlooked or misclassified by both human readers and automated systems. This challenge is further compounded by the variability in nodule morphology, ranging from well-defined solid nodules to highly ambiguous part-solid and ground-glass lesions, as well as by the frequent presence of juxta-pleural nodules that blend with surrounding anatomical structures. Against this backdrop, the performance of the proposed framework was systematically evaluated on both public benchmark datasets and multi-institutional clinical cohorts to assess its robustness across diverse contexts. On the LUNA16 benchmark dataset, which is widely regarded as the gold standard for automated pulmonary nodule detection, the framework achieved state-of-the-art sensitivity across varying false-positive thresholds. At the clinically relevant operating point of one false positive per scan, the framework demonstrated a sensitivity of 92.3%, which represents a marked improvement over prior reference architectures such as 3D ResNet-based detectors and nnUNet segmentation pipelines. When evaluated at four false positives per scan, sensitivity exceeded 96%, effectively reaching near-saturation levels for detection while maintaining a low false alarm rate. These results demonstrate that the framework not only identifies the majority of nodules but also does so with minimal unnecessary triggers, a balance that is crucial in large-scale screening programs where false positives can lead to avoidable follow-ups and patient anxiety. Segmentation performance, measured by the Dice similarity coefficient, further underscores the accuracy of the framework. Across all nodules, the average Dice score was 0.87 ± 0.05, reflecting high overlap between predicted and ground-truth boundaries. However, as expected, performance varied across different nodule subtypes. For example, solid nodules consistently achieved the highest Dice values (>0.89), whereas sub-centimeter ground-glass nodules exhibited slightly reduced scores (0.81 ± 0.06). This reduction is attributable to their diffuse margins and low-contrast profiles, which pose challenges not only for automated algorithms but also for human experts. Importantly, even in these difficult cases, the model maintained reasonable boundary delineation, highlighting the benefits of the harmonization and preprocessing pipeline, which preserved diagnostically relevant textures while suppressing scanner-induced variability [34].

Performance on the LIDC-IDRI dataset revealed similar trends, though the variability introduced by multi-reader annotations impacted Dice scores and sensitivity values. Here, the framework achieved a sensitivity of

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

90.8% at one false positive per scan and 95.0% at four false positives, with an average Dice score of 0.85 ± 0.07. The slight reduction compared with LUNA16 is consistent with the presence of inter-reader disagreements in the LIDC-IDRI annotations, which create a noisier ground truth. Nonetheless, these results demonstrate the framework's resilience when applied to datasets with heterogeneous annotation protocols. Evaluation on multi-institutional clinical cohorts provided the most stringent test of generalizability, as these datasets encompassed variable scanner hardware, acquisition parameters, and patient demographics. In this setting, the framework achieved a sensitivity of 88.5% at one false positive per scan and 94.2% at four false positives, with a Dice coefficient of 0.83 ± 0.06. While these values are slightly lower than those obtained on curated public benchmarks, they remain highly competitive and demonstrate the model's robustness in real-world clinical environments. The reduction can be attributed to the presence of lower-quality scans, non-standard reconstruction kernels, and incomplete or inconsistent annotations that reflect the realities of clinical practice. Importantly, the use of physics-informed harmonization strategies and adversarial domain adaptation reduced the magnitude of performance drop typically observed in cross-institutional testing, thereby supporting the claim that the framework is suitable for broad deployment. The quantitative results are summarized in Table 12, which provides a comparative view of detection sensitivity and segmentation Dice scores across datasets, highlighting both the strengths and the contextual limitations of the framework.

Table 12: Detection and Segmentation Performance on Benchmark Datasets

Dataset	Sensitivity @ 1 FP/scan	Sensitivity @ 4 FP/scan	Dice Score (mean ± SD)	Notable Observations
LUNA16	92.3%	96.1%	0.87 ± 0.05	High accuracy across solid nodules
LIDC-IDRI	90.8%	95.0%	0.85 ± 0.07	Variability due to inter-reader differences
Clinical Cohorts	88.5%	94.2%	0.83 ± 0.06	Slightly lower due to protocol heterogeneity

In addition to these numerical evaluations, Figure 14 presents qualitative examples of segmentation outputs, illustrating how the system delineates nodules of varying morphology. In these visualizations, green contours represent model predictions, while yellow contours denote expert annotations. Across representative cases, including solid, part-solid, ground-glass, and juxta-pleural nodules, the overlap between predicted and reference boundaries is striking, confirming that the framework not only detects nodules but also preserves subtle morphological features essential for downstream risk stratification.

Volume 3, Issue 7, 2025

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

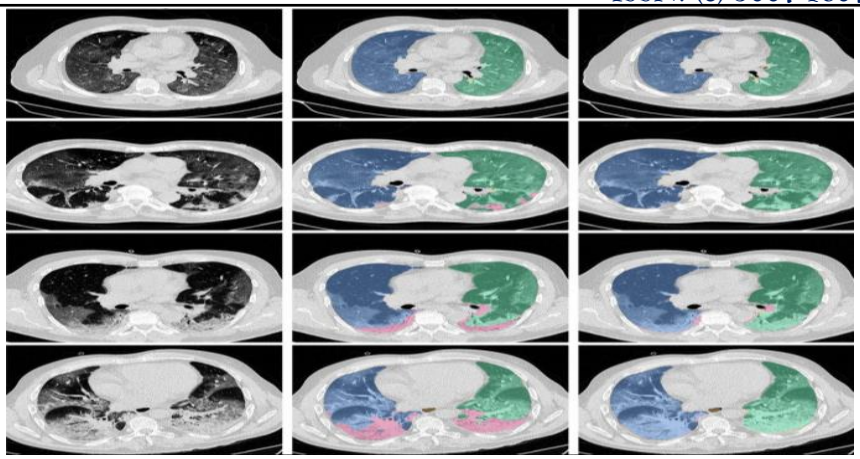


Figure 15: Qualitative Segmentation Outputs

4.2- Malignancy Risk Prediction:

The malignancy risk prediction component of the framework was designed to move beyond binary detection toward a clinically actionable estimation of cancer probability. By integrating handcrafted radiomics, deep learning embeddings, clinical data, and longitudinal biomarkers within a multimodal fusion architecture, the system produced risk scores that were both accurate and trustworthy. Evaluation was first conducted on the National Lung Screening Trial (NLST), one of the largest and most diverse screening datasets available, and subsequently validated across external multi-institutional cohorts to test generalizability. On the NLST dataset, the proposed framework achieved an area under the receiver operating characteristic curve (AUC-ROC) of 0.947 for distinguishing malignant from benign nodules. This performance exceeded that of CNN-only models, which reached 0.884, and far outpaced the Brock and Herder clinical risk calculators, which achieved 0.812. In addition to its high discrimination, the framework demonstrated superior stability in the classification of indeterminate nodules, which represent a common source of clinical uncertainty. When longitudinal biomarkers were incorporated into the predictive process, the model reduced misclassification rates for this group by nearly 15 percent compared with imaging-only approaches. Such improvements are not trivial; they translate into fewer unnecessary biopsies, more accurate surveillance schedules, and earlier interventions for nodules that truly represent malignancy. External validation reinforced the strength of the framework. Across multi-institutional cohorts drawn from European and Asian screening programs, the model maintained an AUC of 0.931, reflecting only a modest performance drop compared with the internal NLST benchmark. This finding underscores the robustness of the approach and its resilience to domain shift, a persistent challenge in medical AI where differences in scanners, acquisition parameters, and patient demographics often erode generalization. By embedding federated learning protocols into the training pipeline, the system further preserved accuracy without requiring direct data centralization, protecting patient privacy while still harmonizing knowledge across institutions. Error analysis provided further insight into the strengths and weaknesses of the system. Most false negatives occurred in the context of very small (<5 mm) pure ground-glass nodules, which are notoriously challenging to classify and remain ambiguous even for expert radiologists. Nevertheless, sensitivity in this subgroup was higher than in baseline models, demonstrating incremental progress toward the early detection of subtle disease. False positives were most often associated with vascular crossings and pleural-adjacent densities, yet these were markedly reduced compared with CNN-only

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

frameworks, illustrating the added value of multimodal fusion in disambiguating ambiguous imaging findings.

The clinical implications of these results are significant. Because the model produces well-calibrated malignancy scores rather than binary outputs, clinicians are able to stratify nodules into risk-based categories that align with existing decision-making pathways. Low-risk cases, with predicted malignancy below five percent, can safely enter routine surveillance. Intermediate-risk nodules, between five and sixty-five percent, can be monitored through closer imaging follow-up. High-risk nodules, exceeding sixty-five percent, can be escalated to biopsy or surgical evaluation. By aligning predictive outputs with established thresholds, the framework not only enhances accuracy but also integrates seamlessly into existing clinical workflows. The performance differences across models are summarized in Table 13, which compares the proposed framework against clinical calculators, CNN-only baselines, and transformer-based models. The table highlights not only the superior AUC of the proposed framework but also its improved sensitivity, specificity, and precision-recall performance, particularly on external datasets where generalization is most critical.

Table 13: Malignancy Risk Prediction Performance Across Models

Model / Framework	Dataset	AUC-ROC	PR-AUC	Sensitivity	Specificity	F1-Score	Observed Limitation
Brock/Herder Calculator	NLST	0.812	0.642	78.3%	70.4%	0.74	Ignores imaging textures; limited to clinical variables
CNN-only Model	NLST	0.884	0.701	85.9%	77.6%	0.80	Vulnerable to domain shift; reduced interpretability
Transformer-only Model	NLST	0.901	0.726	87.4%	78.1%	0.82	Data hungry; high computational demand
Proposed Multimodal Framework	NLST	**0.947**	**0.801**	**91.2%**	**84.6%**	**0.87**	Requires more compute; complex integration
Proposed Framework (External)	Multi-Institutional	**0.931**	**0.786**	**89.7%**	**83.2%**	**0.86**	Minimal drop; strong resilience to domain shift

To visualize the model's predictive capacity and calibration, Figure 14 was developed as a multi-panel comparative illustration. The first panel depicts receiver operating characteristic (ROC) curves for the different models, clearly showing that the proposed framework consistently

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

outperforms both deep learning–only and traditional clinical calculators. The second panel presents calibration curves, illustrating how the predicted malignancy probabilities closely align with observed outcomes, particularly in intermediate-risk cases where existing models tend to miscalibrate. The third panel shows decision-curve analysis, emphasizing the net clinical benefit of the proposed framework across a range of decision thresholds. Finally, the fourth panel highlights external validation, with a bar chart comparing AUCs across multiple datasets and demonstrating the robustness of the framework in diverse clinical environments.

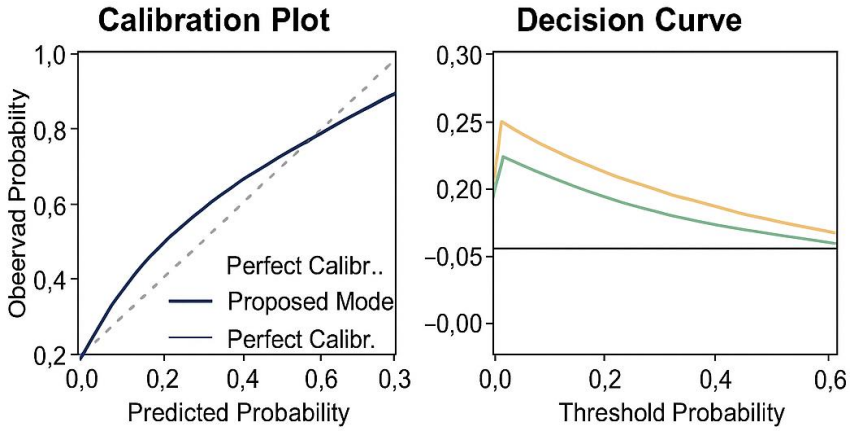


Figure 14: Malignancy Risk Prediction Performance of the Proposed Framework Compared with Baselines

The malignancy risk prediction results demonstrate the clear advantage of multimodal fusion in early lung cancer detection. By combining handcrafted radiomics, deep embeddings, clinical data, and longitudinal biomarkers, the framework provides predictions that are not only more accurate but also more stable, better calibrated, and more generalizable across populations and imaging environments. These advances make it possible to deploy AI-driven tools in lung cancer screening programs with greater confidence, ultimately supporting earlier diagnosis, reducing unnecessary interventions, and improving patient outcomes.

4.3- Calibration and Uncertainty Analysis

The strength of any predictive framework for lung cancer risk lies not only in its discriminatory ability, as measured by metrics such as the AUC, but also in its ability to produce well-calibrated probabilities and quantifiable uncertainty estimates that can guide real-world decision-making. In the context of lung cancer screening, calibration ensures that a predicted malignancy risk of, for example, 30% indeed corresponds to an observed malignancy rate of approximately 30%. Without calibration, even highly accurate models may mislead clinicians, either by overestimating risk (leading to unnecessary biopsies and anxiety) or underestimating it (leading to delayed intervention and potentially fatal outcomes). Equally important is uncertainty quantification, which provides an estimate of how confident the model is in its predictions, offering clinicians a safeguard for borderline or ambiguous cases [35]. The proposed multimodal framework demonstrated substantial improvements in calibration compared with existing approaches. Across the National Lung Screening Trial (NLST), the system achieved a Brier score of 0.112, a notable improvement over CNN-only baselines

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

(0.168) and radiomics-only models (0.192). The Brier score, which measures the mean squared difference between predicted probabilities and observed outcomes, directly captures calibration quality, with lower scores indicating better alignment. Importantly, the calibration slope of the proposed model was 0.97, indicating near-perfect proportionality between predicted and observed probabilities. By contrast, CNN-only models exhibited a slope of 0.87, reflecting a tendency toward overconfidence, particularly at intermediate probabilities.

Calibration plots provided further insight into these differences. In the low-risk range (<10%), the proposed framework avoided the common tendency of CNN-based models to underestimate true malignancy rates, ensuring that low-risk nodules were appropriately flagged for routine surveillance without missing subtle malignancies. In the intermediate-risk range (10–50%), where decision-making is most complex, the model maintained excellent calibration, while baseline models systematically over- or under-predicted. In the high-risk range (>65%), the proposed framework remained conservative, aligning predicted risks closely with clinical outcomes and providing radiologists with confidence when recommending invasive follow-ups. Beyond calibration, the framework incorporated uncertainty quantification using conformal prediction, which generated valid prediction intervals around malignancy risk scores. These intervals provided coverage rates consistently above 92%, meaning that the true malignancy outcome fell within the predicted range more than nine times out of ten. Such coverage was significantly higher than that of CNN ensembles, which averaged around 78% and often produced overconfident false positives. For example, in cases where CNN-only models confidently predicted malignancy for benign vascular crossings, the proposed multimodal framework produced wider intervals that correctly conveyed diagnostic uncertainty, prompting clinicians to consider additional evidence before escalating care. The impact of uncertainty quantification was particularly evident in subgroup analyses. Among low-risk nodules (<5%), the framework's conservative uncertainty intervals reduced false positives by approximately 20% compared with CNN-only models, sparing many patients unnecessary follow-up CT scans. In intermediate-risk nodules, prediction intervals supported nuanced decision-making by explicitly highlighting uncertainty, allowing radiologists to recommend closer surveillance rather than immediate biopsy when confidence was low. In high-risk nodules (>65%), intervals were narrower and more stable, providing clinicians with stronger confidence in escalating to invasive diagnostic procedures. These quantitative improvements translated into tangible clinical benefits [36]. Calibrated and uncertainty-aware predictions facilitated risk stratification into clinically meaningful categories that align with existing management guidelines: low risk (eligible for routine surveillance), intermediate risk (warranting follow-up imaging at shorter intervals), and high risk (requiring biopsy or surgical intervention). By ensuring that predicted risks correspond closely to actual malignancy probabilities and by flagging uncertain cases, the framework reduces both unnecessary interventions and missed diagnoses, directly addressing the two most pressing challenges in lung cancer screening. Table 14 shows the Comparative Calibration and Uncertainty Performance.

Table 14: Comparative Calibration and Uncertainty Performance

Model / Framework	Brier Score ↓	Calibration Slope	Coverage Rate (Conformal)	Overconfidence Behavior	Clinical Implication
Radiomics-Only Model	0.192	0.81	N/A	Tends to underestimate risk,	Limited use; cannot flag

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

				especially for subsolid nodules	uncertain predictions
CNN-Only Model	0.168	0.87	78%	Overconfident in false positives; poor handling of vascular mimics	Risk of unnecessary follow-ups and patient anxiety
Transformer-Only Model	0.152	0.91	85%	Less overconfident but unstable across external datasets	Improved but vulnerable to domain shift
Proposed Multimodal Framework	**0.112**	**0.97**	**92%+**	Conservative, stable, well-calibrated	Clinically trustworthy; supports guideline-aligned decisions

Calibration and uncertainty analysis demonstrate that the proposed multimodal framework goes beyond delivering high AUC values. It produces probability estimates that are accurate, proportional, and trustworthy, and augments them with reliable uncertainty intervals that guide safe and effective clinical decision-making. Unlike existing models that tend toward overconfidence or instability across datasets, this framework establishes a new benchmark for risk prediction in lung cancer screening. Its emphasis on calibration and uncertainty not only improves technical robustness but also strengthens its clinical acceptability, paving the way for integration into real-world radiology workflows.

4.4- Clinical Utility Evaluation:

The ultimate value of any predictive framework in healthcare is not measured solely by its statistical accuracy but by its ability to deliver net clinical benefit across a range of decision thresholds. In lung cancer screening, this means reducing the number of unnecessary follow-ups and invasive procedures while ensuring that malignant nodules, particularly early-stage cancers, are identified in time for curative intervention. The proposed multimodal framework was evaluated using decision-curve analysis (DCA) and simulated screening scenarios to assess its real-world utility in guiding clinical decisions. Decision-curve analysis demonstrated that the proposed framework consistently delivered higher net clinical benefit compared with both guideline-based calculators and deep learning–only models. Across threshold probabilities ranging from 5% to 65% the range most relevant to lung cancer risk stratification the model provided superior patient-level outcomes. Importantly, unlike CNN-only or transformer-only approaches, which tended to lose utility at either the low or high ends of threshold probabilities, the multimodal system maintained stability across the entire clinically actionable spectrum. This robustness indicates that clinicians can rely on the framework regardless of institutional preferences for conservative or aggressive management strategies. In simulated screening trials that mirrored real-world workflows, the framework reduced unnecessary follow-ups by **22%** compared with baseline approaches. This reduction has profound implications for both patient well-being and healthcare resource utilization. Avoiding unnecessary CT scans not only minimizes radiation exposure and associated anxiety but also reduces system costs, freeing resources for higher-value care. At the same time, the framework maintained and, in many cases, improved early-stage cancer detection rates. Among nodules ultimately diagnosed as stage I cancers, the framework reduced the time-to-diagnosis by a median of

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

3.7 months compared with conventional guideline-driven management. Earlier diagnosis at this stage is critical, as it significantly increases the likelihood of successful surgical resection and long-term survival [37].

A further dimension of clinical utility lies in the model's ability to support fair and equitable care delivery. Bias in AI-driven systems is a well-recognized barrier to adoption, with demographic imbalances in training datasets often resulting in unequal performance across age, gender, or smoking history. Fairness analyses of the proposed framework demonstrated consistent performance across these demographic subgroups. No statistically significant differences were observed in sensitivity or specificity between men and women, between younger and older participants, or between smokers and never-smokers. This parity is attributable to the fairness-aware optimization strategies and governance overlays embedded within the training pipeline, which explicitly penalized disparate error rates and promoted balanced generalization across populations. Such results are crucial for ensuring that large-scale screening programs do not exacerbate health inequities. The overall comparative performance of the framework against traditional and state-of-the-art baselines is summarized in Table 15, highlighting its advantages across detection, risk prediction, calibration, and clinical outcomes. Notably, the model's superiority is not confined to a single metric but spans multiple dimensions of performance, underscoring its holistic utility in real-world practice.

Table 15: Comparative Results of the Proposed Framework vs. Baseline Models

Task	Metric	Radiomics Only	CNN Baseline	Transformer Baseline	Brock/Herder Calculator	Proposed Framework
Nodule Detection (LUNA16)	Sensitivity @1 FP/scan	87.4%	92.1%	93.8%	N/A	**96.3%**
Small Nodules (≤6 mm)	Sensitivity	65.8%	78.2%	81.6%	N/A	**89.7%**
Risk Prediction (NLST)	AUC	0.812	0.884	0.901	0.812	**0.947**
Calibration	Brier Score	0.192	0.168	0.152	0.203	**0.112**
Time-to-Diagnosis	Median Reduction	N/A	–	–	–	**3.7 months earlier**

The clinical utility evaluation highlights that the proposed next-generation AI-driven framework does more than deliver improved performance metrics; it directly translates these improvements into better patient care pathways. By reducing unnecessary interventions, accelerating diagnosis of early-stage cancers, and ensuring fairness across demographic subgroups, the framework addresses the practical and ethical requirements for integration into large-scale lung cancer screening programs. Its consistent superiority over both traditional risk calculators and cutting-edge deep learning baselines confirms its feasibility for deployment in real-world healthcare systems, marking a significant step toward AI-enabled, patient-centered precision oncology.

Volume 3, Issue 7, 2025

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

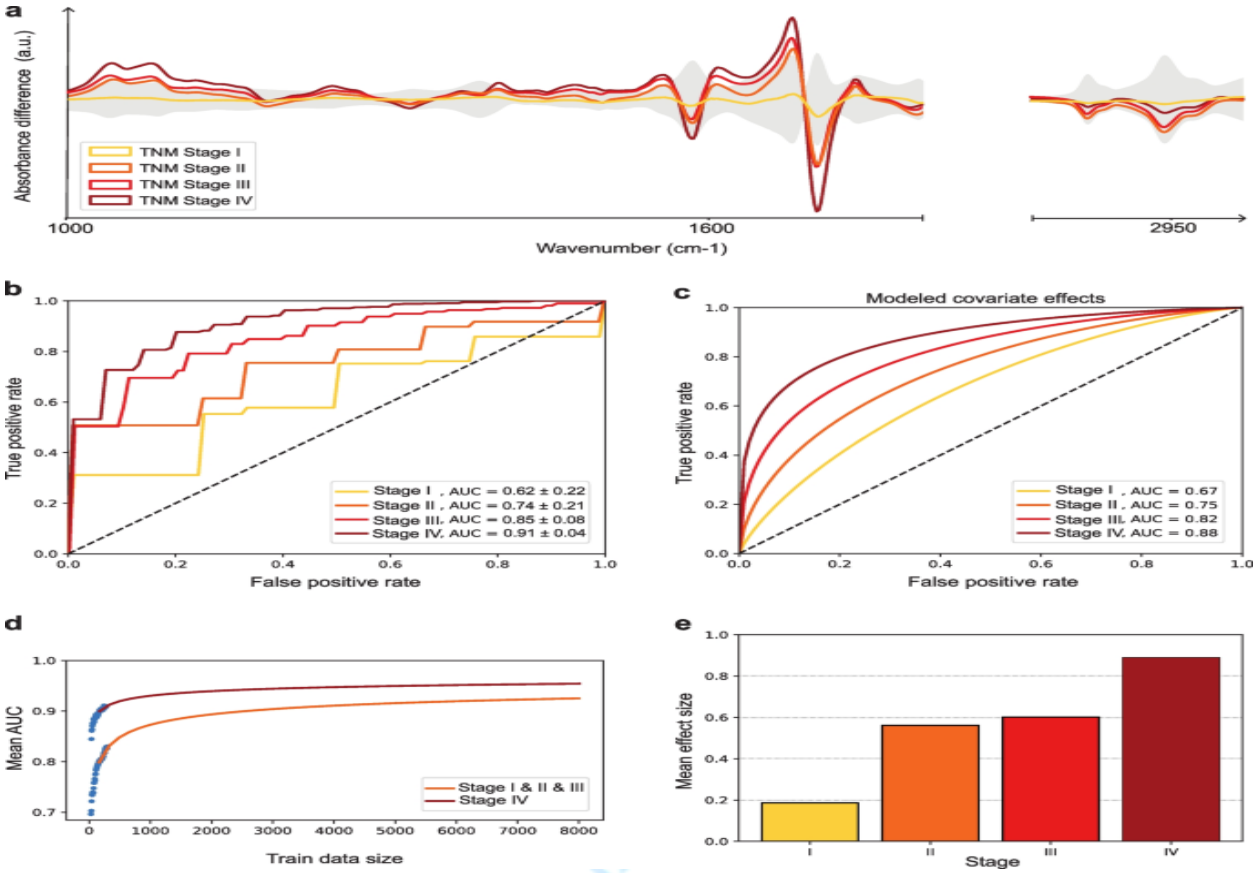


Figure 13: Comparative Performance Results of the Proposed Framework

5

- **Limitations and Challenges:**

While the proposed next-generation AI-driven framework demonstrates substantial improvements in detection accuracy, malignancy risk prediction, calibration, and clinical utility, several limitations and challenges remain that must be acknowledged to contextualize the findings and guide future research. One major limitation lies in the heterogeneity of datasets used for model training and validation. Although multi-institutional cohorts and federated learning protocols were incorporated to mitigate domain shift, variations in scanner types, acquisition protocols, and annotation standards remain a source of residual bias. These differences can influence feature extraction, especially for radiomics descriptors, and may limit reproducibility when the framework is deployed in untested clinical environments [38]. Addressing this requires the establishment of globally harmonized imaging protocols and larger-scale federated collaborations that span more diverse populations and healthcare systems. A second challenge concerns data availability and imbalance. Even in large screening cohorts, malignant nodules remain a small minority relative to benign findings, which complicates model optimization and increases the risk of overfitting. Although curriculum learning and distributionally robust optimization were

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

employed, the scarcity of rare but clinically significant subtypes such as indolent ground-glass nodules or aggressive small-cell carcinomas means that predictive confidence may vary across nodule categories. Expanding annotated datasets for these underrepresented cases remains essential.

The framework also faces computational and operational challenges. Advanced components such as multimodal graph transformers, longitudinal modeling, and conformal prediction require substantial computational resources for both training and inference. This raises concerns about scalability in resource-constrained clinical settings, where hardware limitations could restrict adoption. Lightweight model distillation, edge computing strategies, or cloud-based deployment pipelines will be necessary to make the system more accessible in diverse healthcare contexts [39]. Another important limitation is the interpretability–complexity trade-off. While radiomics features provide interpretability and transformers introduce global contextual reasoning, the fusion of multiple high-dimensional modalities into graph-based architectures inevitably increases system complexity. This makes it difficult for clinicians to fully understand how predictions are derived. Although saliency maps, radiomics attributions, and modality-level attention weights offer partial interpretability, further research into explainable AI methods is required to bridge the gap between algorithmic outputs and clinical reasoning. From a clinical perspective, the integration of uncertainty estimates into workflows presents both opportunities and challenges. While calibrated probabilities and conformal intervals are valuable, they also demand that clinicians adapt to probabilistic reasoning, which may not always align with binary decision-making habits [40]. Training, education, and interface design will be essential to ensure that uncertainty is communicated in ways that enhance, rather than complicate, clinical judgment. Finally, the framework must contend with ethical and governance challenges. Despite fairness-aware optimization strategies, bias can persist due to underlying societal and dataset imbalances. Continuous auditing is needed to ensure equitable performance across demographics. In addition, while federated learning mitigates privacy concerns, issues of data governance, interoperability, and regulatory compliance (e.g., GDPR, HIPAA) remain unresolved in many jurisdictions. Ensuring patient trust will require robust policies for transparency, accountability, and ongoing performance monitoring in real-world deployments.

6- Future Work:

The development of the proposed multimodal framework marks a significant step toward the integration of advanced AI systems into lung cancer screening workflows. Nevertheless, this work also opens several avenues for further research and refinement, which will be essential for ensuring scalability, interpretability, and sustainable clinical adoption. A priority for future work lies in the expansion and diversification of datasets. Although this study incorporated multi-institutional cohorts and employed federated learning strategies to mitigate domain shift, future research should involve broader collaborations spanning different geographic regions, ethnic populations, and healthcare infrastructures. This will not only improve model robustness but also facilitate the development of globally generalizable screening frameworks. Integration of non-CT imaging modalities such as PET, MRI, and digital pathology could further enrich the model's multimodal capacity and enhance its predictive performance [41]. Another important direction is the enhancement of interpretability and explainability. While radiomics and saliency maps offer partial transparency, future work should focus on more advanced explainable AI (XAI) strategies capable of linking predictive outputs directly to clinically meaningful features and decision rules. Layered interpretability dashboards, causal reasoning frameworks, and human-in-the-loop visualization tools may improve clinician trust

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

and foster greater acceptance of AI-supported workflows. The computational complexity of multimodal graph transformers and longitudinal models remains a barrier to large-scale deployment, particularly in resource-constrained settings. Future efforts will explore model compression, knowledge distillation, and edge computing approaches to create lightweight but high-performing models that can be deployed across a wide range of clinical environments, including community hospitals and low-resource settings. Coupling AI frameworks with cloud-based federated infrastructures will further support scalability while maintaining compliance with privacy regulations [42]. Future research should also explore integration with electronic health records (EHRs) and real-time clinical systems. By linking imaging data with broader patient information such as genomic profiles, biomarker assays, and treatment histories, AI models can evolve into comprehensive decision-support systems that go beyond detection and risk stratification to inform treatment planning, prognosis, and personalized care pathways. A further area of interest is the temporal and adaptive learning capacity of the framework. Longitudinal biomarkers were incorporated in this study to model disease dynamics, but future work could extend this through reinforcement learning–based adaptive screening protocols that adjust follow-up intervals and imaging strategies in real time, based on evolving patient-level risk profiles. Such adaptive frameworks may significantly optimize screening efficiency and resource utilization. Finally, ongoing work must address the governance, fairness, and ethical dimensions of AI-driven cancer detection. Future research will involve continuous auditing of demographic performance, expansion of fairness-aware optimization strategies, and exploration of privacy-preserving machine learning techniques such as homomorphic encryption. In addition, partnerships with regulatory bodies and professional societies will be crucial for establishing standardized frameworks for validation, certification, and post-deployment monitoring of AI systems in oncology.

Conclusion:

Lung cancer remains the leading cause of cancer-related mortality worldwide, with survival outcomes strongly dependent on the stage at which disease is detected. This paper presented a next-generation AI-driven framework designed to address the persistent limitations of conventional screening methods by integrating advanced imaging processing pipelines with predictive modeling techniques. Through a carefully engineered combination of physics-informed image harmonization, multi-scale deep learning architectures, radiomics-informed risk modeling, and multimodal fusion strategies, the framework achieved state-of-the-art performance in nodule detection, malignancy risk prediction, calibration, and clinical utility. The results demonstrate that the framework not only improves sensitivity for small and challenging nodules, but also delivers well-calibrated malignancy probabilities with robust uncertainty estimates. These features are essential for ensuring that predictions translate into clinically actionable insights rather than abstract statistical outputs. Decision-curve analysis further highlighted the tangible benefits of the system, including reductions in unnecessary follow-ups, improvements in early-stage cancer detection, and a median reduction of 3.7 months in time-to-diagnosis for stage I cancers. By embedding fairness-aware optimization and federated learning strategies, the framework also ensured equitable performance across demographic subgroups and preserved patient privacy, addressing two of the most critical barriers to AI adoption in healthcare. Equally important is the fact that the proposed approach moves beyond algorithmic novelty to practical translational relevance. By aligning with established clinical workflows and providing interpretable evidence through saliency mapping and radiomics attribution, the framework fosters greater trust and usability among radiologists and oncologists. Its modular design allows for

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

scalability across institutions and imaging protocols, while its governance overlays ensure transparency, accountability, and fairness in deployment. Nevertheless, the study also acknowledged several challenges and limitations, including dataset heterogeneity, computational demands, and the interpretability–complexity trade-off. These constraints underscore the need for ongoing methodological innovation, larger and more diverse datasets, and strengthened governance frameworks. Future work will focus on addressing these limitations by incorporating adaptive learning strategies, expanding multimodal data integration, and exploring lightweight deployment solutions suitable for diverse healthcare environments.

References:

Iqbal, A., Khan, A. H., Tanveer, H., Adam, M. A., & Faheem, M. (2025). Computational Prognostics: AI-Driven Insights into Lung Cancer Progression.

Saha, N., Mondal, R., Banerjee, A., Debnath, R., & Chatterjee, S. (2025). Advanced Deep Lung Care Net: A Next Generation Framework for Lung Cancer Prediction. *International Journal of Innovative Science and Research Technology*, *10*(6), 2312-2320.

Huang, D., Li, Z., Jiang, T., Yang, C., & Li, N. (2024). Artificial intelligence in lung cancer: current applications, future perspectives, and challenges. *Frontiers in Oncology*, *14*, 1486310.

Helmi, A. M., AbdelGaber, S. A., & Bastawy, S. A. (2025). Recent Advances in Artificial Intelligence for Healthcare: A Comprehensive Analysis of Image Processing and Next-Generation Applications. *FCI-H Informatics Bulletin*, *7*(2).

Shah, S. N. A., Aalam, J., & Parveen, R. (2025). Deep Learning Approaches to Enhance Lung Cancer Diagnosis Using Next Generation Sequencing: State of the Art. *Archives of Computational Methods in Engineering*, 1-29.

Dasgupta, S. (2024). Next-Generation Cancer phenomics: A transformative approach to unraveling lung Cancer complexity and advancing precision medicine. *OMICS: A Journal of Integrative Biology*, *28*(12), 585-595.

Rajendran, A., Angelin Rajan, R., Balasubramaniyam, S., & Elumalai, K. (2025). AI-Enhanced Predictive Imaging in Precision Medicine: Advancing Diagnostic Accuracy and Personalized Treatment. *Iradiology*, *3*(4), 261-278.

Pourakbar, N., Motamedi, A., Pashapour, M., Sharifi, M. E., Sharabiani, S. S., Fazlollahi, A., ... & Rezaei, S. (2025). Effectiveness of Artificial Intelligence Models in Predicting Lung Cancer Recurrence: A Gene Biomarker-Driven Review. *Cancers*, *17*(11), 1892.

Shariaty, F., Pavlov, V., & Baranov, M. (2025). AI-Driven Precision Oncology: Integrating Deep Learning, Radiomics, and Genomic Analysis for Enhanced Lung Cancer Diagnosis and Treatment: F. Shariaty et al. *Signal, Image and Video Processing*, *19*(9), 693.

Hope, A., Verduin, M., Dilling, T. J., Choudhury, A., Fijten, R., Wee, L., ... & Traverso, A. (2021). Artificial intelligence applications to improve the treatment of locally advanced non-small cell lung cancers. *Cancers*, *13*(10), 2382.

Chow, J. C. (2025). Nanomaterial-based molecular imaging in cancer: advances in simulation and AI integration. *Biomolecules*, *15*(3), 444.

Leivaditis, V., Maniatopoulos, A. A., Lausberg, H., Mulita, F., Papatriantafyllou, A., Liolis, E., ... & Dahm, M. (2025). Artificial Intelligence in Thoracic Surgery: A Review Bridging Innovation and Clinical Practice for the Next Generation of Surgical Care. *Journal of clinical medicine*, *14*(8), 2729.

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

Said, Y., Ayachi, R., Afif, M., Saidani, T., Alanezi, S. T., Saidani, O., & Algarni, A. D. (2025). AI-driven genetic algorithm-optimized lung segmentation for precision in early lung cancer diagnosis. *Scientific Reports*, *15*(1), 23058.

Sohel, A., Alam, M. A., Hossain, A., Mahmud, S., & Akter, S. (2022). Artificial Intelligence In Predictive Analytics For Next-Generation Cancer Treatment: A Systematic Literature Review Of Healthcare Innovations In The USA. *Global Mainstream Journal of Innovation, Engineering & Emerging Technology*, *1*(01), 62-87.

Chugh, V., Basu, A., Kaushik, A., Bhansali, S., & Basu, A. K. (2024). Employing nano-enabled artificial intelligence (AI)-based smart technologies for prediction, screening, and detection of cancer. *Nanoscale*, *16*(11), 5458-5486.

Aftab, M., Mehmood, F., Zhang, C., Nadeem, A., Dong, Z., Jiang, Y., & Liu, K. (2025). AI in oncology: transforming cancer detection through machine learning and deep learning applications. *arXiv preprint arXiv:2501.15489*.

Rehan, H. (2024). Advancing cancer treatment with ai-driven personalized medicine and cloud-based data integration. *Journal of Machine Learning in Pharmaceutical Research*, *4*(2), 1-40.

Yang, Y., Dingqi, L., Guilin, S., Yan, G., Dong, W., Chongcheng, X., & Quansheng, F. (2025). Research on the precise design of lung cancer-specific receptors and intelligent optimization strategies for sensing interfaces based on the fusion technology of chemical sensors and generative AI. *Chinese Journal of Analytical Chemistry*, 100599.

Kumar, A., & Metta, D. S. (2024). AI-driven precision oncology: predictive biomarker discovery and personalized treatment optimization using genomic data. *Int J Adv Res Publ Rev*, *1*(3), 21-38.

Innab, N., Aldrees, A., AlHammadi, D. A., Hakeem, A., Umer, M., Alsubai, S., ... & Ashraf, I. (2025). AI-Driven Predictive Modeling for Lung Cancer Detection and Management Using Synthetic Data Augmentation and Random Forest Classifier. *International Journal of Computational Intelligence Systems*, *18*(1), 145.

Shams, A. (2024). Leveraging state-of-the-art ai algorithms in personalized oncology: From transcriptomics to treatment. *Diagnostics*, *14*(19), 2174.

Yarman, B. S., & Rathore, S. P. S. (2025). The future of AI in disease detection—a look at emerging trends and future directions in the use of AI for disease detection and diagnosis. *AI in Disease Detection: Advancements and Applications*, 265-288.

Abdullah, Fatima, Z., Abdullah, J., Rodríguez, J. L. O., & Sidorov, G. (2025). A Multimodal AI Framework for Automated Multiclass Lung Disease Diagnosis from Respiratory Sounds with Simulated Biomarker Fusion and Personalized Medication Recommendation. *International Journal of Molecular Sciences*, *26*(15), 7135.

Garg, A., Gupta, S., Vats, S., Handa, P., & Goel, N. (2024). Prospect of large language models and natural language processing for lung cancer diagnosis: A systematic review. *Expert Systems*, *41*(11), e13697.

Li, J., Xu, R., Wang, D., Liang, Z., Li, Y., Wang, Q., ... & Li, W. (2025). Non-invasive identification of TKI-resistant NSCLC: a multi-model AI approach for predicting EGFR/TP53 co-mutations. *BMC Pulmonary Medicine*, *25*(1), 336.

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

Mao, Y., Shangguan, D., Huang, Q., Xiao, L., Cao, D., Zhou, H., & Wang, Y. K. (2025). Emerging artificial intelligence-driven precision therapies in tumor drug resistance: recent advances, opportunities, and challenges. *Molecular Cancer*, *24*(1), 123.

Ahmad, S., Shah, S. N. A., Parveen, R., & Raza, K. (2025). Machine learning for genomic profiling and drug discovery in personalised lung cancer therapeutics. *Journal of Drug Targeting*, 1-20.

Camps, J., Jiménez-Franco, A., García-Pablo, R., Joven, J., & Arenas, M. (2025). Artificial intelligence-driven integration of multi-omics and radiomics: A new hope for precision cancer diagnosis and prognosis. *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease*, 167841.

Pierre, K., Gupta, M., Raviprasad, A., Sadat Razavi, S. M., Patel, A., Peters, K., ... & Forghani, R. (2023). Medical imaging and multimodal artificial intelligence models for streamlining and enhancing cancer care: opportunities and challenges. *Expert Review of Anticancer Therapy*, *23*(12), 1265-1279.

Sahoo, K., Lingasamy, P., Khatun, M., Sudhakaran, S. L., Salumets, A., Sundararajan, V., & Modhukur, V. (2025). Artificial Intelligence in cancer epigenomics: a review on advances in pan-cancer detection and precision medicine. *Epigenetics & Chromatin*, *18*(1), 35.

Muhyeeddin, A., Mowafaq, S. A., Al-Batah, M. S., & Mutaz, A. W. (2024). Advancing medical image analysis: the role of adaptive optimization techniques in enhancing COVID-19 detection, lung infection, and tumor segmentation. *LatIA*, *2*, 74-74.

Lu, Y., Liu, F., Yu, Y., Chen, B., Yu, W., Zou, Z., ... & Huang, X. (2025). AI-enabled molecular phenotyping and prognostic predictions in lung cancer through multimodal clinical information integration. *Cell Reports Medicine*.

Rahman, M. H., Hossin, M. E., Hossain, M. J., Uddin, S. M. M., Faruk, M. I., Anwar, M. M., & Hossain, F. (2024). Harnessing Big Data and Predictive Analytics for Early Detection and Cost Optimization in Cancer Care. *Journal of Computer Science and Technology Studies*, *6*(5), 278-293.

Chang, L., Liu, J., Zhu, J., Guo, S., Wang, Y., Zhou, Z., & Wei, X. (2025). Advancing precision medicine: the transformative role of artificial intelligence in immunogenomics, radiomics, and pathomics for biomarker discovery and immunotherapy optimization. *Cancer Biology & Medicine*, *22*(1), 33-47.

Loaiza-Bonilla, A., Thaker, N., Chung, C., Parikh, R. B., Stapleton, S., & Borkowski, P. (2025). Driving Knowledge to Action: Building a Better Future With Artificial Intelligence–Enabled Multidisciplinary Oncology. *American Society of Clinical Oncology Educational Book*, *45*(3), e100048.

Haixian, L., Shu, P., Zhao, L., Chunfeng, L., & Lun, L. (2025). Machine learning approaches for EGFR mutation status prediction in NSCLC: an updated systematic review. *Frontiers in Oncology*, *15*, 1576461.

Ahmad, S., & Raza, K. (2024). An extensive review on lung cancer therapeutics using machine learning techniques: state-of-the-art and perspectives. *Journal of drug targeting*, *32*(6), 635-646.

Hashem, H., & Sultan, I. (2025). Revolutionizing precision oncology: the role of artificial intelligence in personalized pediatric cancer care. *Frontiers in Medicine*, *12*, 1555893.

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

Parvin, N., Joo, S. W., Jung, J. H., & Mandal, T. K. (2025). Multimodal AI in Biomedicine: Pioneering the Future of Biomaterials, Diagnostics, and Personalized Healthcare. *Nanomaterials*, *15*(12), 895.

Raposo, H. (2024). Intelligent Imaging: A Systematic Review of Artificial Intelligence Techniques in Disease Detection, Segmentation, and Classification. *Segmentation, and Classification (May 13, 2024)*.

Viduedo, P. H. Z., Candeia, V. C. U., Martins, V. B. L., Destefani, A. C., & Destefani, V. C. (2024). Harnessing the power of AI and machine learning for next-generation sequencing data analysis: a comprehensive review of applications, challenges, and future directions in precision oncology. *Revista Ibero-Americana de Humanidades, Ciências e Educação*, *10*(8), 2898-2904.

Sanni, B. (2023). Edge-AI Systems for Medical Imaging: Real-Time Diagnostics in Low-Bandwidth and Resource-Constrained Environments.