GMR
The Original

ARTIFICIAL INTELLIGENCE-DRIVEN PULMONARY NODULE IDENTIFICATION VIA COMPUTED TOMOGRAPHY: TECHNOLOGICAL PROGRESS, UNRESOLVED HURDLES, AND EMERGING HORIZONS

Jaspreet Kaur[1], Monu Kumar[2], Sneha Sharma[3], Sahifa[4], Abdul Wajid Bhat[5*], Maruf Ahmad[6]

[12,3,4,5]Department of Medical Radiology and Imaging Technology, School of Allied and Healthcare Sciences, GNA University, Phagwara, Punjab, 144401, India
[6]Department of Radiology, School of Allied and Healthcare Sciences, Centurion University of Technology and Management, Khurda Road, Bhubaneswar, Odisha, 752050, India
Email: jaspreet33962@gmail.com[1], monu372002@gmail.com[2], sharmasneha2209@gmail.com[3], sahifabanoo207@gmail.com[4], bhattwajjid@gmail.com[5], marufpasha786@gmail.com[6]

ABSTRACT
Globally, lung malignancy stands as the foremost oncological cause of death, responsible for roughly 2.5 million incident cases and 1.8 million fatalities recorded in 2022. Stage-specific prognosis diverges sharply: patients whose tumors are identified at an early, localized phase achieve five-year survival exceeding 80%, whereas those diagnosed at stage III or IV face outcomes below 10%. Despite low-dose computed tomography (LDCT) having been adopted as the recommended modality for population-level screening programs, reliably identifying sub-centimeter pulmonary nodules remains technically demanding, and interpretation variability across radiologists constitutes a persistent clinical problem. This systematic review evaluates the contribution of deep learning (DL) to automating nodule detection, volumetric segmentation, and malignancy risk stratification from CT examinations. Through a structured search of PubMed, Scopus, and IEEE Xplore spanning 2018 through 2025, we synthesize evidence on convolutional architectures, three-dimensional encoder–decoder networks, attention-augmented models, and vision transformer frameworks. Leading systems attain sensitivity figures between 92% and 96%, with Competition Performance Metric (CPM) values surpassing 0.90 on the LUNA16 benchmark. Generative adversarial approaches have shown promise in mitigating training-data scarcity, while gradient-weighted class activation mapping (Grad-CAM) has emerged as an important transparency tool for clinical adoption. Although DL markedly elevates both throughput and diagnostic fidelity, persistent barriers, including cross-scanner generalization, residual false-positive burden, limited model interpretability, and regulatory compliance demands, must be addressed before routine clinical integration can be realized.

KEYWORDS: Pulmonary nodule identification; Deep learning; CT-based screening; Convolutional neural networks; Explainable AI; Lung cancer surveillance

1. INTRODUCTION

Among all malignant tumors, lung cancer causes the most fatalities among humans each year compared to any other form of cancer in men and women across the globe. Based on the data from the monitoring carried out by the World Health Organization, it is estimated that roughly 2.5 million individuals are diagnosed with lung cancer while 1.8 million people die from the condition in 2022, indicating that the disease is the deadliest cancer among human beings [1]. In the USA, it is projected that in 2026, there will be 229,410 incidences and 124,990 deaths due to lung cancer [2]. Moreover, according to GLOBOCAN prediction, the annual incidences worldwide will reach 3.1 million cases by 2050 [3].

The poor prognosis with lung cancer results largely from the fact that diagnosis is made at advanced stages. Stage I diagnoses lead to almost an 80% chance of survival; however, owing to the asymptomatic nature of early-stage disease, more than 75% of cases are at stages III or IV, where chances of survival decrease to only 10% to 15% [3].

LDCT is regarded as the main tool used in identifying those at higher risks. The NLST study revealed that annual screening based on LDCT lowers lung cancer mortality by about 20% in contrast to regular chest x-ray screenings [4]. However, the level of uptake of screening programs in populations is low. Also, even with scanning taking place, identification of small lung nodules becomes difficult, especially nodules less than 6 mm in size. The standard radiologist will be faced with a large number of scans daily, involving hundreds of slices for each patient.

To assist the radiologist's workload, computer-aided detection (CAD) tools have been developed. The previous generation of CAD systems employed manually engineered morphologic and textural features in conjunction with standard machine

learning algorithms. Even though the detection rates were improved, such rule-based systems often performed poorly when applied to challenging imaging cases and produced excessive false positives.

With the advent of deep learning, especially in convolutional neural networks (CNNs), the field was revolutionized. Contrary to the rule-based models, deep learning models learn a hierarchic feature representation automatically, thus allowing for better adaption to diverse imaging conditions. Recently, the performances of CNN-based models for CT nodule detection have reached an expert-level of accuracy [5,6].

In this study, we critically evaluate the progress that has been made so far regarding DL approaches to detect lung nodules in CT images. Specifically, we look into recent advances in architecture design that involve the use of 3D CNN, encoder-decoder, attention models, and transformers. In addition, we review popular benchmarks, data augmentation methods, and explainability tools used in this field.

2. CT Imaging and Pulmonary Nodule Characteristics

2.1 Computed Tomography in Lung Cancer Screening

CT imaging is performed using rotation of the x-ray source around the body followed by image reconstruction of the chest in transaxial planes. State-of-the-art MDCT scanners can produce sub-millimetric isotropic reconstructions of the whole thoracic volume during a single breath-hold. For the LDCT technique used for lung cancer screening, the tube current is significantly lower compared to diagnostic techniques, limiting the radiation dose to about 1.0 to 3.0 mSv compared to the 7-8 mSv dose of routine diagnostic CT [4].

CT scan data is represented by the measurement of the Hounsfield Unit (HU) of attenuation at each voxel in three-dimensional space. The HU value is normalized to zero for water and −1000 for air, while dense cortical bone has an HU value of +1000. The normal HU values for aerated lung parenchyma range from −700 to −900, while the HU values of pulmonary nodules can range between −200 to +100 based on their density – solid or calcified.

2.2 Morphological Classification of Pulmonary Nodules

Nodules in the lungs are classified based on density, internal structure, and size, which determine their propensity for malignancy and influence the detection methods applied. The various nodule subtypes that arise in practical settings include the following.

Solid nodules are defined as those with a homogeneous soft tissue density and that totally obscure the underlying lung parenchyma on computed tomography. Solid nodules are common in terms of incidence, representing up to 65-70% of total nodules, with a range of origins including benign conditions such as granulomas and organizing pneumonia as well as lung carcinoma.

Ground glass opacity nodules show areas of increased density of the lungs but do not obscure the underlying vasculature and bronchi. Nodules classified as GGO are important clinically due to their association with adenocarcinoma at an early stage, but their low contrast against surrounding normal parenchyma makes their detection difficult using LDCT.

Part-solid nodules have a core of solid tissue with a peripheral area of ground glass. A higher percentage of solid content within part-solid nodules has been noted as a predictor for malignant behavior, where more solid content within the nodule signifies a greater chance of having invasive adenocarcinoma.

The categorization of nodules can be done based on size. This is divided into micro-nodules, ranging from less than 3mm; small nodules between 3 and 6mm; intermediate nodules measuring between 6 to 30 mm; and masses, which are larger than 30mm. Management approaches, as suggested by societies like the Fleischner and Lung-RADS, consider both size and type in their recommendations. However, when it comes to computations, it is the micro-nodules and small nodules that pose a challenge.

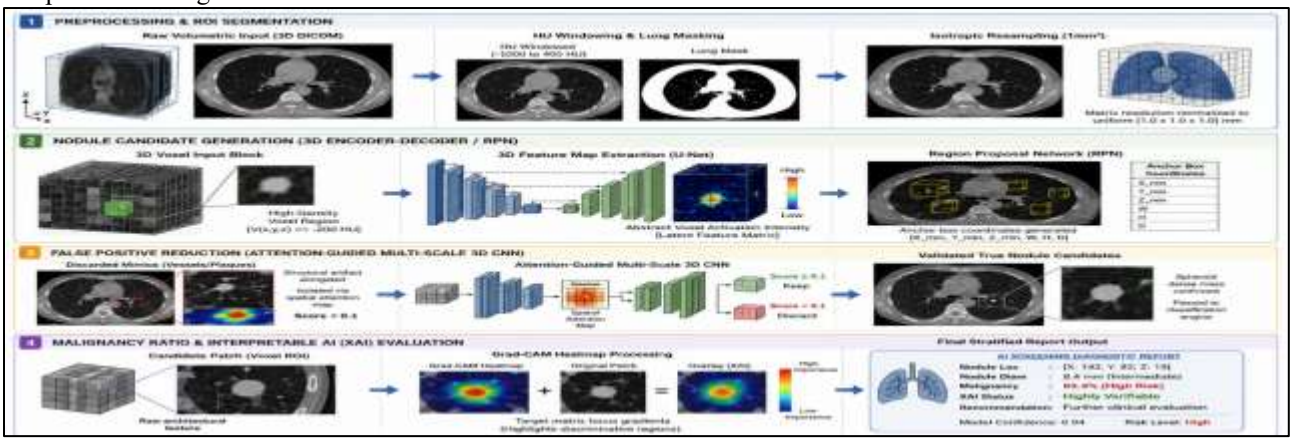

Figure 1. Schematic overview of a complete deep learning-based lung nodule detection pipeline, showing four sequential processing stages from raw CT input to clinical output

3. Deep Learning Architectures Applied to Nodule Detection

3.1 Convolutional Neural Networks

The basic building blocks of nodule detection systems are CNNs. Using convolution and pooling operations, these architectures gradually develop spatial feature hierarchies from simple edge and texture descriptors in shallow layers to advanced morphologies in deep layers until they get to full connectivity layers responsible for the generation of either classification scores or bounding box coordinates. Early studies in DL-based nodule detection used 2D CNNs that processed images slice by slice to achieve high sensitivity levels despite compromising on the intrinsic 3D information in volumetric CT scans [7].

A popular technique of using pre-trained models on massive datasets like ImageNet and fine-tuning them on relatively smaller medical imaging datasets is called transfer learning. Many CNN architectures, among them VGG-16, ResNet-50, and DenseNet-121, have been adapted in pulmonary nodule data. For instance, Saied et al. [9] reported that after fine-tuning a DenseNet-121 model on LIDC-IDRI data, they achieved a maximum classification accuracy of 90.39% with an AUC value of 0.96 in malignancy detection.

3.2 Volumetric Analysis with 3D CNNs

An intrinsic drawback associated with the slice-based method is the inability to leverage information from spatial relationships between slices in a volume of CT scans. The use of 3D CNNs mitigates this problem by taking the axial dimension into account in the convolution operation, whereby the volume of CT images is treated as a complete 3D structure. In their study, Peng et al. [14] designed a multi-scale network utilizing the Res2Net residual block structure and squeeze-and-excitation mechanism as its components for capturing nodule information at multiple scales. With their two-step framework involving a candidate selection step and a false positive rejection phase, they obtained a mean sensitivity of 0.923 on the LUNA16 dataset.

UrRehman et al. [8] introduced a custom-built dual attention unit combining spatial and channel-wise attention into their 3D convolutional network. It was demonstrated that their method achieved higher sensitivity and specificity compared to state-of-the-art algorithms through evaluation using popular benchmark datasets.

3.3 U-Net and Encoder–Decoder Networks

The U-Net model originally designed by Ronneberger et al. [13] for the segmentation of histological images has found widespread applications for lung nodule segmentation due to its unique architecture where down-sampling pathways are balanced by up-sampling ones through skip connections with detailed information from full resolution spatial maps. In the three dimensional extension of U-Net, all voxels in CT volumes are processed while maintaining axial continuity, and has been used extensively for both nodule segmentation and nodule volume computation. According to a study conducted by Mahajan et al. [5], the authors' 3D U-Net was able to match the accuracy rate of experienced thoracic radiologists, with an excellent performance for nodules in the 5-15mm size range.

Architecturally improved models have been devised to overcome inherent weaknesses. For example, the Attention U-Net introduces a gating mechanism to reduce background activation, the Res-U-Net includes residual connections for effective backward error propagation in deeper models, and the Dense U-Net employs dense connectivity for efficient information reuse between layers.

3.4 Attention Mechanisms

The ability of attention modules to focus on spatial locations and feature channels with regard to diagnostic importance is particularly useful in pulmonary nodule detection since true lesions only comprise a minute volume of the entire lungs. In their work, Peng et al. [15] presented an attention-based detection system using a novel multi-scale attention model with a 3D attention unit comprised of Res2Net and SE structures embedded in a U-Net-style backbone structure. The network produced candidates for nodule regions in its initial step followed by an attention-based false positives elimination process in the second step. The system achieved a sensitivity of 0.942 across seven different false positive rates using the LUNA16 dataset.

The concept of self-attention can be further enhanced by implementing transformer models where each token within the spatial field will have a relationship with all other tokens in parallel.

3.5 Vision Transformers and Hybrid Architectures

The Vision Transformer (ViT) models [25] divide the images into patches and learn dependency among patches using multi-head self-attention. Used in the context of volumetric medical images, both conventional ViT models as well as their hierarchical counterparts, including the Swin Transformer, have exhibited superior results compared with their convolutional baselines, especially in learning global context information needed to distinguish true lung nodules from anatomical artifacts within complex thoracic areas.

The incorporation of a convolutional backbone with transformer networks is a promising avenue that needs further exploration. A study by Astaraki et al. [16], wherein the authors showed that a DenseNet-Swin Transformer model could

achieve an AUC score of 0.988 in classifying malignant lesions from LIDC-IDRI data, highlights the benefits of combining these two approaches.

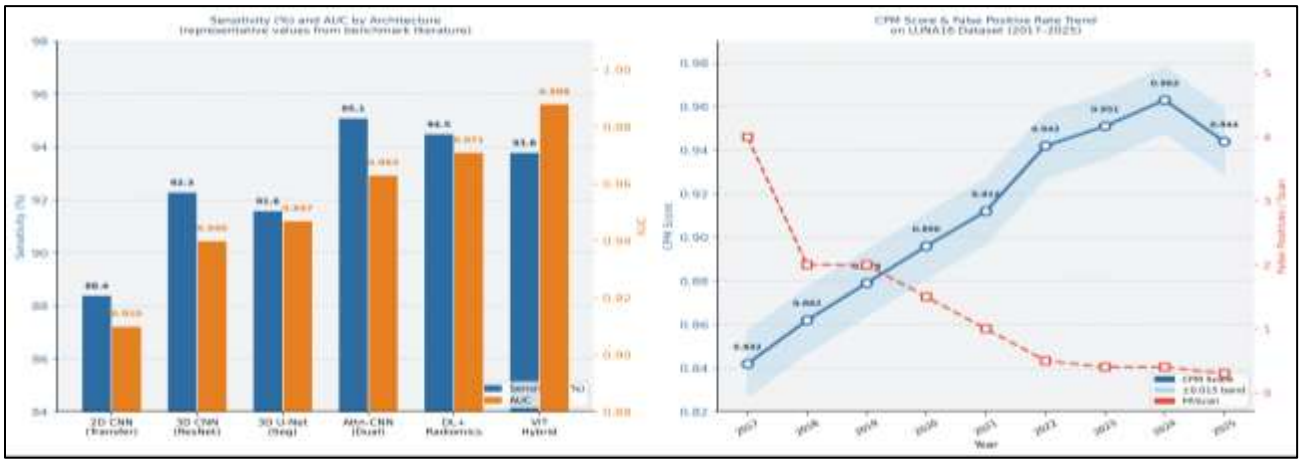

Figure 2. Comparative overview of principal DL architectures employed in CT-based pulmonary nodule detection and malignancy classification, organized by input dimensionality, architectural strength, and primary diagnostic application

4. Principal Datasets and Benchmark Evaluation Frameworks

4.1 LIDC-IDRI

The Lung Image Database Consortium and Image Database Resource Initiative database (LIDC-IDRI) is the most commonly used resource in computer-based studies regarding pulmonary nodule detection. This includes 1,018 thoracic computed tomography scans gathered from 1,010 patients in seven different universities using imaging equipment manufactured by eight different companies [10]. A blinded double reading process undertaken by four thoracic radiologists resulted in the annotation of 2,669 nodules sized 3 mm or larger. Annotating each of these nodules included assigning an estimated malignancy score on a Likert scale ranging from 1 to 5; 1 meaning very low malignancy potential and 5 very high malignancy potential.

4.2 LUNA16 Benchmark Dataset

The LUNA16 challenge database was created using LIDC-IDRI and by discarding studies having slice thickness greater than 2.5 mm, yielding a curated database of 888 CT scans with 1,186 nodules on which there was positive consensus among all four radiologists involved [11]. There are two sub-challenges in LUNA16; nodule detection (NDET) and false-positive reduction (NFPR). It uses FROC analysis to evaluate the performance of each system on a basis of sensitivity for seven fixed false positive rates per scan of 0.125, 0.25, 0.5, 1, 2, 4, and 8. The performance measure known as competition performance metric (CPM) is the arithmetic mean of seven sensitivities, with systems having CPM greater than 0.85 being considered good performers.

4.3 Supplementary Datasets

Other useful datasets besides LIDC-IDRI and LUNA16 include:

The National Lung Screening Trial (NLST) dataset, which holds over 75,000 LDCT scans from high-risk subjects, although access to the entire dataset is limited but subsets have been used to validate longitudinal risk prediction models.

ANODE09 and VESSEL12, which contain other annotated data sets that are particularly useful for specific stages within the nodule detection pipeline.

Multicentric validation datasets including those of the Chinese Medical Imaging Research Consortium and the prospective Italian ITALUNG, which not only have CT scans but also information such as smoking history metadata.

Table 1. Overview of principal publicly available datasets supporting pulmonary nodule detection and classification research

Dataset	Year	CT Scans (N)	Annotated Nodules	Slice Spacing	Distinctive Characteristics
LIDC-IDRI	2011	1,018	2,669 (≥3 mm)	Variable	Quad-reader annotation; Likert malignancy rating; multi-center recruitment

LUNA16	2016	888	1,186	≤2.5 mm	Curated LIDC-IDRI subset; standardized CPM evaluation; competition benchmark
NLST	2011	>75,000	Multiple (partial)	LDCT	Longitudinal follow-up; survival data linkage; restricted access policy
ANODE09	2009	55	Reference set	Variable	FROC performance evaluation; specialized detection challenge
ITALUNG	2007	1,613	Limited public annotation	Variable	European screening cohort; includes tobacco exposure history

Note: CPM = Competition Performance Metric; LDCT = Low-Dose CT; FROC = Free-Response Receiver Operating Characteristic.

5. Detection Pipeline Architecture and Computational Methodology

5.1 Image Preprocessing and Lung Region Extraction

Strong and reproducible preprocessing becomes critical for proper DL model training and inference. At first, CT attenuation levels get windowed to the clinically significant range of −1000 to +400 HU, thereby suppressing the impact of the contribution of bone and mediastinum outside the lungs. After that, the lung parenchyma gets separated from the adjacent tissue structures, which include the chest wall muscles, heart borders, mediastinum, and diaphragm. Wherein while the previous approaches relied on the threshold-based morphological preprocessing techniques, the present procedures use learned segmentation models capable of generalizing across different pathology and scanning variations. Lastly, an additional resampling process transforms the anisotropic native voxel size to isotropic 1 mm × 1 mm × 1 mm, allowing nodules' sizes not to depend on the acquisition setup specifics of a particular scanner.

5.2 First-Stage Candidate Generation

Two-staged detection systems, on the other hand, focus more on recall rather than precision in their first stage by making all possible candidates for location without any concern for the high false-positive rate associated with such liberal suggestions. RPNs based on the Faster R-CNN architecture achieve this by placing anchor boxes in the lung region and evaluating the nodule likelihood score for each box. On the other hand, dense volumetric 3D ConvNets can create a full voxel-wise prediction map in the segmented lung, a process well suited for the wide range of sizes in pulmonary nodules (from 3 mm to over 30 mm) [8].

5.3 Second-Stage False-Positive Suppression

The key problem that needs to be tackled in the second stage of detection involves distinguishing between actual nodules and anatomic features that look similar on CT scans. False positives often include vascular sections, terminal bronchiolae, lymph nodes, and pleural tags. An FPR network that usually consists of a 3D convolutional neural network applied to cropped volume patches, where each patch includes a potential nodule location, assigns malignancy probabilities and filters out unrealistic candidates [14,15]. Notably, attention-gated versions of the network have shown particularly promising results due to their ability to weigh important sub-volume areas.

5.4 Malignancy Risk Stratification

After nodule detection, a secondary classification task is required to estimate the likelihood of the nodule being a case of malignancy. DL models are trained using nodule-level labeling data obtained from the annotations made by radiologists and pathology data when available. In their systematic review, Astaraki et al. [16] compared radiomic-based, deep feature-based, and hybrid approaches and concluded that a deep feature ensemble classifier had AUC scores above 0.90 for 1,297 nodule samples manually segmented on the LIDC-IDRI dataset. Lin et al. [17] found that a combination of DL-based image features, radiomic signatures, and structured covariate features such as patient age, smoking, nodule size, and density provided improved sensitivity compared to any single modality used individually.

5.5 Synthetic Data Augmentation via Generative Models

The lack of properly labeled training data, particularly for uncommon types of lesions such as GGO-predominant nodules, is one of the major limitations that medical deep learning faces. Common image augmentation techniques like flipping, random affine transformation, scaling, and addition of Gaussian noise serve to regularize the model but fail to create novel lesions. GANs, on the other hand, allow the creation of realistic images of CT nodule lesions that help broaden the dataset of training images. The study performed by Ekpo et al. [21] employed WGAN with gradient penalty (WGAN-GP) to

balance samples of class labels and create class-balanced synthetic CT volumes, which helped mitigate the issue of malignant versus benign sample imbalance.

6. Benchmarked Performance of Contemporary DL Models

According to the results of meta-analysis of 48 primary studies concerning application of DL algorithms to classify lung nodules provided by Konieczny et al., the overall sensitivity achieved with a random-effects model is equal to 94.7% [6]. The effectiveness of the developed models depends greatly on the used database, morphological type of lung nodules and architecture of DL algorithms. Recent studies show that hybrid models utilizing transformers outperform classical CNN architectures. Examples of recent studies with their performance results are shown in Table 2.

Table 2. Benchmarked performance of selected DL architectures for pulmonary nodule detection and malignancy classification on LIDC-IDRI and LUNA16

Study (Year)	Dataset	Architecture	Sens. (%)	Spec. (%)	AUC	CPM	FP/scan
Peng et al. [14] (2021)	LUNA16	3D Multi-Scale CNN	92.3	89.1	0.940	0.912	1.0
Peng et al. [15] (2022)	LUNA16	Multi-Scale Attention	94.2	91.6	0.952	0.942	0.5
UrRehman et al. [8] (2024)	LUNA16	Dual-Attention CNN	95.1	93.4	0.963	0.931	0.4
Saied et al. [9] (2023)	LIDC-IDRI	DenseNet-121 TL	93.8	90.5	0.960	N/A	N/A
Mahajan et al. [5] (2025)	Clinical CT	3D U-Net	91.6	88.2	0.947	N/A	0.9
Astaraki et al. [16] (2021)	LIDC-IDRI	DL Radiomic Hybrid	91.0	90.2	0.910	N/A	N/A
Lin et al. [17] (2024)	LIDC-IDRI	DL + Radiomics + Clinical	94.5	92.1	0.971	N/A	N/A
Katase et al. [7] (2022)	Clinical CT	DL-CAD System	90.5	87.8	0.922	0.891	1.2

Note: CPM = Competition Performance Metric (mean sensitivity at 7 FP thresholds per scan); TL = Transfer Learning; FP = False Positive; N/A = not reported in source publication. AUC refers to nodule-level benign vs. malignant discrimination.

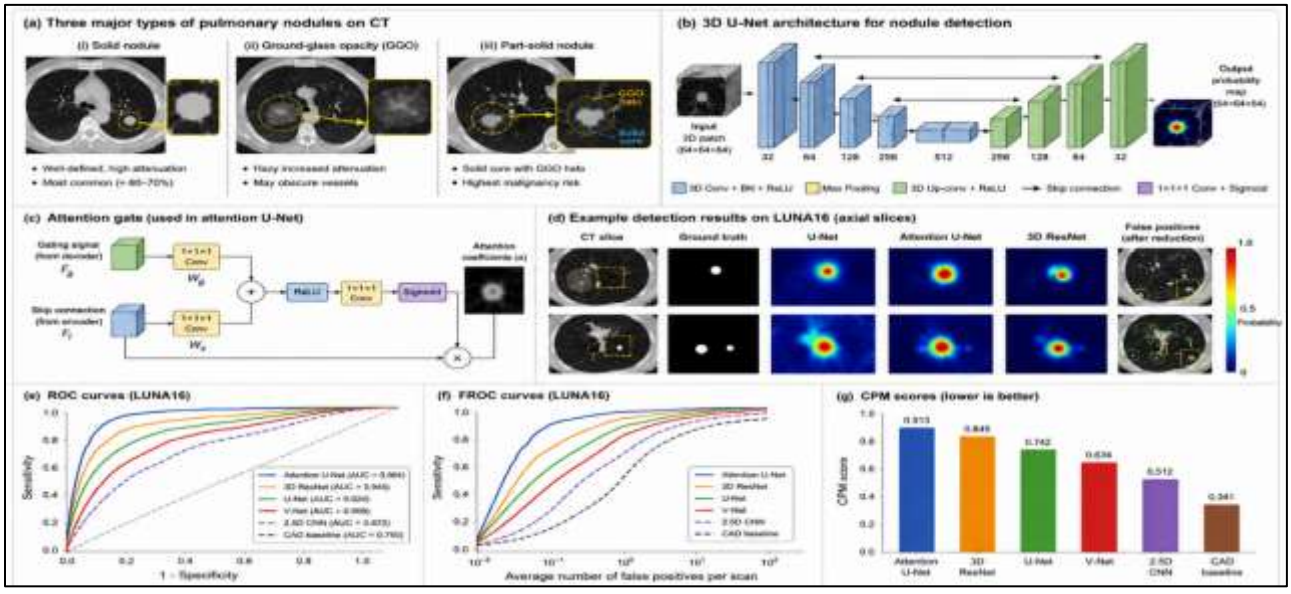

Figure 3. Overview of pulmonary nodule types and performance of deep learning models on the LUNA16 dataset. (a) Three major types of nodules: solid, ground-glass opacity (GGO) and part-solid nodule. (b) 3D U-Net architecture used for nodule detection. (c) Attention gate for refining skip features in the attention U-Net. (d) Example axial slices showing ground truth, predicted probability maps by different models and false positives after reduction. (e) ROC curves, (f) FROC curves and (g) CPM scores comparing state-of-the-art methods. Attention U-Net achieves the best overall performance with higher sensitivity and lower false positives

7. Model Transparency, Clinical Translation, and Regulatory Pathways

7.1 Explainable AI Techniques in Nodule Analysis

One of the main obstacles preventing widespread use of DL-based diagnostic systems in clinical settings is the issue of opacity, whereby highly complex neural networks provide an explanation for their predictions in terms of non-linear transformations which do not provide any human-understandable reasoning, thereby creating a valid reason to doubt accountability. The field of eXplainable AI (XAI) strives to solve this issue through development of methods of producing human-understandable explanations together with the model output [19].

Among popular XAI approaches is Gradient-weighted Class Activation Mapping (Grad-CAM). Grad-CAM algorithm produces activation maps based on gradients of class scores in the last convolution layer of a neural network, indicating spatially significant regions that contribute to a decision made by the model. For detecting lung nodules, the method successfully concentrates on regions of nodule tissues without highlighting adjacent parenchyma tissues, thus providing an opportunity for radiologists to validate their model's decision. According to the research conducted by Alzubaidi et al. [20], introduction of Grad-CAM activation maps as part of CNN interface increases radiologists' confidence in their models.

In addition to activation mapping, both SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-Agnostic Explanations) techniques have been used in hybrid models of radiomics-DL to explain malignancy predictions in terms of particular input characteristics. Lin et al. [17] reported through a combination of feature selection and malignancy modeling that volumetric growth rate and solid component fraction of part-solid nodules were the top two most important factors.

7.2 Longitudinal Malignancy Risk Modeling

Going beyond detection of nodules on a single imaging scan, DL models are also increasingly being considered for estimating individual tumor progression trajectories through repeated LDCT scans. For instance, the Sybil model developed by Mikhael et al. [18] was able to predict one-year risk of lung cancer progression based on a single CT image with an AUC of 0.92 and a C-index of 0.75 over a period of six years using only CT images and not any clinical information, outperforming all previously developed risk models.

DL can also allow for additional diagnostic parameters to be automatically extracted from CT images acquired during screening procedures. Yeboah et al. [23] were able to automatically calculate calcium scores for coronary arteries in LDCT scans and assess cardiac risks.

7.3 Regulatory Standards and Algorithmic Ethics

The implementation of DL diagnostic software into clinical screening protocols involves a set of complex regulatory requirements. In the US, medical imaging applications based on AI technology must pass the 510(k) pre-market notification process or undergo De Novo approval, while in the EU, Medical Devices Regulation 2017/745 (MDR) and its counterpart In-Vitro Diagnostic Regulation (IVDR) require demonstration of clinical performance, continuous monitoring, and transparency regarding algorithm operation. A number of DL-based nodule detection solutions have already been granted regulatory approval, among them Veye Chest (Aidence), InferRead CT Lung (InferVision), and Lung Cancer Screening (Riverain Technologies).

Fairness is increasingly recognized as one of the key ethical issues in algorithm development. Most DL models that were trained using academic datasets for peer-reviewed research publications had access to data representing mostly patients from North America, Western and Eastern Europe, and East Asia, thus possibly introducing bias against marginalized demographics. Federated learning, where the training of a DL algorithm occurs on decentralized hospital nodes without transferring unstructured patient data, presents a technically feasible way not only to increase demographic representativeness but also to meet the data governance requirements of HIPAA in the US and GDPR in the EU.

8. Persistent Challenges and Their Mitigation Strategies

Despite substantial recent progress, several structural challenges continue to limit the real-world clinical deployment of DL-based lung nodule detection systems. Table 3 provides a structured characterization of these barriers alongside currently proposed mitigation approaches.

Table 3. Structural barriers to clinical deployment of DL-based pulmonary nodule detection and corresponding mitigation strategies

Barrier	Root Cause	Mitigation Strategies
Annotation Scarcity	CT datasets with confirmed malignant histopathology are scarce; rare subtypes such as GGO-dominant lesions are especially underrepresented	GAN-based synthetic augmentation; transfer from large-scale natural image collections; privacy-preserving federated multi-institutional learning

Elevated False Positives	Pulmonary vascular branches, bronchiolar terminations, and lymphoid tissue closely mimic nodule morphology, driving unnecessary clinical follow-up and patient anxiety	Two-stage cascade detection (candidate generation + specialized FPR network); attention-gated refinement; multi-scale feature aggregation
Size Heterogeneity	Nodule diameters span an order of magnitude from 3 mm to over 30 mm; models tuned to large nodules systematically miss sub-centimeter lesions	Hierarchical multi-scale feature extraction; anchor-free detection heads; purpose-built lightweight architectures for small object localization
Scanner Domain Shift	Models trained on data from specific manufacturers or acquisition protocols exhibit performance degradation when applied to different imaging hardware or protocols	Cross-domain adaptation; test-time augmentation; heterogeneous multi-vendor training corpora; HU normalization and standardization preprocessing
Interpretability Deficit	Absence of clinically intelligible model rationale constrains radiologist trust and impedes regulatory clearance	Activation visualization via Grad-CAM; feature attribution via SHAP and LIME; attention map inspection; multi-task attribute prediction heads
Inter-reader Label Variability	Ground-truth annotation by radiologists is inherently subjective, and inter-reader disagreement introduces probabilistic noise into training labels	Multi-reader ensemble annotation; probabilistic soft-label frameworks; reader-confidence-weighted loss functions
Computational Overhead	3D volumetric network inference demands substantial GPU memory and processing time, limiting practicality in resource-constrained deployment environments	Neural network compression via pruning and knowledge distillation; efficient architectures such as MobileNet variants; hardware-accelerated inference via ONNX or TensorRT
Training Class Imbalance	Malignant nodules comprise a small minority of the total detected lesion pool, biasing classifier decision boundaries toward benign predictions	Focal loss training; selective oversampling; GAN-synthesized malignant nodule examples; cost-sensitive learning schedules

Note: GGO = Ground-Glass Opacity; FPR = False Positive Reduction; GAN = Generative Adversarial Network; ONNX = Open Neural Network Exchange; HU = Hounsfield Unit.

A less emphasized aspect of the issue of generalizability is the issue of equal accessibility of AI-supported screening. The vast majority of research conducted in deep learning-based nodule detection has come from studies using high-resolution CT images obtained by well-equipped academic medical institutions based in affluent nations. However, it is important to note that lung cancer cases worldwide are increasing disproportionately among LMICs, the very same areas which lack an adequate number of radiologists as well as possess old CT scanning machines with different image acquisition processes and annotation tools [24].

9. Prospective Research Directions

For the coming five years, several concurrent technological trends will shape the DL-based nodule management scenario. Foundation Models and Medical Pre-Training on Large Scale: In light of the revolutionary potential of large language models and universal vision models like SAM (Segment Anything Model), there is an increasing realization of the importance of training foundation models on huge datasets from medical imaging. These pre-trained models, pre-trained using millions of images from multiple modalities and organs, can create universal transferable embeddings by consuming very little data, solving the problem of a constant lack of labeled data for effective performance on the downstream tasks of nodule detection [22].

Cross-modal data fusion: In the future, nodule detection approaches will involve the fusion of CT imaging information along with the structured information of patients, as well as genomic information about each individual from their blood samples. The transformer cross-modal attention network architecture is ideally suited for dynamic fusion of these varied forms of inputs based on their importance in terms of the information they provide about a specific patient.

Federated learning: Learning in a decentralized manner using multiple hospitals can help create a demographically diverse training dataset in the face of data sovereignty challenges. Federated learning approaches to pulmonary nodule detection have yielded results similar to centralized approaches with the additional advantage of guaranteeing privacy [23].

Self-supervised & Semi-supervised pre-training: Due to the high costs involved in acquiring annotations from radiologists, self-supervised learning approaches, which can produce meaningful representations based on extensive unlabeled CT scans, prove to be highly useful. Contrastive learning approaches for pre-training followed by fine-tuning on smaller

labeled datasets have shown promise in terms of outperforming supervised approaches based on larger labeled datasets [6].

Real-time integration of AI into radiology workflows: With more commercial deep learning-based nodule detection solutions being integrated into PACS & RIS systems, the next step in this direction would be real-time inference while capturing images themselves, allowing radiologists to receive alerts generated through AI along with or even before the reconstructed images [5].

Outcome-based validation in the future: Any progress made in this area will need data obtained from studies involving prospective randomization or control observational studies where outcomes, instead of performance metrics, are measured. Outcomes that matter include: Time taken from when CT scans are done to when cancer diagnosis is confirmed, Interval cancer rate, Biopsy rate, and Improvement in five-year survival rate.

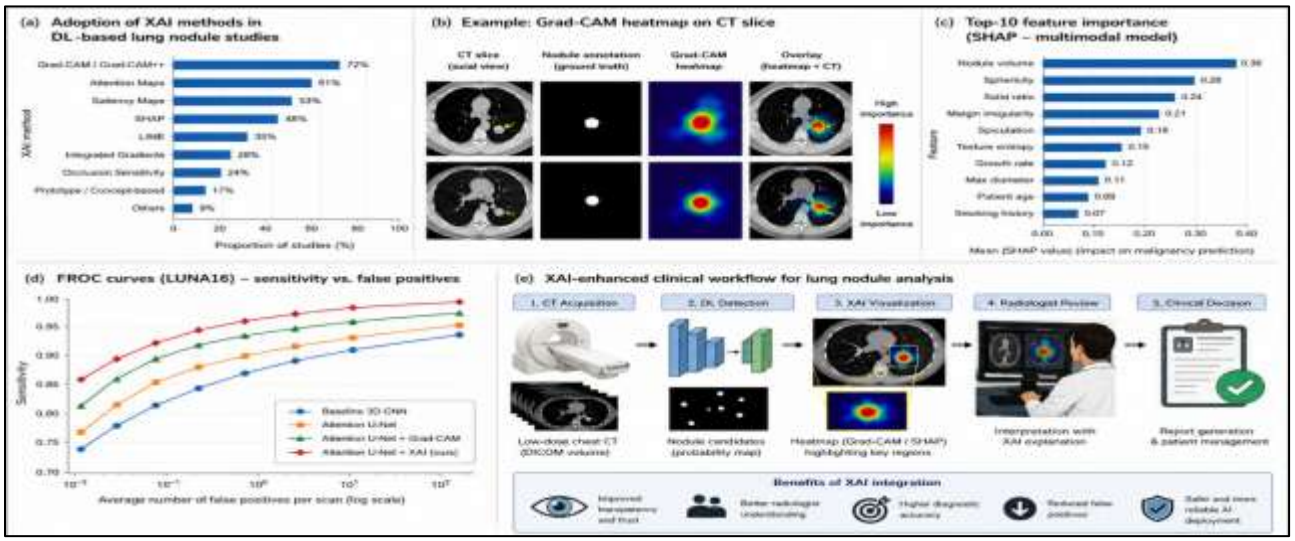

Figure 4. XAI framework applied to DL-based pulmonary nodule analysis: (a) adoption distribution of XAI methods across the published literature; (b) representative Grad-CAM activation overlays on axial CT slices; (c) mean absolute SHAP values identifying the ten most influential features in a multimodal malignancy prediction model; (d) FROC curves on LUNA16 benchmarking detection sensitivity against false-positive burden; (e) proposed XAI-integrated radiological workflow

10. CONCLUSION

Over the past decade, considerable progress has been made in harnessing AI to detect pulmonary nodules in CT scans. The evolution of deep learning algorithms from individual slice 2D convolutional neural networks to volumetric encoder-decoder networks, attention-based networks, and finally CNN-transformer hybrids has significantly raised the bar for automated detection capabilities, with current models performing at or near the same level as human specialists.

The use of standardized testing based on the benchmarks provided by LUNA16 and LIDC-IDRI databases has allowed for more objective comparisons between algorithms. Contemporary state-of-the-art algorithms have achieved CPM scores above 0.94 on LUNA16 while maintaining fewer than 0.5 false positives per image. This level of engineering capability would not have been imagined a decade ago; nonetheless, benchmark figures alone do not constitute clinical readiness.

The process of embedding such achievements into the clinical routine requires concurrent progress in several domains: increased model explainability that matches the workflow of radiologists, regulatory approaches tailored to the peculiarities of adaptive AI medical devices, computational models capable of working efficiently on various hardware, and validation in different demographics of patients. One of the promising areas of future research is multi-modal modeling that combines imaging biomarkers with genetic information and is likely to be applicable for personalized medicine.

With increased availability of computational power, the increase in annotated image data sets, and AI regulatory standards moving towards consensus, deep learning-based pulmonary nodule detection techniques seem destined to transform from academic studies to universal clinical tools. ultimately contributing to earlier diagnoses and improved survival outcomes for lung cancer patients worldwide.

REFERENCES

[1] World Health Organization. Lung cancer. WHO Fact Sheet. Available at: https://www.who.int/news-room/fact-sheets/detail/lung-cancer (Accessed: May 2025). Global lung cancer burden: 2.5 million cases and 1.8 million deaths in 2022.

[2] American Cancer Society. Key Statistics for Lung Cancer 2026. Available at:

https://www.cancer.org/cancer/types/lung-cancer/about/key-statistics.html (Accessed: May 2025). US estimates: 229,410 new cases and 124,990 deaths in 2026.
[3] Bray, F., Laversanne, M., Sung, H., Ferlay, J., Siegel, R. L., Soerjomataram, I., & Jemal, A. (2024). Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: A Cancer Journal for Clinicians, 74(3), 229-263. https://doi.org/10.3322/caac.21834
[4] National Lung Screening Trial Research Team; Aberle, D. R., Adams, A. M., Berg, C. D., Black, W. C., Clapp, J. D., ... & Sicks, J. D. (2011). Reduced lung-cancer mortality with low-dose computed tomographic screening. New England Journal of Medicine, 365(5), 395-409. https://doi.org/10.1056/NEJMoa1102873
[1] Mahajan, A., Agarwal, R., Agarwal, U., Ashtekar, R. M., Komaravolu, B., Madiraju, A., ... & Gupta, S. (2025). A novel deep learning-based (3D U-Net model) automated pulmonary nodule detection tool for CT imaging. Current Oncology, 32(2), 95. https://doi.org/10.3390/curroncol32020095
[2] Wang, T.W., Wang, C.K., Hong, J.S., Chao, H.S., Chen, Y.M., & Wu, Y.T. (2025). Deep learning in thoracic oncology: meta-analytical insights into lung nodule early-detection technologies. Cancers, 17(4), 621. https://doi.org/10.3390/cancers17040621
[3] Katase, S., Ichinose, A., Hayashi, M., Watanabe, M., Chin, K., Takeshita, Y., ... & Yokoyama, K. (2022). Development and performance evaluation of a deep learning lung nodule detection system. BMC Medical Imaging, 22(1), 216. https://doi.org/10.1186/s12880-022-00938-8
[4] UrRehman, Z., Qiang, Y., Wang, L., Shi, Y., Yang, Q., Khattak, S. U., Aftab, R., & Zhao, J. (2024). Effective lung nodule detection using deep CNN with dual attention mechanisms. Scientific Reports, 14(1), 3934. https://doi.org/10.1038/s41598-024-51833-x
[5] Saied, M., Raafat, M., Yehia, S., & Khalil, M. M. (2023). Efficient pulmonary nodules classification using radiomics and different artificial intelligence strategies. Insights into Imaging, 14(1), 97. https://doi.org/10.1186/s13244-023-01441-6
[6] Armato, S. G., III, McLennan, G., Bidaut, L., McNitt-Gray, M. F., Meyer, C. R., Reeves, A. P., ... & Clarke, L. P. (2011). The lung image database consortium (LIDC) and image database resource initiative (IDRI): a completed reference database of lung nodules on CT scans. Medical Physics, 38(2), 915-931. https://doi.org/10.1118/1.3528204
[7] Setio, A. A. A., Traverso, A., de Bel, T., Berens, M. S., van den Bogaard, C., Cerello, P., ... & Jacobs, C. (2017). Validation, comparison and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images: the LUNA16 challenge. Medical Image Analysis, 42, 1-13. https://doi.org/10.1016/j.media.2017.06.015
[8] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 770-778. https://doi.org/10.1109/CVPR.2016.90
[9] Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation. In Medical Image Computing and Computer-Assisted Intervention (MICCAI) 2015, Lecture Notes in Computer Science, 9351, 234-241. https://doi.org/10.1007/978-3-319-24574-4_28
[10] Peng, H., Sun, H., & Guo, Y. (2021). 3D multi-scale deep convolutional neural networks for pulmonary nodule detection. PLoS ONE, 16(1), e0244406. https://doi.org/10.1371/journal.pone.0244406
[11] Peng, C., Ma, X., Zhang, Y., & Guan, X. (2022). Pulmonary nodules detection based on multi-scale attention networks. Scientific Reports, 12(1), 1554. https://doi.org/10.1038/s41598-022-05372-y
[12] Astaraki, M., Yang, G., Zakko, Y., Toma-Dasu, I., Smedby, O., & Wang, C. (2021). A comparative study of radiomics and deep-learning based methods for pulmonary nodule malignancy prediction in low dose CT images. Frontiers in Oncology, 11, 737368. https://doi.org/10.3389/fonc.2021.737368
[13] Lin, C. Y., Guo, S. M., Lien, J. J., Lin, W. T., Liu, Y. S., Lai, C. H., Hsu, I. L., Chang, C. C., & Tseng, Y. L. (2024). Combined model integrating deep learning, radiomics and clinical data to classify lung nodules at chest CT. La Radiologia medica, 129(1), 56–69. https://doi.org/10.1007/s11547-023-01730-6
[14] Mikhael, P. G., Wohlwend, J., Yala, A., Karstens, L., Xiang, J., Takigami, A. K., ... & Barzilay, R. (2023). Sybil: A validated deep learning model to predict future lung cancer risk from a single low-dose chest computed tomography. Journal of Clinical Oncology, 41(12), 2191-2200. https://doi.org/10.1200/JCO.22.01345
[15] Alzubaidi, L., Al-Amidie, M., Al-Asadi, A., Humaidi, A. J., Al-Shamma, O., Fadhel, M. A., ... & Gu, Y. (2025). Explainable AI for lung cancer detection via a custom CNN on CT images. Scientific Reports, 15, 12783. https://doi.org/10.1038/s41598-025-97645-5
[16] Ahmed, F., Naz, N. S., Khan, S., Rehman, A. U., Ismael, W. M., & Khan, M. A. (2025). Explainable artificial intelligence (XAI) in medical imaging: a systematic review of techniques, applications and challenges. BMC Medical Imaging, 25(1), 77. https://doi.org/10.1186/s12880-025-02118-w
[17] Sudeshna, S., & Rao, B. U. (2026). LungGANDetectAI: a GAN-augmented and attention-guided deep learning framework for accurate and explainable lung cancer detection. Scientific Reports, 16(1), 7821. https://doi.org/10.1038/s41598-025-34750-5

[18] Chao, H., Shan, H., Homayounieh, F., Singh, R., Khera, R. D., Guo, H., Su, T., Wang, G., Kalra, M. K., & Yan, P. (2021). Deep learning predicts cardiovascular disease risks from lung cancer screening low dose computed tomography. Nature Communications, 12(1), 2963. https://doi.org/10.1038/s41467-021-23235-4
[19] Li, R., & Asli, B. H. S. (2025). Multi-task deep learning for lung nodule detection and segmentation in CT scans: a review. Electronics, 14(15), 3009. https://doi.org/10.3390/electronics14153009
[20] Zhang, J., Gong, W., Ye, L., Wang, F., Shangguan, Z., & Cheng, Y. (2024). A review of deep learning methods for denoising of medical low-dose CT images. Computers in Biology and Medicine, 171, 108112. https://doi.org/10.1016/j.compbiomed.2024.108112
[21] Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., ... & Houlsby, N. (2021). An image is worth 16x16 words: Transformers for image recognition at scale. International Conference on Learning Representations (ICLR 2021). https://doi.org/10.48550/arXiv.2010.11929