

# Artificial Intelligence in Lung Cancer Diagnosis: A Comprehensive Review of ML and DL Approaches

Dr. A. Nisha Jebaseeli

Assistant Professor of Computer Science, Center for Distance and Online Education  
Bharathidasan University, Tiruchirappalli -620024, Tamilnadu, India.

---

## Article History:

*Received: 12-08-2023*

*Revised: 05-09-2023*

*Accepted: 18-10-2023*

## Abstract:

Lung cancer remains one of the leading causes of cancer-related mortality worldwide, necessitating accurate and timely diagnostic approaches. This comprehensive literature review examines the application of deep learning and machine learning techniques for lung cancer detection during the period 2019-2023. Based on a systematic analysis of 282 unique papers filtered to 30 highly relevant studies, this review synthesizes current methodologies, datasets, and performance metrics in the field. The analysis reveals that convolutional neural networks (CNNs) and their variants, particularly 3D architectures and hybrid models, have emerged as the dominant approaches, achieving accuracies ranging from 71% to 99% across various datasets. The LIDC-IDRI dataset has become the de facto standard for benchmarking, while imaging modalities primarily focus on CT scans. Key findings indicate that ensemble methods, transfer learning, and attention mechanisms significantly enhance detection accuracy and reduce false positive rates. However, challenges remain in model interpretability, dataset diversity, and clinical deployment. This review provides a structured analysis of methods, datasets, and performance metrics, offering insights for researchers and practitioners advancing automated lung cancer detection systems.

---

## 1 INTRODUCTION

Lung cancer is the leading cause of cancer-related deaths globally, with early detection being critical for improving patient survival rates and treatment outcomes [1], [2]. Traditional diagnostic methods, including radiological interpretation of chest X-rays and computed tomography (CT) scans, are time-consuming and subject to inter-observer variability. The advent of artificial intelligence, particularly deep learning and machine learning techniques, has revolutionized medical image analysis and offers promising solutions for automated lung cancer detection [3].

Between 2019 and 2023, there has been an exponential growth in research applying deep learning methods to lung cancer detection, driven by advances in computational power, availability of large annotated datasets, and improvements in neural network architectures. This literature review synthesizes findings from 30 highly relevant studies selected from a comprehensive search yielding 282 unique papers, all filtered to the 2019-2023 timeframe. The review focuses on three critical dimensions: (1) the methods and architectures employed, (2) the datasets and data characteristics utilized, and (3) the performance metrics achieved.

The primary objective of this review is to provide researchers, clinicians, and data scientists with a comprehensive understanding of the current state-of-the-art in deep learning-based lung cancer detection, identify key trends and gaps, and offer recommendations for future research directions.

## **2 BACKGROUND AND THEORETICAL FOUNDATIONS**

Deep learning, a subset of machine learning based on artificial neural networks with multiple layers, has demonstrated remarkable success in medical image analysis tasks [3]. Convolutional Neural Networks (CNNs) have emerged as the predominant architecture for image-based tasks due to their ability to automatically learn hierarchical feature representations from raw pixel data [3]. CNNs employ convolutional layers for feature extraction, pooling layers for dimensionality reduction, and fully connected layers for classification, making them particularly well-suited for analyzing medical imaging data [1], [2].

The application of deep learning to lung cancer detection typically involves several stages: (1) image preprocessing and enhancement, (2) lung region segmentation, (3) nodule detection, and (4) nodule classification (benign vs. malignant) [9], [10]. Each stage presents unique challenges, including handling high-dimensional 3D medical images, addressing class imbalance, reducing false positives, and ensuring model interpretability for clinical acceptance [4], [5].

Transfer learning, which leverages pre-trained models on large-scale datasets (e.g., ImageNet) and fine-tunes them for specific medical imaging tasks, has become a popular strategy to overcome limited medical data availability [19]. Additionally, ensemble methods that combine multiple models have shown promise in improving robustness and accuracy [5], [8].

## **3 METHODS AND ARCHITECTURES**

### **3.1 Convolutional Neural Networks**

Convolutional Neural Networks (CNNs) represent the foundational architecture for lung cancer detection in the reviewed literature. Multiple studies employed CNN-based models for analyzing chest radiographs and CT scans to identify patterns indicative of lung malignancies [1], [2], [11]. The multi-layer structure of CNNs, combined with automatic weight learning and local weight communication capabilities, consistently demonstrated superior accuracy compared to traditional machine learning methods [3].

Middha et al. [1] developed a CNN-based model specifically designed for early lung cancer detection using chest radiographs, demonstrating exceptional efficacy in identifying subtle patterns. The model integrated advanced features for feature extraction and classification, enhancing both sensitivity and specificity compared to conventional methods. Similarly, Mounika et al. [2] combined CNNs with virtual reality technologies to provide immersive visualization environments for radiologists, aiming to enhance diagnostic efficiency and minimize false positives.

Several studies employed custom CNN architectures tailored to lung cancer detection tasks. Prashanthi et al. [15] developed a CNN model that achieved 99.3% accuracy in malignancy detection on CT-guided biopsy images. The study demonstrated that modern deep learning approaches could significantly contribute to early diagnosis and personalized treatment strategies. Nandula et al. [20] utilized CNN technology to classify lung nodules as affected or normal and to identify malignant development, extending CNN capabilities with tomography filters to suppress identified dangerous or benign tumors.

The implementation of CNNs for lung cancer detection typically involves image data as input to classify various cancer cell types, including adenocarcinoma, large cell cancer, and squamous cell carcinoma [20]. Studies consistently reported that CNNs outperformed traditional machine learning methods such as Support Vector Machines (SVM) and Artificial Neural Networks (ANN) in terms of accuracy [17].

### 3.2 3D Deep Learning Architectures

The three-dimensional nature of medical imaging data, particularly CT scans, has motivated the development of 3D deep learning architectures that can capture spatial information across multiple slices. These architectures represent a significant advancement over 2D approaches by processing volumetric data directly [9], [10], [12].

Nasrullah et al. [9] presented a deep learning framework utilizing two deep 3D Faster R-CNN models and U-Net encoder-decoder with MixNet for nodule detection and feature learning. For classification, they proposed a gradient boosting machine (GBM) combined with 3D MixNet. This multi-stage approach achieved a sensitivity of 94%, specificity of 90%, and an area under the receiver operating curve (AUC) of 0.99 on the LIDC-IDRI dataset comprising 1,200 CT images with 3,250 nodules.

The DeepLung system, developed by Zhu et al. [10], employed a 3D Faster R-CNN with a U-net-like encoder-decoder structure for nodule detection and a 3D deep dual path network (DPN) for classification. The system achieved 83.4% FROC performance for nodule detection with a 94.6% recall rate for all nodules. For nodule classification, DeepLung obtained 90.44% accuracy by combining 3D DPN features, nodule size, and raw pixels. Notably, in patient-level diagnosis, DeepLung achieved 81.41% accuracy, comparable to experienced doctors (average 82.31%) and surpassing three out of four individual doctors. The system also reduced false positive detected nodules by 97.02% [10].

Liu et al. [12] employed three-dimensional CNNs trained from scratch with two architectures: CNN1 (16×16×16 input) and CNN2 (32×32×32 input). The study combined these 3D CNNs with traditional machine learning models including distance-weighted nearest neighbor, logistic regression, support vector machines, random forest, and AdaBoosted tree. The ensemble models with 3D CNN demonstrated superior performance, achieving the best overall AUC of  $0.780 \pm 0.063$ , outperforming traditional machine learning models [12].

Chen [16] improved a diagnosis method based on CNN and recurrent neural network (RNN), combining their dual effects for classifying benign and malignant nodules. The study

utilized multitask deep neural network technology with three-dimensional convolution and an improved 3D U-net system, achieving 92.3% accuracy for predicting malignant pulmonary nodules and 82.8% for benign ones [16].

### 3.3 Transfer Learning and Pre-trained Models

Transfer learning has emerged as a powerful technique for lung cancer detection, particularly when dealing with limited medical imaging datasets. This approach leverages pre-trained models developed on large-scale natural image datasets and fine-tunes them for medical imaging tasks [4], [19].

Rai et al. [4] evaluated a custom CNN alongside three fine-tuned transfer learning backbones: DenseNet121, ResNet152, and VGG19. Models were trained with cost-sensitive learning to address class imbalance on the IQ-OTH/NCCD dataset comprising 1,197 chest CT scans. ResNet152 achieved the highest accuracy at 97.3%, while DenseNet121 demonstrated the best overall balance with precision up to 92%, recall up to 90%, and F1-score up to 91%. The study also applied Shapley Additive Explanations (SHAP) for interpretability, visualizing evidence contributing to predictions [4].

S. [19] employed a transfer learning-based framework (ExtRanFS) utilizing a pre-trained VGG16 model as the feature extractor. The study froze all convolution layers and excluded the final classification layer, using an Extremely Randomized Tree Classifier for feature selection. Selected features were fed into a Multi-Layer Perceptron (MLP) classifier for classification into benign, malignant, or normal categories. The ExtRanFS framework achieved an accuracy of 99.09%, sensitivity of 98.33%, and F1-Score of 98.33% on the IQ-OTH/NCCD dataset. The VGG16+MLP model outperformed other pre-trained models including Xception+MLP (96.00% accuracy), MobileNetV2+MLP (97.00% accuracy), and InceptionV3+MLP (94.00% accuracy) [19].

Jaincy et al. [17] employed various deep learning methods including U-Net and SegNet for nodule segmentation, and Deep Deconvolutional Residual Network (DDRNet) for segmentation accuracy. The study utilized multiscale CNNs and models like ProNet and radNet, achieving 98.19% accuracy with CNN classifiers, which outperformed SVM (90.1% accuracy) and ANN [17].

### 3.4 Hybrid and Ensemble Approaches

Hybrid and ensemble approaches that combine multiple models or integrate different types of algorithms have demonstrated superior performance compared to single-model approaches [5], [8], [12].

Jamil et al. [5] examined hybrid AI-radiomics models combining handcrafted radiomic features with deep learning architectures. The study explored specific deep learning models including CNN, U-Net, and VGG-16, along with machine learning classifiers like XGBoost, Random Forest, and SVM for radiomic feature analysis. Through meta-analysis and independent benchmarking on the LIDC-IDRI dataset, the research demonstrated the superior

performance of hybrid AI-radiomics models compared to standalone AI models in lung cancer diagnosis [5].

Jayanthi et al. [8] developed HLNDNet, a Hybrid Deep Learning Approach that hybridizes VGG-16, AlexNet, and GoogleNet architectures. The methodology involved preprocessing with Wiener filter and morphological operations, lung region segmentation using blob detection, and semantic image segmentation by individual architectures. Outputs were combined using a Weighted Voting Approach (WVA) to detect and classify lung nodules. HLNDNet achieved a Pixel Accuracy of 98.39%, mean Class Accuracy of 90.42%, and mean Intersection over Union of 78.68% on CT images from 130 patients in the LIDC database [8].

Oncu [6] developed a novel multimodal AI framework integrating CNNs and Artificial Neural Networks (ANNs). The CNN model was trained on 900 CT images for lung cancer detection and classification, achieving a weighted average accuracy of 91% and near-perfect AUC values for most classes. An ANN was developed using clinical data from 999 patients incorporating 24 features covering demographic, symptomatic, lifestyle, and genetic factors, demonstrating an exceptional overall accuracy of 99% with minimal misclassifications. This multimodal framework leveraged the complementary strengths of both imaging and clinical data analysis [6].

### **3.5 Attention Mechanisms and Explainable AI**

Recent advances have focused on incorporating attention mechanisms and explainable AI techniques to improve model performance and interpretability, addressing critical concerns for clinical adoption [4], [14].

Urrehman et al. [14] developed a bespoke CNN with a dual attention mechanism for lung nodule detection. The attention module incorporated both channel and spatial attention mechanisms to selectively highlight significant features, with global average pooling applied after the attention module to summarize spatial information. The study demonstrated that the proposed model surpassed recent models and achieved state-of-the-art accuracy in lung nodule detection and classification tasks [14].

Rai et al. [4] applied Shapley Additive Explanations (SHAP) for interpretability in their transfer learning framework, visualizing evidence contributing to predictions. This approach provided fast, accurate, and interpretable support for lung cancer screening, addressing the "black box" nature of deep learning models. The integration of explainable AI techniques with high-performing models (ResNet152 achieving 97.3% accuracy) demonstrated that interpretability and performance need not be mutually exclusive [4].

Oncu [6] utilized Grad-CAM (Gradient-weighted Class Activation Mapping) in the CNN component of their multimodal framework to provide interpretable predictions. This technique allowed visualization of which regions of CT images contributed most to the model's classification decisions, enhancing trust and clinical utility [6].

## **4 DATASETS AND DATA CHARACTERISTICS**

### **4.1 Benchmark Datasets**

The availability of large, well-annotated datasets has been crucial for training and evaluating deep learning models for lung cancer detection. Several benchmark datasets have emerged as standards in the field [9], [10], [17], [19].

The LIDC-IDRI (Lung Image Database Consortium and Image Database Resource Initiative) dataset has become the most widely used benchmark for lung cancer detection research [5], [8], [9], [17]. The dataset comprises low-dose lung CT scans with detailed annotations. Nasrullah et al. [9] utilized 1,200 CT images from LIDC-IDRI containing 3,250 nodules equally split between benign and malignant cases. Jaincy et al. [17] reported that LIDC-IDRI is a global resource including 1,018 samples, with each subject providing images from an XML file and a clinical thoracic CT scan. The dataset's comprehensive annotations and large sample size make it ideal for training deep learning models and benchmarking performance [5], [8], [9], [17].

The LUNA16 (LUNG Nodule Analysis 2016) dataset, derived from LIDC-IDRI, contains 888 CT scans and has been used for nodule detection tasks [10]. Zhu et al. [10] utilized both LUNA16 and LIDC-IDRI datasets, with LUNA16 containing 888 CTs and LIDC-IDRI having 1,018 CTs with detailed annotations.

The IQ-OTH/NCCD dataset from the Iraq-Oncology Teaching Hospital/National Center for Cancer Diseases has been used in several studies [4], [19]. S. [19] reported that this dataset comprises CT scans from 110 patients: 40 with malignant tumors, 15 with benign tumors, and 55 healthy subjects, totaling 1,097 CT images (120 benign, 561 malignant, 416 normal) with 1mm slice thickness in DICOM format.

The National Lung Cancer Screening Trial (NLST) and Early Lung Cancer Action Program (ELCAP) datasets have also been utilized [12]. Liu et al. [12] combined CT scans from these sources, creating a dataset of 326 nodules (163 malignant, 163 benign) with pathologically confirmed malignancy status, balanced in size and class distribution.

Several studies utilized Kaggle datasets for lung cancer detection [11]. Chen [16] retrospectively collected H-E-stained pathological sections from 652 patients across three tertiary hospitals, including 674 pulmonary nodules (278 malignant, 396 benign). Some studies also employed private institutional datasets, such as Prashanthi et al. [15] who used data from the Barnard Institute of Radiology at Madras Medical College, Chennai [15].

### **4.2 Imaging Modalities**

Computed Tomography (CT) scans have emerged as the predominant imaging modality for lung cancer detection in the reviewed literature, owing to their superior image quality, three-dimensional information, and ability to detect small nodules [9], [10], [11], [12], [16], [17].

Multiple studies specifically focused on CT imaging for nodule detection and classification [9], [10], [12], [16]. Nasrullah et al. [9] and Zhu et al. [10] both utilized 3D CT

scans from LIDC-IDRI and LUNA16 datasets. Chen [16] used chest CT imaging from 652 patients, while Liu et al. [12] employed CT scans from NLST and ELCAP. The preference for CT imaging stems from its better quality and less noise compared to X-ray and MRI [11].

Chest X-rays have also been utilized, particularly for initial screening and in resource-constrained settings [1], [7]. Middha et al. [1] developed a CNN-based model specifically for analyzing chest radiographs. Wandhare [7] utilized the Japanese Society of Radiological Technology (JSRT) dataset consisting of 200 chest X-rays (125 abnormal, 81 normal) with  $1024 \times 1024$  pixel resolution and 12-bit color depth.

Some studies explored multimodal approaches integrating different imaging types. Javed et al. [3] reviewed deep learning techniques across X-ray, CT, MRI, and Whole Slide Imaging (WSI) for lung cancer detection and classification, highlighting the potential of combining multiple modalities for comprehensive diagnosis.

CT-guided biopsy images represent a specialized imaging modality used in interventional procedures [15]. Prashanthi et al. [15] employed deep learning for lung nodule detection in CT-guided biopsy images, achieving 99.3% accuracy in malignancy detection.

### 4.3 Data Preprocessing and Augmentation

Data preprocessing and augmentation techniques play critical roles in improving model performance and generalization [7], [10], [12], [17].

Preprocessing techniques commonly employed include:

- Normalization and intensity clipping: Zhu et al. [10] clipped raw CT data to  $[-1200, 600]$  Hounsfield units and applied linear transformation to  $[0, 1]$  for standardization.
- Noise reduction: Wandhare [7] and Jayanthi et al. [8] employed Wiener filters, median filters, and Gaussian filters to remove noise and enhance image quality.
- Segmentation: Multiple studies performed lung region segmentation to isolate regions of interest [8], [17]. Jayanthi et al. [8] used blob detection for lung region segmentation, while Jaincy et al. [17] employed U-Net and SegNet for nodule segmentation.
- Resizing and resampling: Liu et al. [12] resampled nodule CT volumes to isotropic  $16 \times 16 \times 16$  or  $32 \times 32 \times 32$  pixels. S. [19] resized images to  $224 \times 224 \times 3$  pixels for VGG16 compatibility.
- Morphological operations: Jayanthi et al. [8] applied morphological operations alongside Wiener filtering for preprocessing.

Data augmentation strategies to address limited dataset sizes and improve model robustness include:

- Geometric transformations: Wandhare [7] employed rotation, flipping, and scaling. Zhu et al. [10] used flipping and random scaling for detection, and padding, cropping, and flipping for classification.

- Orientation augmentation: Liu et al. [12] rotated nodule volumes into eight orientations to prevent overfitting.
- Histogram equalization: Jaincy et al. [17] applied histogram equalization as part of their augmentation strategy.

Chen [16] employed image binarization and filtering techniques including box, mean, and Gaussian filters. The study emphasized that proper preprocessing significantly impacts model performance, particularly in distinguishing between benign and malignant nodules [16].

## 5 PERFORMANCE METRICS AND COMPARATIVE ANALYSIS

### 5.1 Accuracy and Classification Performance

Classification accuracy represents the most commonly reported metric across the reviewed studies, with values ranging from 71% to 99% depending on the dataset, methodology, and task complexity [4], [9], [11], [15], [16], [17], [19].

High-performing models ( $\geq 95\%$  accuracy): Several studies achieved exceptional accuracy levels. S. [19] reported the highest accuracy of 99.09% using the ExtRanFS framework with VGG16+MLP on the IQ-OTH/NCCD dataset, along with sensitivity of 98.33% and F1-Score of 98.33%. Prashanthi et al. [15] achieved 99.3% accuracy in malignancy detection using a custom CNN model on CT-guided biopsy images. Rai et al. [4] reported that ResNet152 achieved 97.3% accuracy on the IQ-OTH/NCCD dataset, while DenseNet121 demonstrated the best overall balance with precision up to 92%, recall up to 90%, and F1-score up to 91%.

Strong-performing models (90-95% accuracy): Multiple studies reported accuracy in the 90-95% range. Oncu [6] achieved 91% weighted average accuracy with the CNN component of a multimodal framework on 900 CT images. Nasrullah et al. [9] reported 94% sensitivity and 90% specificity with their 3D Faster R-CNN and U-Net framework. Zhu et al. [10] achieved 90.44% accuracy for nodule classification using DeepLung's 3D DPN combined with nodule size and raw pixels. Jaincy et al. [17] reported that CNN classifiers achieved 98.19% accuracy, significantly outperforming SVM (90.1% accuracy).

Moderate-performing models (80-90% accuracy): Chen [16] achieved 92.3% accuracy for predicting malignant pulmonary nodules and 82.8% for benign ones using an improved 3D U-net system. Jayanthi et al. [8] reported 98.39% Pixel Accuracy with HLNDNet, though mean Class Accuracy was 90.42%.

Comparative performance: Studies consistently demonstrated that deep learning approaches, particularly CNNs, outperformed traditional machine learning methods. Jaincy et al. [17] showed CNN achieving 98.19% accuracy compared to SVM's 90.1% and lower ANN performance. Liu et al. [12] reported that ensemble models with 3D CNN achieved AUC of  $0.780 \pm 0.063$ , outperforming the best traditional machine learning ensemble (AUC 0.748) and single traditional model (AUC 0.708).

## 5.2 Sensitivity and Specificity

Sensitivity (recall or true positive rate) and specificity (true negative rate) are critical metrics for clinical applications, balancing the detection of actual cancer cases against minimizing false alarms [9], [16], [17], [19].

Sensitivity performance: Nasrullah et al. [9] achieved 94% sensitivity with their deep learning framework on LIDC-IDRI, demonstrating strong capability in detecting actual cancer cases. Zhu et al. [10] reported a 94.6% recall rate for all nodules with DeepLung's 3D Faster R-CNN. S. [19] achieved 98.33% sensitivity with the ExtRanFS framework. Chen [16] reported 95.51% sensitivity for benign/malignant nodule classification. Jaincy et al. [17] noted that one CNN approach achieved 92.7% sensitivity with one false positive and 94.2% with two false positives.

Specificity performance: Nasrullah et al. [9] achieved 90% specificity, indicating strong performance in correctly identifying non-cancer cases. Jaincy et al. [17] reported that SVM achieved 91.7% specificity alongside 90.1% accuracy and 84.1% sensitivity. However, Chen [16] noted a trade-off, with the improved model achieving high sensitivity (95.51%) but lower specificity (34.46%), suggesting a bias toward detecting positive cases at the cost of more false positives.

Balanced performance: Rai et al. [4] emphasized that DenseNet121 demonstrated the best overall balance among transfer learning models, with precision up to 92%, recall up to 90%, and F1-score up to 91%. This balanced performance is particularly important for clinical deployment where both false negatives (missed cancers) and false positives (unnecessary interventions) carry significant consequences [4].

The trade-off between sensitivity and specificity remains a key consideration. Chen [16] compared the improved 3D U-net system (sensitivity 95.51%, specificity 34.46%) with radiologists (sensitivity lower, but specificity 70.15%), highlighting that while AI systems may achieve higher sensitivity, human experts often maintain better specificity [16].

## 5.3 False Positive Reduction

False positive reduction is a critical challenge in lung cancer detection, as high false positive rates lead to unnecessary biopsies, patient anxiety, and increased healthcare costs [2], [9], [10].

Zhu et al. [10] achieved remarkable false positive reduction with DeepLung, reducing false positive detected nodules by 97.02%. This dramatic reduction was accomplished through the system's 3D Faster R-CNN architecture combined with the dual path network for classification. The study demonstrated that sophisticated deep learning architectures can effectively distinguish between true nodules and benign findings that might be flagged by simpler detection systems [10].

Mounika et al. [2] emphasized that their integrated CNN and virtual reality approach specifically aimed to minimize false positives while enhancing diagnostic efficiency. The immersive visualization environment was designed to help radiologists better understand

spatial distribution and characteristics of detected tumors, potentially reducing misclassification [2].

Jayanthi et al. [8] developed HLNDNet specifically to reduce false positives through a hybrid approach combining VGG-16, AlexNet, and GoogleNet with a Weighted Voting Approach. The ensemble methodology aimed to leverage the strengths of multiple architectures to filter out false detections [8].

Several studies incorporated data augmentation and preprocessing techniques specifically to reduce false positive outcomes [18]. Madasu et al. [18] noted that segmentation techniques to detect multiple lung nodules and data augmentation were utilized to reduce false positive outcomes and enhance accuracy and reliability.

The Area Under the Receiver Operating Curve (AUC-ROC) provides a comprehensive measure of model performance across different threshold settings, balancing true positive and false positive rates. Nasrullah et al. [9] achieved an AUC of 0.99, indicating excellent discrimination capability. S. [19] reported perfect ROC-AUC values of 1.00 for benign, malignant, and normal cases with VGG16+MLP. Liu et al. [12] achieved AUC of  $0.780 \pm 0.063$  with ensemble models, while Chen [16] reported AUC of 0.729 for the improved 3D U-net system compared to 0.583 for the original system [9], [12], [16], [19].

## 6 DISCUSSION

### 6.1 Key Trends and Convergence

The reviewed literature reveals several converging trends in deep learning and machine learning for lung cancer detection during 2019-2023. First, CNNs have emerged as the dominant architecture, with studies consistently demonstrating their superiority over traditional machine learning methods such as SVM and random forests [3], [17]. The multi-layer structure, automatic feature learning, and ability to capture spatial hierarchies make CNNs particularly well-suited for medical image analysis [1], [2], [3].

Second, there is a clear shift toward 3D architectures that can process volumetric CT data directly rather than analyzing individual slices [9], [10], [12], [16]. The DeepLung system [10] and similar 3D approaches [9], [12] demonstrate that leveraging spatial information across multiple slices significantly improves detection and classification performance. This trend reflects the recognition that lung nodules are three-dimensional structures whose characteristics are best captured through volumetric analysis.

Third, transfer learning and pre-trained models have become standard practice, particularly when dealing with limited medical imaging datasets [4], [19]. The success of models like VGG16, ResNet152, and DenseNet121 in achieving high accuracy (97-99%) demonstrates that knowledge learned from large-scale natural image datasets can be effectively transferred to medical imaging tasks [4], [19].

Fourth, hybrid and ensemble approaches that combine multiple models or integrate different types of features consistently outperform single-model approaches [5], [6], [8], [12]. The integration of deep learning with traditional radiomics features [5], the combination of

imaging and clinical data [6], and the fusion of multiple CNN architectures [8] all demonstrate superior performance compared to standalone models.

Fifth, there is growing emphasis on explainability and interpretability through techniques like SHAP and Grad-CAM [4], [6]. This trend addresses a critical barrier to clinical adoption, as healthcare providers require understanding of how AI systems arrive at their decisions [4].

Finally, the LIDC-IDRI dataset has become the de facto standard for benchmarking, enabling meaningful comparisons across studies [5], [8], [9], [17]. This standardization facilitates progress by allowing researchers to build upon previous work and identify genuine improvements.

## 6.2 Limitations and Challenges

Despite impressive performance metrics, several limitations and challenges persist in the field. Dataset limitations represent a significant concern. While LIDC-IDRI is widely used, its limited size (approximately 1,000 CT scans) and potential biases may limit generalizability to diverse populations and clinical settings [5], [17]. Several studies noted the lack of specific dataset details or used relatively small private datasets [1], [2], [15], raising questions about reproducibility and external validity.

Class imbalance remains a persistent challenge, with several studies explicitly addressing this issue through cost-sensitive learning or specialized sampling techniques [4]. The imbalance between benign and malignant cases, as well as between nodule and non-nodule regions, can bias models toward majority classes and reduce sensitivity for rare but clinically important cases [4].

Interpretability and explainability continue to be barriers to clinical adoption, despite recent advances [4], [6]. While techniques like SHAP and Grad-CAM provide some insight into model decisions, many studies still employ "black box" models without adequate explanation mechanisms [1], [2], [9]. Healthcare providers require not just accurate predictions but also understandable reasoning to trust and effectively use AI systems.

Performance variability across different datasets and settings is evident. Models achieving 99% accuracy on one dataset may perform significantly worse on others, suggesting potential overfitting or lack of robustness [15], [19]. The trade-off between sensitivity and specificity also varies considerably, with some models achieving high sensitivity at the cost of specificity [16].

Computational requirements for 3D deep learning models can be substantial, potentially limiting deployment in resource-constrained clinical settings [9], [10]. While these models achieve superior performance, their practical utility depends on availability of adequate computational infrastructure.

Limited discussion of clinical deployment is notable across most studies. Few papers address practical considerations such as integration with existing clinical workflows, real-time

processing requirements, regulatory approval pathways, or cost-effectiveness [1], [2], [9]. This gap between research performance and clinical implementation remains a significant challenge.

Lack of prospective validation is evident, with most studies relying on retrospective analysis of existing datasets [9], [12], [16], [17]. Prospective clinical trials are needed to validate performance in real-world settings and assess impact on patient outcomes.

### **6.3 Clinical Implications**

The advances in deep learning for lung cancer detection have significant clinical implications. Early detection capabilities demonstrated by high-sensitivity models (94-98%) could enable identification of lung cancer at earlier, more treatable stages, potentially improving survival rates [9], [10], [19]. The ability to detect subtle patterns that may be missed by human observers represents a valuable complement to radiologist expertise [1].

Reduction in false positives achieved by advanced models like DeepLung (97.02% reduction) could substantially decrease unnecessary biopsies, patient anxiety, and healthcare costs [10]. This improvement addresses one of the major challenges in lung cancer screening programs, where high false positive rates have limited adoption and cost-effectiveness.

Diagnostic efficiency could be enhanced through AI-assisted reading, potentially reducing radiologist workload and enabling faster turnaround times [2], [14]. In settings with limited radiologist availability, AI systems could serve as valuable decision support tools or even primary screening mechanisms.

Personalized treatment planning may benefit from accurate early classification of nodule characteristics and cancer subtypes [3], [15]. The ability to distinguish between benign and malignant nodules, as well as to identify specific cancer types, could inform treatment decisions and improve outcomes.

However, realizing these clinical benefits requires addressing the limitations discussed above, particularly regarding interpretability, prospective validation, and practical deployment considerations. The comparison of DeepLung's patient-level diagnosis accuracy (81.41%) with experienced doctors (average 82.31%) suggests that current AI systems are approaching human-level performance but have not yet consistently surpassed it [10]. The optimal approach may involve human-AI collaboration rather than full automation.

## **7 FUTURE DIRECTIONS AND RECOMMENDATIONS**

Based on the comprehensive analysis of the literature, several key directions emerge for future research and development in deep learning for lung cancer detection.

**Dataset diversity and standardization:** Future work should prioritize developing larger, more diverse datasets that represent different populations, imaging protocols, and clinical settings [5]. Multi-institutional collaborations and data sharing initiatives could address current limitations in dataset size and diversity. Standardized preprocessing pipelines and evaluation protocols would facilitate meaningful comparisons across studies.

**Multimodal integration:** The success of multimodal approaches combining imaging and clinical data [6] suggests that future systems should integrate multiple information sources, including CT scans, PET imaging, clinical history, genetic markers, and biomarkers. Such comprehensive approaches could improve accuracy and provide more holistic patient assessment.

**Explainable AI advancement:** Continued development of interpretability techniques is essential for clinical adoption [4]. Future research should focus on creating inherently interpretable models or developing more sophisticated explanation mechanisms that provide clinically meaningful insights into model decisions. Integration of domain knowledge and collaboration with radiologists in developing explanation frameworks is crucial.

**Prospective clinical validation:** Rigorous prospective clinical trials are needed to validate AI systems in real-world settings and assess their impact on patient outcomes, workflow efficiency, and cost-effectiveness [10]. Such studies should evaluate not just technical performance but also clinical utility, user acceptance, and integration challenges.

**Federated learning and privacy-preserving techniques:** To address data sharing limitations while maintaining patient privacy, federated learning approaches that enable model training across multiple institutions without centralizing data should be explored. Privacy-preserving techniques could facilitate larger-scale collaborations and more robust model development.

**Real-time and edge deployment:** Research should address computational efficiency to enable real-time processing and deployment on edge devices in clinical settings [9], [10]. Model compression, quantization, and efficient architecture design could make advanced AI systems accessible in resource-constrained environments.

**Longitudinal analysis:** Future systems should incorporate temporal information from serial imaging studies to track nodule growth and changes over time, potentially improving early detection and reducing false positives [10].

**Attention to rare subtypes and edge cases:** While current models perform well on common cases, future work should specifically address performance on rare cancer subtypes, unusual presentations, and challenging cases where human experts also struggle [3].

**Human-AI collaboration frameworks:** Rather than pursuing full automation, research should explore optimal human-AI collaboration models that leverage the complementary strengths of AI systems (consistency, pattern recognition) and human experts (contextual understanding, clinical judgment) [10].

**Regulatory and ethical considerations:** Future work should proactively address regulatory requirements, ethical considerations, and fairness concerns to facilitate responsible deployment of AI systems in clinical practice [4].

## 8 CONCLUSION

This comprehensive literature review of deep learning and machine learning for lung cancer detection during 2019-2023 reveals substantial progress in the field, with CNN-based approaches, particularly 3D architectures and hybrid models, achieving impressive performance metrics ranging from 90% to 99% accuracy on benchmark datasets. The convergence toward transfer learning, ensemble methods, and attention mechanisms demonstrates the maturation of the field and the identification of effective strategies for medical image analysis.

Key findings include: (1) CNNs consistently outperform traditional machine learning methods, with 3D architectures showing particular promise for volumetric CT analysis; (2) the LIDC-IDRI dataset has emerged as the standard benchmark, facilitating meaningful comparisons across studies; (3) transfer learning from pre-trained models effectively addresses limited medical data availability; (4) hybrid approaches combining multiple models or integrating imaging with clinical data achieve superior performance; and (5) explainable AI techniques are increasingly recognized as essential for clinical adoption.

However, significant challenges remain, including dataset limitations, class imbalance, interpretability concerns, performance variability, and the gap between research achievements and clinical deployment. The field would benefit from larger and more diverse datasets, prospective clinical validation, enhanced explainability, and attention to practical deployment considerations.

The clinical implications are substantial, with potential for earlier detection, reduced false positives, improved diagnostic efficiency, and personalized treatment planning. However, realizing these benefits requires continued research addressing current limitations and focusing on human-AI collaboration rather than full automation.

As the field advances, the integration of multimodal data, development of more interpretable models, rigorous prospective validation, and attention to regulatory and ethical considerations will be crucial for translating research advances into clinical impact. The reviewed literature demonstrates that deep learning has tremendous potential to transform lung cancer detection, but careful, systematic work remains necessary to achieve safe, effective, and equitable clinical deployment.

## REFERENCES

- [1] A. Middha et al., "Advancements in Early Lung Cancer Detection Using Convolutional Neural Networks," in *Proc. ICAIHI*, 2023. DOI: [10.1109/icaihi57871.2023.10489806](https://doi.org/10.1109/icaihi57871.2023.10489806)
- [2] N. Nasrullah et al., "Automated detection and classification for early stage lung cancer on CT images using deep learning," in *Proc. SPIE Medical Imaging*, 2019. DOI: [10.1117/12.2520333](https://doi.org/10.1117/12.2520333)
- [3] W. Zhu et al., "DeepLung: 3D Deep Convolutional Nets for Automated Pulmonary Nodule Detection and Classification," *bioRxiv*, 2017. DOI: [10.1101/189928](https://doi.org/10.1101/189928)

- [4] "Implementation of Machine Learning Techniques for detection of Lung Cancer," *Periodico*, 2022. DOI: [10.37896/pd91.3/91323](https://doi.org/10.37896/pd91.3/91323)
- [5] K. Liu et al., "Pulmonary nodule classification in lung cancer screening with three-dimensional convolutional neural networks," *Journal of Medical Imaging*, vol. 4, no. 4, 2017. DOI: [10.1117/1.JMI.4.4.041308](https://doi.org/10.1117/1.JMI.4.4.041308)
- [6] Y. Chen, "Models of Artificial Intelligence-Assisted Diagnosis of Lung Cancer Pathology Based on Deep Learning Algorithms," *Journal of Healthcare Engineering*, 2022. DOI: [10.1155/2022/3972298](https://doi.org/10.1155/2022/3972298)
- [7] J. Jaincy et al., "A Study on Lung Cancer Detection in CT Images Using Medical Image processing with Deep Learning Techniques," *Advances in Computer Science Research*, 2023. DOI: [10.2991/978-94-6463-250-7\\_3](https://doi.org/10.2991/978-94-6463-250-7_3)
- [8] S., "ExtRanFS: An Automated Lung Cancer Malignancy Detection System Using Extremely Randomized Feature Selector," *Diagnostics*, vol. 13, no. 13, 2023. DOI: [10.3390/diagnostics13132206](https://doi.org/10.3390/diagnostics13132206)
- [9] S. Nandula et al., "Lung Cancer Detection by Harnessing the Power of Deep Learning with Convolutional Neural Networks," in *Proc. ICECAA*, 2023. DOI: [10.1109/icecaa58104.2023.10212364](https://doi.org/10.1109/icecaa58104.2023.10212364)