A Systematic Review of Deep Learning Architectures for Lung Cancer Detection and Classification

Mudit Mittal[1], Prabhakar Semwal[2], Pallaw Singh Aswal[3]

[1,2,3]Assistant Professor, Department of Computer Science Engineering,
THDC Institute of Hydropower Engineering and Technology, New Tehri, Uttarakhand, India

DOI: 10.64823/ijter.2606012

Abstract: Lung cancer is the leading cause of cancer-related mortality worldwide, responsible for approximately 18% of all cancer deaths globally. The absence of early clinical symptoms significantly delays diagnosis, making timely and accurate automated detection systems a critical public health necessity. The rapid growth of deep learning technologies has opened transformative opportunities for automated lung cancer detection and classification from medical imaging data, particularly CT scans. However, despite impressive benchmark results, numerous methodological challenges and research gaps persist that hinder the transition of these models from laboratory settings to clinical practice.

This paper presents a comprehensive systematic review of six representative deep learning-based studies for lung cancer detection and classification, published between 2018 and 2025. The reviewed works include four original research papers — spanning hybrid CNN-SVM models, transfer learning frameworks, CNN-LSTM architectures, and transformer-based segmentation systems — alongside two systematic review and survey papers that collectively examine over 30 models from the literature. A structured comparative analysis is conducted across eight key dimensions: paper type, datasets used, model architectures, classification task complexity, imaging modality, best reported accuracy, and other quantitative performance metrics including AUC, sensitivity, specificity, F1-score, and Dice coefficient.

The critical analysis of each paper reveals recurring limitations across the field, including over-reliance on single benchmark datasets, restriction to binary classification tasks, absence of integrated segmentation and multi-class classification pipelines, minimal adoption of transformer-based models for classification, near-complete lack of model explainability mechanisms, and a universal absence of prospective clinical validation. Extending this analysis across all six reviewed papers, nine significant and cross-validated research gaps are systematically identified and documented. These gaps collectively define the design requirements for a next-generation lung cancer detection model: one that integrates 3D segmentation with multi-class subtype classification, employs transformer-based attention mechanisms, incorporates model interpretability, supports Low-Dose CT inputs, and is validated across diverse datasets and clinical environments.

The findings of this review establish a rigorous evidence-based foundation for the development of a novel deep learning and image processing model for lung cancer detection, contributing to the ongoing effort to bridge the gap between computational performance and real-world clinical applicability.

Keywords: Convolutional Neural Networks, deep learning, lung cancer detection, medical image analysis, CT scan, systematic review, image segmentation, transfer learning, transformer models, self-supervised learning, computer-aided diagnosis, LIDC-IDRI.

I. INTRODUCTION

Lung cancer considered as one of the most devastating and life-threatening diseases globally, consistently ranking as the leading cause of cancer-related mortality across both developed and developing nations. According to the Global Cancer Statistics (GLOBOCAN 2020), near about 193 lakh population were recorded as a new cases of cancer across the globe, of which lung cancer accounted for approximately 11.4% of new cases and 18% of all cancer-related deaths — surpassing all other cancer types in mortality rate [Sung et al., 2021]. The 05 year survival rate in the patient with lung cancer is at critically low side, at approximately 10–20%, primarily because lung cancer is most frequently detected at advanced stages. And then curative intervention is almost no longer feasible.

The primary reason behind such alarming mortality figures is the late-stage diagnosis of the disease. In the initial stage, lung cancer shows minimal or almost no visible clinical signs, which makes timely identification extremely challenging without systematic screening. If detected at an early localized stage, the survival rate for 05-years can be improved to over 60%, underscoring the clinical imperative for effective early detection systems. Early and accurate identification of this severe disease is therefore not merely a clinical priority but a pressing public health necessity, as it can significantly enhance the patient survival rates.

Traditional diagnostic approaches for lung cancer rely heavily on the manual interpretation of medical imaging data by trained radiologists, primarily through chest X-rays, computed tomography (CT) scans, magnetic resonance imaging (MRI), and positron emission tomography (PET) scans. CT imaging, in particular, has become the preferred modality for lung cancer screening due to its high sensitivity in detecting small pulmonary nodules and its ability to provide detailed cross-sectional images of the lung parenchyma. Low-Dose CT (LDCT) screening, demonstrated by The National Lung Screening Trial (NLST), reduced lung cancer mortality by 20%.

Despite these advances in imaging technology, the manual interpretation of CT scans remains inherently time-consuming, subjective, and prone to inter-observer variability. Studies have shown that even experienced radiologists can miss up to 30% of small pulmonary nodules on CT images, particularly when nodule size is below 6mm or when the nodule is located adjacent to blood vessels or the pleural wall. Under conditions of high clinical workload, diagnostic fatigue further increases the risk of missed or misclassified lesions. These limitations have motivated the research community to develop automated, intelligent diagnostic solutions capable of providing consistent, objective, and high-throughput analysis of medical imaging data.

In recent years, the emergence of deep learning as a subfield of machine learning has fundamentally transformed the landscape of medical image analysis. Deep learning models, particularly Convolutional Neural Networks (CNNs) and their increasingly sophisticated variants, possess the remarkable ability to automatically learn hierarchical feature representations directly from raw imaging data without requiring manual feature engineering. This capability makes them especially well-suited for the complex pattern recognition tasks inherent in lung cancer detection, segmentation, and classification. The availability of large annotated medical imaging datasets — most notably the Lung Image Database Consortium and Image Database Resource Initiative (LIDC-IDRI), containing CT images from over 1,000 patients with expert annotations of pulmonary nodules — has further accelerated the development and validation of deep learning approaches in this domain.

A wide range of deep learning architectures has been applied to the problem of lung cancer detection and classification, reflecting the rapid evolution of the field. Early approaches employed standard CNN

architectures such as AlexNet and VGGNet for nodule classification. Subsequent work introduced deeper and more efficient architectures including ResNet, GoogLeNet (Inception), and DenseNet, which demonstrated significant improvements in feature extraction capability. Encoder-decoder architectures such as U-Net became widely adopted for lung nodule segmentation tasks, enabling pixel-level delineation of tumour regions. More recently, transformer-based models — originally developed for natural language processing tasks — have been adapted for medical image analysis, with architectures such as UNETR (UNEtTRansformers) demonstrating state-of-the-art performance in 3D volumetric segmentation tasks by capturing global contextual information through multi-head self-attention mechanisms.

Six studies form the core of this systematic review, representing a diverse cross-section of deep learning approaches for lung cancer detection. Ansari et al. (2025) proposed the SVMVGGNet-16 model, a hybrid architecture combining VGGNet-16 for deep feature extraction with a Support Vector Machine (SVM) classifier, applied to four-class lung cancer subtype classification on the LIDC-IDRI dataset. Zhang et al. (2018) provided a comprehensive survey of deep learning-based computer-aided diagnosis approaches, covering the full pipeline from data preprocessing and lung segmentation to nodule detection and pathological diagnosis. Fei et al. (2024) introduced the RGIV3 architecture, an enhanced GoogLeNet Inception V3 model augmented with a custom feature fusion layer, specifically designed to address the source-target domain gap in transfer learning for pulmonary nodule detection. Devarajan et al. (2023) proposed a CNN-LSTM hybrid model that exploits both spatial features through CNNs and sequential inter-slice temporal dependencies through Long Short-Term Memory networks. Said et al. (2023) developed a complete end-to-end diagnosis system combining UNETR-based 3D segmentation with a self-supervised classification network, achieving state-of-the-art performance. Finally, Thanoon et al. (2023) provided the most comprehensive review of CT-based deep learning approaches for lung cancer screening, systematically evaluating over 30 models and explicitly documenting eight research challenges in the field.

Despite these considerable advances, a thorough cross-paper analysis reveals that several fundamental challenges and research gaps remain unaddressed, limiting the clinical applicability and generalizability of current models. This review systematically compares all six studies, critically evaluates their individual strengths and limitations, and identifies nine key research gaps that collectively motivate the development of a novel deep learning and image processing model for lung cancer detection that transcends the limitations of existing approaches.

The remainder of this paper is organized as follows. Section 2 presents the search methodology and inclusion criteria following the PRISMA framework. Section 3 provides the detailed comparative analysis of all six reviewed papers. Section 4 presents an in-depth critical analysis of each study individually. Section 5 documents nine systematic research gaps identified across the reviewed literature. Section 6 discusses observed trends and future research directions. Section 7 concludes the paper. A complete reference list is provided in Section 8.

II. Search Methodology

A. Literature Search Strategy

A systematic and technical literature search was performed to select peer-reviewed publications reporting the application of deep learning and image processing techniques to lung cancer detection, classification, or segmentation from medical imaging data. The search was performed over 04 major academic and scientific databases: IEEE Xplore, PubMed (MEDLINE), Scopus, and Google-Scholar. These databases were selected to ensure broad coverage of both engineering/computer science publications (IEEE Xplore, Scopus) and biomedical literature (PubMed), as well as grey literature and conference proceedings (Google Scholar).

The following primary search query terms and their Boolean combinations were used to maximize retrieval coverage:

- "deep learning" AND "lung cancer detection"
- "convolutional neural network" AND "lung nodule classification"
- "medical image segmentation" AND "lung cancer"
- "computer-aided diagnosis" AND "lung CT scan"
- "transfer learning" AND "pulmonary nodule detection"
- "hybrid deep learning" AND "lung cancer"
- "transformer" AND "medical image" AND "lung cancer"
- "deep learning" AND "lung cancer" AND "CT scan" AND "review"

Boolean operators AND and OR were systematically applied to combine primary search terms. The search was restricted to publications in the English language and covered the period from January 2017 to April 2025, to ensure both historical context and contemporary relevance.

B. Inclusion Criteria

Studies were included in this review if they satisfied all of the following criteria:

- The study proposed, evaluated, or systematically reviewed a deep learning or machine learning model specifically applied to lung cancer detection, classification, or segmentation.
- The study used medical imaging data as the primary input, including CT scans, chest X-rays, PET scans, MRI images, or histopathological images.
- The study reported quantitative performance metrics such as accuracy, sensitivity, specificity, AUC, F1-score, Dice coefficient, or precision-recall.
- The study was published in a peer-reviewed journal or reputable conference proceedings with available full text.
- The study provided sufficient methodological detail to allow meaningful comparative analysis of the approach and results.

C. Exclusion Criteria

Studies were excluded from this review if they met any of the following conditions:

- The study focused exclusively on cancer types other than lung cancer without a dedicated lung cancer component.
- The study employed only traditional machine learning methods (e.g., SVM, Random Forest, k-NN alone) without any deep learning component.
- The study did not use medical imaging as the primary data source (e.g., studies based exclusively on genomic or blood marker data).
- The study was a duplicate publication, abstract-only submission, editorial, opinion piece, or commentary without original experimental content.
- The study lacked sufficient methodological or experimental detail to permit reproducibility assessment or meaningful comparative analysis.

D. Study Selection Process — PRISMA Framework

The study selection process was conducted in accordance with the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines, which ensure transparency, reproducibility, and methodological rigour in the literature review process. The selection proceeded through four sequential, hierarchical stages:

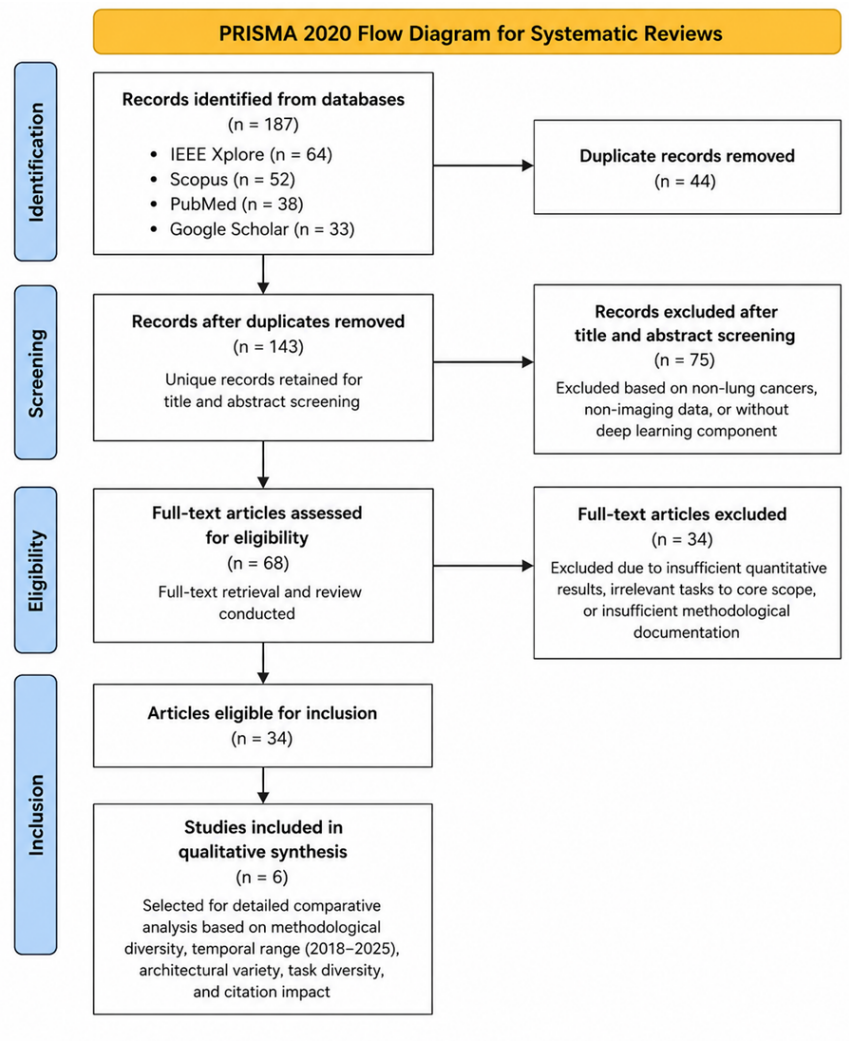

Fig 2a. PRISMA Framework

E. Data Extraction

For each of the six selected studies, the following information was systematically extracted and recorded in a structured format to enable comparative analysis:

1. Author(s), institution, and year of publication;
2. Type of paper (original research, survey, or review)
3. Dataset(s) used, including size and annotation type
4. Deep learning architecture and model type
5. Imaging modality (CT, X-ray, MRI, PET)
6. Classification task type and number of classes
7. Preprocessing and data augmentation techniques employed
8. Training, validation, and testing strategy
9. Reported performance metrics including accuracy, AUC, sensitivity, specificity, precision, recall, F1-score, and Dice coefficient where applicable
10. Stated limitations and future work recommendations.

III. Comparative Analysis of Reviewed Studies

This section presents a structured comparative analysis of all six reviewed studies. **Table 1** provides a comprehensive side-by-side comparison across eight key dimensions, followed by a detailed performance comparison in **Table 2**. The analysis reveals significant diversity in architectural approaches, task formulations, dataset choices, and evaluation protocols, making direct numerical comparison non-trivial but informative as a characterization of the current state of the field.

Table 1: Structured Comparative Analysis of Six Deep Learning Papers for Lung Cancer Detection (2018–2025)

Paper / Authors	Year	Type	Dataset(s)	Key Model(s)	Task	Best Accuracy	Key Metrics
Ansari et al. (2025) SVMVGG Net-16 Curr. Med. Imaging	2025	Research	LIDC-IDRI	VGGNet-16 + SVM (Hybrid)	4-class: ADC, LCC, SCC, Normal	96.72% (test) 97.22% (train)	AUC 96.87% Recall 84.67% Precision 87.40% F1 85.73%
Zhang et al. (2018) CAD Lung Cancer Survey J. CAD & CG	2018	Survey	LIDC-IDRI, LUNA16, JSRT	CNN, ResNet, U-Net, DBN, SAE (reviewed)	Full pipeline: pre-processing, segmentation, nodule detection, diagnosis	~95.5% (best cited model: SAE+SVM)	Sensitivity 0.907 AUC 0.987
Fei et al. (2024) DL Lung Image Recognition IJIRCST	2024	Research	LUNA16 (888 patients)	GoogleNet InceptionV3 + Feature Fusion Layer (RGIV3)	Binary: nodule vs. normal	88.78%	Sensitivity 87.18% Specificity 80.6%
Devarajan et al. (2023) DL Automated Detection ICAAIHI	2023	Research	Proprietary CT dataset (undisclosed)	CNN-LSTM Hybrid	Binary: cancer vs. non-cancer	87.00%	Precision 88% Recall 84% F1 86% AUC-ROC 0.92
Said et al. (2023) Medical Image Segmentati	2023	Research	Decathlon (96 CT scans) + LIDC-IDRI	UNETR (Transformer + U-Net) + Self-	Segmentation + binary classification (benign/malignant)	98.77% (classif.) 97.83% (seg.)	Dice 96.42% Sensitivity 96.85% Specifici

Paper / Authors	Year	Type	Dataset(s)	Key Model(s)	Task	Best Accuracy	Key Metrics
on Diagnostics				Supervised Classifier			ty 97.12%
Thanoon et al. (2023) DL Review for Lung Cancer Diagnostics	2023	Review	LIDC-IDRI, LUNA16, NLST, Tianchi AI, CIA, ImageNet + others	CNN, 3D CNN, U-Net, ResNet, Mask R-CNN, DenseNet, DCNN (reviewed — 30+ models)	Comprehensive review: classification + segmentation (CT-based)	99.45% (best cited: CNN-based model)	Sensitivity up to 98.3% Dice 0.9502 Seg. acc. 97.68%

* Papers marked as Survey or Review report performance figures from cited models rather than original experiments. Direct accuracy comparison with Research papers is methodologically approximate.

A. Performance Summary

Following table consolidates the primary performance metrics reported across all six papers to facilitate direct comparison. Where papers are surveys or reviews, the best-cited model's performance is recorded.

Table 2: Performance Metrics Comparison across All Six Reviewed Papers

Paper	Modality	Model	Accuracy	Sensitivity	Specificity	Dice	Key Note
Said et al. (2023)	CT (3D)	UNETR+SS	98.77%	96.85%	97.12%	96.42 %	Highest classif. acc. Small test set (32 scans)
Ansari et al. (2025)	CT	SVMVGGNet-16	96.72%	84.67%	87.40%	—	Only 4-class subtype paper. Mild overfitting gap.
Fei et al. (2024)	CT	RGIV3	88.78%	87.18%	—	—	Low specificity (80.6%). Binary only.
Devarajan et al. (2023)	CT	CNN-LSTM	87.00%	84.00%	—	—	Proprietary dataset. Lowest accuracy.

Paper	Modality	Model	Accuracy	Sensitivity	Specificity	Dice	Key Note
Zhang et al. (2018)*	CT/X-Ray	Various (survey)	~95.5%	0.907	—	—	Survey. Best cited model. 2018 coverage.
Thanoon et al. (2023)*	CT	Various (review)	99.45%	98.30%	—	0.9502	Review. Best cited CNN model. No original results.

B. Dataset Coverage

The LIDC-IDRI dataset is the most consistently used across the reviewed papers, appearing in four of the six studies. It remains the gold standard for lung nodule research, comprising CT images from 1,010 patients with expert annotations from multiple radiologists. The LUNA16 dataset, a curated subset of LIDC-IDRI focusing on nodules 3mm or larger with consensus annotations, is used by two papers. Said et al. (2023) uniquely employ the Medical Segmentation Decathlon dataset, which provides 3D volumetric CT scans, making it particularly suitable for transformer-based 3D architectures. Devarajan et al. (2023) use an undisclosed proprietary dataset, which is a significant methodological concern with respect to reproducibility and external validation. Thanoon et al. (2023) cover the widest dataset range across their reviewed literature, including NLST, Tianchi AI, Cancer Imaging Archive (CIA), ImageNet (for pre-training), and several private clinical datasets.

A notable finding is that no paper in the reviewed set validates its proposed model across more than one dataset simultaneously. This single-dataset dependency is a universal limitation that raises legitimate questions about model generalizability to different scanner types, imaging protocols, and patient demographics.

C. Architecture Evolution across Papers

The reviewed papers collectively trace the evolution of deep learning architectures for lung cancer detection across an eight-year period. Zhang et al. (2018) document the landscape prior to the widespread adoption of deep networks, when CNN variants such as AlexNet, GoogLeNet, ResNet, and deep belief networks represented the frontier. Fei et al. (2024) build on the GoogLeNet Inception V3 architecture with a custom feature fusion layer, demonstrating that targeted modifications to established transfer learning architectures can yield meaningful performance gains. Ansari et al. (2025) combine CNN feature extraction with SVM classification in a hybrid framework, reflecting growing interest in integrating traditional machine learning strengths with deep feature learning. Devarajan et al. (2023) advance to hybrid recurrent-convolutional architectures, exploiting the sequential nature of CT slice stacks through LSTM layers. Said et al. (2023) represent the current architectural frontier, employing UNETR — a transformer-based encoder fused with a U-Net decoder — for 3D volumetric segmentation followed by self-supervised classification. Thanoon et al. (2023) synthesize developments up to 2023, covering architectures from simple 2D CNNs to 3D CNNs, Mask R-CNN, DenseNet, and residual architectures, while notably not yet covering transformer-based classifiers.

D. Task Complexity Comparison

The six reviewed papers span a spectrum of task complexity. The simplest formulation is binary classification — cancer vs. non-cancer or nodule vs. normal tissue — adopted by Fei et al. (2024) and Devarajan et al. (2023). Said et al. (2023) advance to a two-stage task combining binary-class 3D segmentation with binary classification (benign vs. malignant). Ansari et al. (2025) address the most clinically informative task: four-

class subtype classification distinguishing Adenocarcinoma (ADC), Large Cell Carcinoma (LCC), Squamous Cell Carcinoma (SCC), and Normal tissue. Both Zhang et al. (2018) and Thanoon et al. (2023), as review/survey papers, cover the full task spectrum across their surveyed literature, from candidate detection through segmentation to malignancy classification and staging. The disparity in task complexity across papers makes direct accuracy comparison unreliable — a model classifying into four subtypes faces a fundamentally harder problem than one performing binary detection.

IV. Critical Analysis of Each Study

A. Ansari et al. (2025) — SVMVGGNet-16

- **Study Overview:** Ansari et al. (2025) propose the SVMVGGNet-16 model, a hybrid deep learning and machine learning framework that integrates VGGNet-16 — a 16-layer deep CNN pre-trained on ImageNet — with a Support Vector Machine (SVM) classifier. The model is applied to the task of four-class lung cancer subtype classification on the LIDC-IDRI dataset. The preprocessing pipeline employs median filtering for noise reduction, histogram equalization for contrast enhancement, and Canny edge detection for tumour boundary delineation. Geometric features (area, perimeter, eccentricity, compactness, circularity) are extracted from segmented tumour regions and used as SVM inputs, while VGGNet-16 handles hierarchical feature extraction from the CT image patches.

- **Strengths:** This is the only paper in the reviewed set that addresses four-class lung cancer subtype classification, distinguishing between ADC, LCC, SCC, and Normal tissue simultaneously. This multi-class formulation is of high clinical significance, as different subtypes require fundamentally different treatment protocols. The hybrid architecture exploits the complementary strengths of CNNs (rich spatial feature extraction) and SVMs (robust high-dimensional classification), addressing the overfitting tendencies of deep networks on small medical datasets. The preprocessing pipeline is well-documented and computationally appropriate for clinical CT data. Performance is comprehensively evaluated using five metrics (accuracy, AUC, recall, precision, F1-score) across training, validation, and testing phases, and a confusion matrix and ROC curves are provided per class. The model achieves strong testing accuracy (96.72%) and AUC (96.87%), establishing a competitive baseline for four-class classification.

- **Weaknesses:** The gap between training accuracy (97.22%) and testing accuracy (96.72%), while modest, indicates mild overfitting, particularly notable for the ADC class which shows the highest misclassification rate in the confusion matrix. The study relies exclusively on the LIDC-IDRI dataset with no cross-dataset validation, limiting confidence in generalizability to different scanner models or imaging centres. The comparative benchmarking is restricted to only three prior works (ResNet 2017, Deep Autoencoder 2019, MV-KBC 2022), providing an incomplete view of the model's position in the broader literature. No tumour segmentation is performed prior to classification, which may mean the model is partially learning from surrounding tissue context rather than tumour-intrinsic features. Model interpretability mechanisms, such as Grad-CAM or attention visualization, are entirely absent, making clinical trust and validation difficult. The dataset split (70/20/10) is not cross-validated, leaving uncertainty about performance stability across different data partitions.

B. Zhang et al. (2018) — Computer-Aided Lung Cancer Diagnosis Survey

- **Study Overview:** Zhang et al. (2018) provide a comprehensive survey of deep learning-based computer-aided lung cancer diagnosis methods, covering the complete clinical pipeline from CT image preprocessing through lung parenchyma segmentation, nodule candidate detection and segmentation, to pathological diagnosis and malignancy classification. The survey is organized around four processing stages and systematically reviews traditional methods alongside deep learning approaches published through 2017–2018, including CNNs, ResNets, deep belief networks (DBNs), stacked autoencoders (SAEs), and recurrent architectures. The paper covers competitive methods from the LUNA16 challenge

and Kaggle Data Science Bowl 2017, providing performance data across sensitivity, specificity, accuracy, and AUC metrics.

- **Strengths:** This is the most pipeline-complete survey among the reviewed papers, offering a structured overview of the entire lung cancer detection workflow. The paper documents the key architectural advantage of 3D CNNs over 2D approaches in capturing inter-slice spatial context, a finding directly validated by Said et al. (2023) in the same review set. It covers transfer learning strategies, curriculum learning, and ensemble methods systematically, providing methodological breadth not found in any individual research paper. The survey presents performance data across multiple architectures and datasets, offering a useful comparative baseline. It identifies early formulations of several key research challenges — particularly regarding dataset size, overfitting, and clinical applicability — that remain open in 2025.

- **Weaknesses:** Being a 2018 survey, it predates the emergence of transformer-based architectures (ViT, UNETR), self-supervised learning, and sophisticated attention mechanisms — all of which have substantially advanced the field post-2021 and are already operationalized by Said et al. (2023) in the same review set. The paper presents no original experimental results, and the performance metrics reported across surveyed models are non-comparable due to differing datasets, task formulations, and evaluation protocols. The primary language of the full paper text is Chinese, which limits accessibility and reproducibility for the broader international research community. The survey does not explicitly formalize or enumerate research gaps, making it less actionable as a guide for future model development compared to Thanoon et al. (2023).

C. Fei et al. (2024) — RGIV3 Enhanced GoogLeNet Inception V3

- **Study Overview:** Fei et al. (2024) address a practical and often-overlooked challenge in medical image transfer learning: the significant domain gap between ImageNet-pretrained deep networks and the target medical imaging domain (lung CT). The authors propose RGIV3 — an enhanced GoogLeNet Inception V3 architecture augmented with a custom feature fusion layer comprising three fully connected layers and a dropout layer. The fusion layer is designed to enable non-linear fitting of network-learned features, improving the extraction of relevant diagnostic patterns from the LUNA16 dataset. The model is evaluated on binary pulmonary nodule classification (nodule vs. healthy tissue) and tested across multiple train-test split ratios to assess generalization stability.

- **Strengths:** The paper addresses a genuine and clinically important challenge — source-target domain mismatch in transfer learning — and provides a concrete, lightweight architectural solution. The fusion layer design is computationally efficient, adding minimal parameters while meaningfully improving feature extraction quality. Testing across three train-test split ratios (80/20, 70/30, 60/40) provides stronger evidence of generalization capability than single-split evaluations used by most compared papers. The model achieves meaningful improvements over baseline GoogLeNet Inception V3 (+2.7% accuracy, +2.22% sensitivity), validating the design choice. An empirical search over node count and dropout rate combinations (30 experimental groups) is used to optimize fusion layer configuration, reflecting methodological thoroughness.

- **Weaknesses:** The model is limited to binary classification of pulmonary nodules, which is clinically insufficient for treatment planning that requires knowledge of cancer subtype. The dataset size is relatively modest — 870 patients total, with 5,000 augmented images per class — limiting statistical confidence in the reported results. The specificity of 80.6% indicates a meaningful false positive rate (approximately 1 in 5 healthy tissue patches misclassified as nodules), which Thanoon et al. (2023) identify as a field-wide challenge with direct clinical consequences including unnecessary biopsies and patient anxiety. The paper does not compare against more recent architectures such as EfficientNet, Vision Transformers, or hybrid CNN-Transformer models, making it difficult to assess the model's

relative standing in the 2024 landscape. Model interpretability is entirely absent, and there is no discussion of the clinical deployment pathway.

D. Devarajan et al. (2023) — CNN-LSTM Hybrid Architecture

- **Study Overview:** Devarajan et al. (2023) propose a hybrid architecture combining a Convolutional Neural Network (CNN) for spatial feature extraction with a Long Short-Term Memory (LSTM) network for sequential inter-slice dependency analysis, applied to automated binary lung cancer detection from CT scans. The CNN component is based on a pre-trained architecture with fully connected layers removed to retain spatial feature maps, while the LSTM component models the temporal/sequential relationships between adjacent CT slices. The combined model fuses CNN and LSTM outputs through feature concatenation, followed by fully connected layers and sigmoid activation for binary classification. The model is trained and evaluated using binary cross-entropy loss and the Adam optimizer, with standard train-validation-test splits.

- **Strengths:** The integration of LSTM to exploit the sequential nature of CT slice stacks is a conceptually innovative design choice that addresses a genuine limitation of purely spatial 2D CNN approaches. Sequential CT slices contain meaningful inter-slice structural information — particularly for nodule boundary delineation and shape characterization — that standard CNNs process independently, potentially missing important 3D context. The paper is notable for being the only reviewed study to address model interpretability, providing Class Activation Maps (CAMs) and LSTM attention visualizations to help explain the model's diagnostic decisions, directly responding to the interpretability gap identified by Thanoon et al. The paper also raises ethical considerations around data privacy, potential model bias, and the clinical implications of automated diagnostic deployment — issues that are rarely discussed in engineering-focused medical AI papers.

- **Weaknesses:** The most critical limitation is the use of a proprietary and entirely undisclosed dataset, which renders the results non-reproducible and prevents independent validation. This directly contradicts the open-data recommendations made by Thanoon et al. (2023) and undermines the scientific credibility of the reported accuracy figures. The model achieves the lowest accuracy (87%) among all research papers in the reviewed set, which is particularly concerning given the binary (rather than multi-class) nature of the task. The paper does not establish through ablation studies that the LSTM component provides meaningful performance gains over a standard 3D CNN, which would capture inter-slice context natively. The task scope is limited to binary classification, and there is no discussion of cancer subtype differentiation. The description of the dataset as having 'N images' without specifying the actual count or patient demographics severely limits the paper's scientific contribution.

E. Said et al. (2023) — UNETR and Self-Supervised Classification

- **Study Overview:** Said et al. (2023) present the most architecturally advanced original research paper in the reviewed set, proposing a complete end-to-end lung cancer diagnosis system consisting of two main components. The first component is a UNETR (UNEtTRansformers) network used for 3D CT scan segmentation. UNETR employs a Vision Transformer (ViT) as the encoder, processing 3D input volumes as sequences of non-overlapping patches through multi-head self-attention layers, directly connected to a U-Net-style decoder via skip connections at multiple resolutions. The second component is a Self-Supervised Classification Network that receives the UNETR segmentation output and classifies each segmented region as either benign or malignant. The self-supervised approach eliminates the need for large labeled datasets by simultaneously learning class labels and representations in a single-stage end-to-end process, using a cross-entropy loss with uniform prior to prevent degenerate constant-prediction solutions.

- **Strengths:** Said et al. (2023) is the only paper in the reviewed set that delivers a complete end-to-end pipeline combining segmentation and classification in a single deployable system. The adoption of

UNETR represents the most architecturally sophisticated approach across all reviewed papers, exploiting transformer-based global context capture for 3D volumetric segmentation — an approach consistently recommended but not implemented in any of the review papers. The model achieves the highest classification accuracy (98.77%) and competitive segmentation performance (97.83% accuracy, 96.42% Dice) among research papers. The self-supervised learning paradigm addresses the labeled data scarcity challenge that constrains most supervised approaches. Systematic evaluation of optimizer choice (AdamW vs. Nadam) and activation function (ReLU vs. Leaky ReLU) provides methodologically rigorous ablation insights. The model outperforms all five state-of-the-art segmentation baselines on the Decathlon dataset.

- **Weaknesses:** The most significant limitation is the extremely small dataset size. The model is trained on only 64 CT scan volumes and tested on 32 — a total of 96 cases from the Decathlon dataset. The reported 98.77% classification accuracy based on 32 test cases is statistically unreliable; a difference of just one correctly/incorrectly classified scan would change accuracy by 3.1%. This severely limits confidence in the generalizability of the reported results and makes it inappropriate to claim state-of-the-art performance without validation on substantially larger datasets. The model has 93.32 million parameters and requires 10.23 seconds per scan for inference, which is computationally prohibitive for real-time or high-throughput clinical screening applications. The classification task is limited to binary benign/malignant discrimination and does not differentiate between lung cancer subtypes, which is essential for treatment planning. No model interpretability or explainability mechanisms are provided.

F. Thanoon et al. (2023) — Comprehensive Review of DL Techniques

- **Study Overview:** Thanoon et al. (2023) provide the most comprehensive systematic review of CT-based deep learning approaches for lung cancer screening and diagnosis. The review covers the full spectrum of deep learning methodologies organized into four learning paradigms (supervised, unsupervised, semi-supervised, and reinforced learning), two primary task categories (classification and segmentation), and five imaging modalities (CT, 3D CT, LDCT, MRI, X-ray). The review examines over 30 models from 2018 to 2023, provides a structured summary table of dataset and method characteristics, compares performance metrics across reviewed models, and discusses the advantages and disadvantages of each approach. The paper concludes with an explicit enumeration of eight research gaps and a comprehensive set of future research direction recommendations.

- **Strengths:** This paper offers the broadest and most systematic scope of all six reviewed studies, providing a comprehensive taxonomy of deep learning approaches in the lung cancer detection domain. It is the only paper to systematically compare five imaging modalities (X-ray, CT, MRI, PET-CT, and Ultrasound) in terms of clinical applications, advantages, and limitations, providing essential context for modality selection decisions. The review covers 30+ models across a five-year window (2018–2023), making it the most temporally comprehensive source in the reviewed set. The explicit identification of eight research challenges — covering dataset heterogeneity, small nodule detection, CNN interpretability, LDCT screening, and others — provides the most structured and actionable gap analysis of any reviewed paper. The discussion of future research directions spans a uniquely broad range, covering data standardization, cloud computing, feature fusion with clinical data, and novel optimization strategies.

- **Weaknesses:** As a review paper, Thanoon et al. (2023) presents no original experimental results. The performance comparison table (Table 3 in the paper) compares models trained on different datasets performing different tasks, making the accuracy figures across rows non-comparable without appropriate methodological caveats. The review notably does not cover transformer-based architectures — particularly Vision Transformers (ViT), UNETR, or hybrid CNN-Transformer models — despite these having been published and gaining significant traction by 2022–2023. Self-supervised learning approaches, which Said et al. (2023) demonstrate effectively in the same reviewed set, are also absent from the survey coverage. The review focuses exclusively on CT imaging, with limited substantive

discussion of practical multi-modal fusion implementation strategies, despite acknowledging PET-CT as superior for staging and metabolic characterization. The review coverage effectively ends in early 2023, missing significant architectural advances published in late 2023 and 2024.

V. Research Gaps in Existing Literature

The cross-paper comparative analysis of the six reviewed studies reveals nine persistent and critical research gaps that collectively limit the clinical applicability, generalizability, and trustworthiness of current deep learning models for lung cancer detection and classification. These gaps are not merely acknowledged in individual papers as future work but emerge as systematic, cross-validated deficiencies spanning the entire body of reviewed literature. Table 3 presents these gaps alongside supporting evidence and proposed directions for a novel model. Each gap is further elaborated in the subsections below.

Table 3: Research Gaps Identified Across All Six Reviewed Papers — Evidence and Proposed Directions

#	Research Gap	Evidence from Reviewed Literature	Proposed Direction for Novel Model
1	Dataset Diversity & Cross-Dataset Generalization	All 6 papers rely on 1–3 benchmark datasets. Thanoon et al. document 10+ datasets used across literature but no paper validates cross-dataset performance. Devarajan et al. use a fully undisclosed proprietary dataset.	Train and validate across multiple datasets simultaneously (LIDC-IDRI, LUNA16, Decathlon, NLST). Include prospective hospital imaging data to test real-world generalizability.
2	Unified Segmentation + Multi-Class Classification Pipeline	Said et al. combine segmentation + binary classification. Ansari et al. deliver 4-class classification without segmentation. Thanoon et al. call for pipeline integration. No paper delivers both tasks end-to-end.	Develop an end-to-end pipeline: 3D volumetric segmentation followed by multi-class subtype classification (ADC, LCC, SCC, Normal) within a single unified model.
3	Transformer-Based Classification	Said et al. use UNETR transformers for segmentation only. Thanoon et al. do not cover ViT or hybrid CNN-Transformer classifiers. No paper applies transformers to multi-class lung cancer subtype classification.	Apply Vision Transformers (ViT) or hybrid CNN-Transformer architectures to the multi-class classification task to exploit global context capture and long-range dependency modelling.
4	Multi-Modal Imaging Fusion	Thanoon et al. compare 5 imaging modalities but no research paper fuses more than one. All 5 research papers use CT only. PET-CT fusion is noted as beneficial but never implemented.	Integrate CT, PET, and/or MRI data using multi-modal fusion frameworks. Include complementary metabolic and structural imaging signals for improved diagnostic accuracy.
5	Model Explainability & Clinical Interpretability	Thanoon et al. identify CNN interpretability as a core challenge. Only Devarajan et al. attempt CAMs and LSTM attention visualization. The	Integrate Grad-CAM, SHAP values, or attention heatmaps into the model architecture to produce transparent, radiologist-readable

#	Research Gap	Evidence from Reviewed Literature	Proposed Direction for Novel Model
		remaining 5 papers present entirely black-box models.	diagnostic outputs for clinical adoption.
6	Small Nodule & Early-Stage Detection	Thanoon et al. specifically cite detection of sub-3mm nodules as critical. Most reviewed models are benchmarked on larger nodules. No paper reports performance stratified by nodule size.	Specifically design and benchmark model performance on sub-centimeter nodules. Incorporate multi-scale feature extraction to improve sensitivity for small early-stage lesions.
7	Low-Dose CT (LDCT) Compatibility	Thanoon et al. note LDCT is the preferred clinical screening modality to reduce radiation. All 5 research papers train on standard-dose CT. No paper addresses LDCT-specific preprocessing.	Train models on LDCT images with dedicated denoising (e.g., Gaussian/Median filtering, GAN-based enhancement) to match real clinical LDCT screening conditions.
8	Patient Metadata & Clinical Data Integration	Thanoon et al. explicitly recommend combining imaging features with medical history and genetic reports. No reviewed paper incorporates clinical metadata of any kind alongside imaging data.	Incorporate patient age, smoking history, genetic risk markers, and blood biomarker data alongside imaging for multi-source fusion, improving both diagnostic accuracy and model credibility.
9	Cloud Computing & Lightweight Deployment	Said et al.'s model has 93.32M parameters and 10.23s inference time — clinically undeployable without high-end GPU. No paper proposes model compression or cloud-based inference.	Develop lightweight model variants via pruning, quantization, or knowledge distillation, and design cloud-based inference pipelines for deployment in resource-constrained clinical environments.

VI. Discussion and Future Directions

A. Key Trends Observed Across the Reviewed Literature

A synthesis of the six reviewed papers reveals several significant trends that collectively characterize the current state and trajectory of deep learning for lung cancer detection.

- The most prominent trend is the progressive architectural shift toward hybrid models that integrate the strengths of multiple learning paradigms. While early works (Zhang et al., 2018) reviewed CNN-based approaches as largely independent architectures, more recent research papers in this review (Ansari et al., 2025; Devarajan et al., 2023) demonstrate deliberate, principled fusion of complementary models. The SVMVGGNet-16 combination of deep feature extraction with SVM classification, and the CNN-LSTM integration of spatial and temporal analysis, both reflect the field's growing recognition that no single architecture is sufficient for the diagnostic complexity of lung cancer imaging.

- A second clear trend is the move from 2D slice-based analysis toward 3D volumetric processing. Zhang et al. (2018) documented the theoretical advantages of 3D CNNs in capturing inter-slice spatial dependencies, and Said et al. (2023) operationalize this at the architectural frontier with full 3D transformer-based processing through UNETR. Thanoon et al. (2023) corroborate this trend in their

review, noting that models using 3D input consistently outperform 2D slice-based equivalents in segmentation and detection tasks. This convergence suggests that 3D volumetric processing should be considered the default rather than the optional enhancement in future model designs.

- Transformer-based architectures represent the most significant recent architectural development in the field. Said et al.'s (2023) successful application of UNETR to 3D lung segmentation demonstrates that the multi-head self-attention mechanism — originally developed for NLP — provides tangible benefits for medical volumetric image analysis through global context capture that CNN receptive fields cannot match. The fact that this advancement is confined to segmentation in the reviewed literature, with no transformer-based classification work present, highlights both the recency of this trend and the opportunity it presents.

- A fourth trend, visible most clearly in the two review papers, is the growing awareness of the gap between benchmark performance and clinical deployability. Both Zhang et al. (2018) and Thanoon et al. (2023) conclude with extensive future direction recommendations that are focused not on further accuracy improvements but on the non-accuracy dimensions of clinical readiness: interpretability, dataset diversity, computational efficiency, integration with clinical workflows, and regulatory validation. This shift in emphasis reflects the field's maturation from proof-of-concept research toward clinical translation.

B. Design Principles for a Novel Model

The gaps and limitations identified in the reviewed literature collectively define a coherent set of design requirements for a next-generation lung cancer detection and classification model. A novel model proposed in response to these gaps should embody the following design principles:

- **3D Volumetric Processing:** The model should natively process 3D CT scan volumes rather than individual 2D slices, exploiting the full spatial context across slice dimensions for both segmentation and classification tasks.

- **Unified Pipeline:** The model should integrate tumour segmentation and multi-class subtype classification (ADC, LCC, SCC, Normal) within a single end-to-end trainable framework, eliminating the disconnect between these two clinically essential tasks.

- **Transformer-Based Attention:** A hybrid CNN-Transformer encoder should be employed, combining CNN-based local feature extraction with transformer-based global context modelling through multi-head self-attention to capture both fine-grained texture and large-scale structural patterns.

- **Multi-Scale Feature Extraction:** The architecture should incorporate feature pyramid networks or multi-scale pooling to ensure sensitivity to both small early-stage nodules and larger, more established tumors.

- **Explainability:** Grad-CAM or attention heatmap visualization should be integrated into the inference pipeline to produce interpretable, radiologist-readable diagnostic maps alongside the classification output.

- **Multi-Dataset Validation:** The model should be trained and validated across at minimum three benchmark datasets (LIDC-IDRI, LUNA16, Decathlon) to establish cross-dataset generalizability.

- LDCT Compatibility: Preprocessing modules should include LDCT-specific noise reduction strategies (adaptive filtering, GAN-based enhancement) to ensure robust performance on low-radiation screening CT images.

- **Computational Efficiency:** Model architecture search and post-training compression techniques should be applied to produce a deployment-ready variant with inference time below 3 seconds per scan on standard clinical hardware.

C. Recommended Future Research Directions

Based on the identified research gaps and the design principles articulated above, the following specific future research directions are recommended for the advancement of deep learning-based lung cancer detection:

- **Multi-Modal Fusion Architectures:** Future work should develop deep learning architectures that natively fuse CT structural information with PET metabolic data and/or MRI soft tissue contrast through cross-modal attention mechanisms. Early fusion (at input level), intermediate fusion (at feature level), and late fusion (at decision level) strategies should be systematically evaluated and compared for their impact on detection sensitivity and subtype classification accuracy.

- **Federated Learning for Multi-Institutional Training:** Federated learning frameworks, which enable model training across multiple hospital systems without requiring centralized data sharing, offer a promising solution to the dataset scarcity and patient privacy constraints that restrict current research. Future models should be designed for federated deployment, with differential privacy guarantees and robust aggregation strategies to handle institutional data heterogeneity.

- **GAN-Based Data Augmentation and Synthetic Data Generation:** The persistent challenge of limited labeled medical imaging data can be partially addressed through generative adversarial network-based synthetic CT generation. Future work should develop lung cancer CT synthesis pipelines that generate realistic nodule augmentations across the full size, morphology, and subtype spectrum, enabling training on substantially larger effective datasets.

- **Prospective Clinical Validation Trials:** The most critical gap between current research and clinical deployment is the absence of prospective validation studies. Future research should include rigorous prospective clinical trials comparing automated model performance against expert radiologist diagnosis on blinded patient data, with pre-specified performance thresholds required for regulatory submission. Reader studies involving multiple radiologists and direct comparison of AI-assisted versus unassisted reading should be prioritized.

- **Clinical Metadata Fusion:** Future models should incorporate structured clinical data — including smoking history, genetic risk markers, blood biomarkers, and prior imaging — alongside imaging features through multi-branch architecture designs that learn optimal feature fusion weights through joint training. This multi-source approach more closely mirrors the actual clinical decision-making process.

- **Regulatory-Compliant Explainability:** Future research should develop standardized explainability pipelines that generate regulatory-grade diagnostic reports — combining Grad-CAM heatmaps, attention maps, feature attribution scores, and natural language explanations — that meet the documentation requirements of medical device regulatory bodies (FDA, CE, CDSCO).

VII. CONCLUSION

This paper has presented a comprehensive systematic review of deep learning-based approaches for lung cancer detection and classification, with detailed comparative and critical analysis of six significant studies spanning the period from 2018 to 2025. The reviewed methodologies collectively represent the current state

of the art in intelligent lung cancer diagnosis, encompassing hybrid CNN-SVM models (Ansari et al., 2025), comprehensive field surveys (Zhang et al., 2018), transfer learning with domain adaptation (Fei et al., 2024), hybrid recurrent-convolutional architectures (Devarajan et al., 2023), transformer-based segmentation with self-supervised classification (Said et al., 2023), and comprehensive literature reviews covering 30+ models (Thanoon et al., 2023).

The comparative analysis presented in Section 3 reveals significant diversity in architectural strategies, task formulations, dataset choices, and evaluation protocols across the reviewed literature. While reported accuracy figures range from 87% to 99.45%, direct numerical comparison is constrained by fundamental differences in task complexity, dataset characteristics, and evaluation methodology. The critical analysis in Section 4 documents specific strengths and weaknesses for each paper, revealing that no single reviewed study simultaneously addresses the full set of clinical requirements for a deployable lung cancer detection system.

Nine critical and cross-validated research gaps have been systematically identified in Section 5: (1) dataset diversity and cross-dataset generalization; (2) unified segmentation and multi-class classification pipeline; (3) transformer-based classification; (4) multi-modal imaging integration; (5) model explainability and clinical interpretability; (6) small nodule and early-stage detection; (7) LDCT compatibility; (8) patient metadata and clinical data integration; and (9) cloud computing and lightweight deployment. These gaps are not isolated limitations of individual papers but represent systematic deficiencies across the entire reviewed literature, confirmed by cross-paper analysis. They provide a rigorous, evidence-based specification for the design of a novel deep learning and image processing model for lung cancer detection.

The design principles and future research directions articulated in Section 6 translate these gaps into a coherent development roadmap, emphasizing 3D volumetric processing, hybrid CNN-Transformer architectures, unified multi-task pipelines, multi-dataset validation, LDCT compatibility, explainability integration, and deployment efficiency. Collectively, these principles define a model that would meaningfully advance beyond the current state of the art — not merely in benchmark accuracy, but in clinical relevance, trustworthiness, and practical deployability.

The findings of this review underscore a fundamental truth of the current state of the field: while deep learning has achieved remarkable performance on established benchmarks, the gap between benchmark performance and clinical utility remains substantial. Bridging this gap requires not only architectural innovation but also methodological rigor in dataset construction, clinical validation design, regulatory compliance strategy, and deployment engineering. This review provides the foundational analysis and directional guidance necessary for the next generation of research to make meaningful progress toward clinically deployable, interpretable, and globally accessible lung cancer detection systems.

VIII. References

[1] M. M. Ansari, S. Kumar, M. B. B. Heyat, H. Ullah, M. A. Bin Hayat, S. Sumbul, S. Parveen, A. Ali, and T. Zhang, “SVMVGGNet-16: A Novel Machine and Deep Learning Based Approaches for Lung Cancer Detection using Combined SVM and VGGNet-16,” Current Medical Imaging, vol. 21, Art. no. e15734056348824, 2025, doi: 10.2174/0115734056348824241224100809.

[2] P. Zhang, X. Xu, H. Wang, Y. Feng, H. Feng, J. Zhang, S. Yan, Y. Hou, Y. Song, J. Li, and X. Liu, “Computer-Aided Lung Cancer Diagnosis Approaches Based on Deep Learning,” Journal of Computer-Aided Design & Computer Graphics, vol. 30, no. 1, 2018, doi: 10.3724/SP.J.1089.2018.16919.

[3] X. Fei, Y. Wang, L. Dai, and M. Sui, “Deep Learning-Based Lung Medical Image Recognition,”

International Journal of Innovative Research in Computer Science and Technology, vol. 12, no. 3, pp. 100–105, 2024, doi: 10.55524/ijircst.2024.12.3.16.

[4] H. R. Devarajan, P. Kumar, S. Balasubramanian, V. R. Jallepalli, and S. K. Swarnkar, "Deep Learning for Automated Detection of Lung Cancer from Medical Imaging Data," in Proc. Int. Conf. Artificial Intelligence for Innovations in Healthcare Industries (ICAIIHI), 2023, doi: 10.1109/ICAIIHI57871.2023.10488962.

[5] Y. Said, A. A. Alsheikhy, T. Shawly, and H. Lahza, "Medical Images Segmentation for Lung Cancer Diagnosis Based on Deep Learning Architectures," Diagnostics, vol. 13, no. 3, Art. no. 546, 2023, doi: 10.3390/diagnostics13030546.

[6] M. A. Thanoon, M. A. Zulkifley, M. A. A. Mohd Zainuri, and S. R. Abdani, "A Review of Deep Learning Techniques for Lung Cancer Screening and Diagnosis Based on CT Images," Diagnostics, vol. 13, no. 16, Art. no. 2617, 2023, doi: 10.3390/diagnostics13162617.

[7] H. Sung, J. Ferlay, R. L. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal, and F. Bray, "Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries," CA: A Cancer Journal for Clinicians, vol. 71, no. 3, pp. 209–249, 2021, doi: 10.3322/caac.21660.

[8] R. L. Siegel, K. D. Miller, N. S. Wagle, and A. Jemal, "Cancer Statistics, 2023," CA: A Cancer Journal for Clinicians, vol. 73, no. 1, pp. 17–48, 2023, doi: 10.3322/caac.21763.

[9] S. G. Armato III et al., "The Lung Image Database Consortium (LIDC) and Image Database Resource Initiative (IDRI): A Completed Reference Database of Lung Nodules on CT Scans," Medical Physics, vol. 38, no. 2, pp. 915–931, 2011, doi: 10.1118/1.3528204.

[10] A. A. A. Setio et al., "Validation, Comparison, and Combination of Algorithms for Automatic Detection of Pulmonary Nodules in Computed Tomography Images: The LUNA16 Challenge," Medical Image Analysis, vol. 42, pp. 1–13, 2017, doi: 10.1016/j.media.2017.06.015.

[11] A. Hatamizadeh, Y. Tang, V. Nath, D. Yang, A. Myronenko, B. Landman, H. R. Roth, and D. Xu, "UNETR: Transformers for 3D Medical Image Segmentation," in Proc. IEEE/CVF Winter Conf. Applications of Computer Vision (WACV), 2022, pp. 574–584.

[12] D. Ardila et al., "End-to-End Lung Cancer Screening with Three-Dimensional Deep Learning on Low-Dose Chest Computed Tomography," Nature Medicine, vol. 25, no. 6, pp. 954–961, 2019, doi: 10.1038/s41591-019-0447-x.

[13] The National Lung Screening Trial Research Team, "Reduced Lung-Cancer Mortality with Low-Dose Computed Tomographic Screening," New England Journal of Medicine, vol. 365, no. 5, pp. 395–409, 2011, doi: 10.1056/NEJMoa1102873.

[14] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), 2016, pp. 770–778.

[15] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in Proc. Int. Conf. Medical Image Computing and Computer-Assisted Intervention (MICCAI), 2015, pp. 234–241.

[16] G. Litjens et al., "A Survey of Deep Learning in Medical Image Analysis," Medical Image Analysis, vol. 42, pp. 60–88, 2017, doi: 10.1016/j.media.2017.07.005.

[17] A. Dosovitskiy et al., "An Image Is Worth 16×16 Words: Transformers for Image Recognition at Scale," in Proc. Int. Conf. Learning Representations (ICLR), 2021.

[18] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization," in Proc. IEEE Int. Conf. Computer Vision (ICCV), 2017, pp. 618–626.

[19] M. Antonelli et al., "The Medical Segmentation Decathlon," Nature Communications, vol. 13, no. 1, pp. 1–13, 2022, doi: 10.1038/s41467-022-30695-9.

[20] N. Horeweg et al., "Detection of Lung Cancer through Low-Dose CT Screening (NELSON): A Prespecified Analysis of Screening Test Performance and Interval Cancers," The Lancet Oncology, vol. 15, no. 12, pp. 1342–1350, 2014, doi: 10.1016/S1470-2045(14)70387-0.

[21] E. Amrani, L. Karlinsky, and A. Bronstein, "Self-Supervised Classification Network," in Proc. European Conf. Computer Vision (ECCV), 2022, pp. 116–132.

[22] M. Avanzo, J. Stancanello, G. Pirrone, and G. Sartor, "Radiomics and Deep Learning in Lung Cancer," Strahlentherapie und Onkologie, vol. 196, no. 10, pp. 879–887, 2020, doi: 10.1007/s00066-020-01625-9.