applied sciences

MDPI

Review

Segmentation and Classification of Lung Cancer Images Using Deep Learning

Xiaoli Yang *, Angchao Duan, Ziyan Jiang, Xiao Li, Chenchen Wang, Jiawen Wang and Jiayi Zhou

College of Medical Technology and Engineering, Henan University of Science and Technology, Luoyang 471023, China; 240320221383@stu.haust.edu.cn (A.D.); 250320221520@stu.haust.edu.cn (Z.J.); 240320221380@stu.haust.edu.cn (X.L.); 240320221388@stu.haust.edu.cn (C.W.); 230320221322@stu.haust.edu.cn (J.W.); 230320221329@stu.haust.edu.cn (J.Z.)

* Correspondence: yxl@haust.edu.cn

Abstract

Lung cancer ranks among the world's most prevalent and deadly diseases. Early detection is crucial for improving patient survival rates. Computed tomography (CT) is a common method for lung cancer screening and diagnosis. With the advancement of computer-aided diagnosis (CAD) systems, deep learning (DL) technologies have been extensively explored to aid in interpreting CT images for lung cancer identification. Therefore, this review aims to comprehensively examine DL techniques developed for lung cancer screening and diagnosis. It explores various datasets that play a crucial role in lung cancer CT image segmentation and classification tasks, analyzing their differences in aspects such as scale. Next, various evaluation metrics for measuring model performance are discussed. The segmentation section details convolutional neural network-based (CNN-based) segmentation methods, segmentation approaches using U-shaped network (U-Net) architectures, and the application and improvements of Transformer models in this domain. The classification section covers CNN-based classification methods, classification methods incorporating attention mechanisms, Transformer-based classification methods, and ensemble learning approaches. Finally, the paper summarizes the development of segmentation and classification techniques for lung cancer CT images, identifies current challenges, and outlines future research directions in areas such as dataset annotation, multimodal dataset construction, multi-model fusion, and model interpretability.

Keywords: lung cancer; deep learning technology; classification; segmentation

Check for updates

Academic Editor: Thomas Lindner

Received: 17 December 2025
Revised: 3 January 2026
Accepted: 5 January 2026
Published: 7 January 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

On a global scale, lung cancer constitutes a major portion of deaths attributed to cancer. Its prevalence and the severity of its progression make it a leading threat to global public health [1]. Approximately 18% of all cancer-attributable deaths worldwide are caused by lung cancer. The etiology of lung cancer is multifactorial, encompassing factors such as tobacco use, exposure to airborne pollutants, gender, genetic predisposition, and advancing age [2]. Chronic cigarette smoking significantly elevates the risk of developing lung cancer [3]. Smoking is a primary cause of lung cancer among smokers. The risk factors for lung cancer extend beyond active smoking to include passive inhalation of tobacco smoke and environmental pollutants, as well as inherent factors such as genetic susceptibility. In some countries, lung cancer incidence has plateaued due to historical trends, while in others it continues to rise, reflecting a growing public health concern.

This indicates that lung cancer will become more prevalent, at least over the next several decades [4]. Beyond common respiratory symptoms such as hemoptysis, chest pain, and wheezing, lung cancer may also trigger hoarseness and dysphagia due to local invasion or distant metastasis. Metastatic disease in advanced lung cancer is typically characterized by localized symptoms, such as bone pain and headaches, and accompanied by broader systemic effects.

Lung cancer can be histopathologically classified into two main types: small cell lung cancer (SCLC) and non-small cell lung cancer (NSCLC). Among these, non-small cell lung cancer accounts for approximately 80% to 85% of all lung cancer cases, making it the most common type, while small cell lung cancer represents about 10% to 15% of lung cancer diagnoses. Unfortunately, lung cancer is often diagnosed at an advanced stage, and about 70% of patients are diagnosed when the disease has already progressed significantly. This leads to a low overall survival rate, with the five-year survival rate for lung cancer patients being approximately 16%. However, early detection of lung cancer can significantly improve treatment outcomes, increasing the five-year survival rate to around 70% [5].

Accurate classification of the lung cancer type critically guides the selection of an effective treatment regimen [6]. Currently, the detection, segmentation, and classification of lung cancer nodules can be performed using various medical imaging modalities, such as Computed tomography (CT) [7], X-ray images [8], histopathological images [9], and sputum smear microscopy images [10]. CT and magnetic resonance imaging (MRI) are routinely employed in early disease detection, contributing to enhanced patient survival rates [11,12].

Currently, lung lesion screening primarily relies on CT technology, but traditional manual image interpretation has inherent limitations. Each patient's CT data consists of hundreds of slices, where lesions often exhibit variable morphologies, small sizes, and blurred boundaries. This makes manual lesion delineation by physicians, based on experience, not only extremely labor-intensive but also prone to misdiagnosis and missed diagnoses. In response to this challenge, deep learning (DL) techniques dedicated to CT image segmentation and classification have consequently emerged. These approaches can automatically learn key lesion features from large volumes of CT images, establishing efficient segmentation and classification paradigms. This assists physicians in rapidly and accurately identifying lesions, significantly enhancing diagnostic efficiency and reliability.

Early cancer detection is crucial for improving recovery prospects [12], thus demanding precise diagnosis and timely intervention. Ultimately, irrespective of the tools used, disease diagnosis hinges on the accurate interpretation of medical information. It is essential that this interpretation is carried out by qualified specialists who possess the expertise to analyze and make informed decisions based on the data. Due to the complexity of medical images, disagreements among experts occasionally occur, which is a common challenge. As a result, intelligent diagnostic assistance systems have become essential in the medical field. Increasingly, both traditional machine learning (ML) and DL algorithms are being adopted in the analysis and interpretation of medical images for diagnostic purposes [13].

While several review articles have surveyed the application of DL in lung cancer CT image analysis, most prior works tend to focus either on segmentation or classification in isolation, or they emphasize specific architectural advancements without systematically integrating dataset characteristics, evaluation metrics, and methodological evolution into a unified analytical framework. For instance, earlier reviews such as those by Naik et al. [14] primarily addressed nodule detection and classification using convolutional neural networks (CNNs), while more recent surveys have highlighted Transformer-based approaches or attention mechanisms in medical image analysis. However, a comprehensive synthesis

that concurrently examines dataset diversity, metric selection, segmentation networks, and classification paradigms remains underrepresented.

To bridge this gap, this review provides an integrated and structured analysis of DL-driven lung cancer image segmentation and classification, with the following distinctive contributions:

1. Holistic methodological taxonomy: We systematically categorize and compare segmentation and classification techniques across three evolutionary stages: CNN-based foundations, U-shaped network (U-Net) and its attention-enhanced variants, and Transformer-integrated architectures. This dual-task perspective enables a clearer understanding of model progression and task-specific adaptations.
2. Dataset and metric-centered analysis: Unlike reviews that treat datasets merely as data sources, we provide a detailed discussion of widely used public datasets and private collections, highlighting their scale, modality, annotation quality, and impact on model generalizability. Additionally, we link evaluation metrics to clinical and technical requirements, providing guidance for metric selection in different experimental settings.
3. Comparative tabular synthesis: Through consolidated tables, we offer a comparison of datasets, methods, and performance metrics across cited studies, enabling rapid insights into trends, dataset dependencies, and advanced results.
4. Integrated challenges and future directions: We identify cross-cutting challenges, such as annotation scarcity, multimodal fusion, model interpretability, and computational efficiency, and propose cohesive future research pathways that span both segmentation and classification tasks.

By offering this integrated narrative, this review not only updates the literature with the latest advances but also provides a structured reference for researchers and clinicians seeking to develop, evaluate, or deploy DL-based tools for lung cancer image analysis. Our goal is to foster a more systematic understanding of how dataset properties, model architectures, and evaluation practices jointly shape the performance and clinical applicability of DL systems in this critical domain.

Review Methodology

To ensure a comprehensive and systematic coverage of the literature on deep learning-based segmentation and classification of lung cancer CT images, we conducted a structured review following a predefined search strategy. The literature search was performed across several major academic databases, including PubMed, IEEE Xplore, Web of Science, and Google Scholar, covering publications from 2017 to 2025. The search keywords included combinations of terms such as “lung cancer,” “CT image,” “deep learning,” “segmentation,” “classification,” “U-Net,” “Transformer,” “attention mechanism,” and “ensemble learning.”

The initial search yielded a large number of articles, which were subsequently screened based on the following inclusion criteria:

1. Studies focusing on the application of deep learning techniques to lung cancer CT image analysis.
2. Publications presenting original research, including methodological innovations, comparative evaluations, or clinical validations.
3. Articles published in English in peer-reviewed journals or conferences.

Exclusion criteria included:

1. Review articles, editorials, and commentaries.
2. Works that did not provide sufficient technical or evaluative details.

2. Dataset Discussion

In the research on lung cancer segmentation and classification using CT images, datasets play a foundational role. High-quality data directly determines model performance by providing researchers with rich, reliable imaging samples and forming the essential basis for effective model training and algorithm optimization. This enhances segmentation and classification accuracy, ultimately improving clinical diagnostic efficiency. The following sections will highlight the key characteristics of relevant datasets.

- LIDC-IDRI: The Lung Image Database Consortium and Image Database Resource Initiative (LIDC-IDRI) has become a standard benchmark dataset in lung cancer research, offering annotated CT scans for the detection and classification of pulmonary nodules.
- National Lung Screening Trial (NLST): NLST is a lung screening program using low-dose CT (LDCT) and chest X-ray (CXR), based on outcomes from a large high-risk cohort, collecting a significant volume of lung screening images. The dataset consists of CT and chest X-ray images from over 53,000 participants aged 55 to 74, primarily used to assess the effectiveness of low-dose CT in detecting lung cancer at an early stage.
- The Alibaba Cloud TianChi Medical Competition dataset: The Tianchi dataset has served as a key resource in several studies aimed at developing and evaluating deep learning models for lung cancer detection [15].
- Lung-PET-CT-Dx Dataset: Sourced from The Cancer Imaging Archive (TCIA), this dataset comprises 251,135 CT and Positron emission tomography and computed tomography (PET/CT) images from 355 patients. The cancer images were sourced from patients with biopsy-confirmed diagnoses across four major histopathological subtypes, including adenocarcinoma (ADC), Large Cell Carcinoma (LCC), Squamous Cell Carcinoma (SQC) and small cell carcinoma (SCC), providing a foundational resource for multidimensional lung cancer research. This dataset is frequently used in studies developing and evaluating DL models for lung cancer diagnosis [16].
- Private Datasets: Several investigations have relied on private datasets for evaluating DL technologies, whose origins remain undisclosed in the public domain [17].

Despite providing essential benchmarks, public datasets often originate from specific institutions or standardized protocols, limiting their generalizability to in real and complex contexts clinical environments. Variations in scanners, imaging parameters, and patient populations can lead to distribution shifts that degrade model performance when deployed in practice.

Public datasets, such as those used in Ref. [18], offer advantages such as high data quality, standardized annotations, large sample sizes, and support for reproducibility and fair benchmarking, which facilitate method comparison and promote open science. However, they may lack certain labels, differ from real clinical distributions. Private datasets, such as those in Ref. [19], on the other hand, are often more clinically representative, contain clinically validated labels, and present greater diagnostic challenges, making them valuable for evaluating model robustness and generalizability in real and complex contexts settings. Yet, they are typically smaller, harder to acquire due to privacy and regulatory constraints, costly to annotate, and may introduce selection bias, limiting reproducibility and broad applicability. Researchers often use both to complement each other: public datasets for benchmarking and private ones for clinical validation.

Most studies rely on single or few datasets, increasing the risk of domain shift when models are applied across institutions, devices, or populations. Future work should prioritize multi-center, large-scale, and multimodal data collection to enhance model robustness and clinical applicability.

Overview of Preprocessing Pipelines

A standardized preprocessing pipeline is essential for ensuring the consistency and comparability of deep learning models trained on CT images. Although specific implementations may vary across studies, the following steps constitute a common checklist adopted in most lung cancer image analysis workflows:

1. Resampling: CT scans are resampled to a uniform voxel spacing to mitigate variations in resolution across different scanners and acquisition protocols.
2. Hounsfield Unit (HU) Windowing: Voxel intensities are clipped and normalized to a specific window to enhance the contrast between pulmonary tissues and lesions.
3. Lung Masking: The lung parenchyma is segmented from the thoracic cavity using automated or semi-automated methods to exclude extraneous structures such as the chest wall, mediastinum, and airways.
4. Patching/Tiling: For memory efficiency and to handle large volumetric data, three-dimensional (3D) CT volumes are often divided into smaller 3D patches or 2D slices, especially when using 2.5D or multi-view approaches.
5. Intensity Normalization: Voxel values are further normalized to stabilize training and improve convergence.

3. Model Evaluation Metrics

Evaluation metrics serve as essential tools for measuring model performance and effectiveness. Their selection must align with the specific requirements of the experimental tasks and the inherent characteristics of the processed images, typically achieved through a multi-metric collaborative assessment to comprehensively reflect model performance. These metrics provide researchers and clinicians with objective references for evaluating algorithm reliability and application value, enabling relevant personnel to make precise and efficient decisions. Given the unique nature of medical imaging tasks, this study specifically selected multiple evaluation metrics to conduct a comprehensive, multi-dimensional assessment of model performance.

For segmentation tasks, IoU (Intersection over Union) is a metric used to evaluate the degree of overlap between a predicted region and the ground truth annotation region. It is defined as the ratio of the area of intersection to the area of union of the two regions. The Dice Similarity Coefficient (DSC) is the harmonic mean of precision and recall, defined as twice the area of overlap divided by the total number of pixels in both the predicted and ground truth segments. Both of these metrics quantify how well the predicted boundaries match the true boundaries, with higher values indicating more accurate segmentation results. A higher IoU or Dice coefficient signifies that the model has captured the target region more precisely, minimizing discrepancies between predicted and true areas. Sensitivity is a standardized metric defined as the proportion of actual positive cases that are correctly identified by a diagnostic test or classifier. Specificity is a standardized performance metric that measures the proportion of actual negative cases correctly identified as negative by a test or model.

In classification tasks, metrics such as accuracy and the Area Under the ROC Curve (AUC) are employed to evaluate the model's ability to discriminate between different classes. Accuracy is the primary and most intuitive metric for classification models, defined as the proportion of correct predictions among the total number of cases examined. Accuracy, as the most intuitive metric, is highly effective on datasets with balanced class distribution. AUC is the definitive, threshold-agnostic metric for evaluating the overall discriminative power of a binary classifier. Defined as the probability that the model ranks a randomly chosen positive instance higher than a randomly chosen negative instance, it is mathematically equivalent to the area enclosed by the ROC curve and the false positive

rate axis. An AUC approaching 1 denotes enhanced capability in discriminating between the target classes.

Clinical Benchmarks and Model Reliability Considerations

When evaluating the performance of deep learning models for lung cancer image analysis, beyond the general metrics discussed above, it is crucial to consider established clinical baselines and reliability assessments.

Baseline Models: In image segmentation tasks, U-Net and its variants have become the standard architecture and a common baseline for performance comparison in medical image segmentation, owing to their encoder–decoder structure and skip connections that excel at capturing multi-scale features. In recent years, no-new-Net (nnU-Net) has demonstrated strong performance by automating data preprocessing, network architecture design, and training pipelines, and is frequently used as a benchmark for validating new methods. In classification tasks, classical CNN architectures remain the baseline models for most comparative studies due to their mature training paradigms, powerful feature extraction capabilities, and abundant pre-trained weights. Transformer-based models, with their global modeling capabilities, are emerging as new baselines for tasks requiring the handling of complex contextual relationships.

Class Imbalance Handling: Lung cancer lesions typically occupy only a very small region in CT images, leading to a severe class imbalance problem. Commonly used mitigation strategies include weighted loss functions, data augmentation, and threshold adjustment during post-processing.

Model Calibration and Uncertainty Estimation: In clinical applications, models should not only provide predictions but also indicate their confidence in those predictions. Calibration techniques can be employed to make the model's output probabilities better reflect true confidence levels. Furthermore, methods such as deep ensembles can be used to estimate model uncertainty, assisting clinicians in assessing the reliability of AI-generated outputs.

4. Segmentation Methods

Image segmentation refers to the process of partitioning an image and delineating the boundaries of specific organs or anatomical structures. In semantic segmentation tasks, DL technologies have demonstrated significant advancements, making them a powerful tool in medical diagnostics. This technology focuses on identifying organs or lesions in medical imaging modalities like CT and MRI, providing key information about organ size and shape [20,21]. Numerous automated segmentation systems have been proposed in the literature, often using traditional methods such as edge detection and filtering as preprocessing steps. The use of DL to extract complex features has become increasingly popular. While manual feature design and extraction are critical for such systems, their complexity limits the practical deployment of these technologies. Medical researchers extensively leverage DL capabilities in image processing, particularly in image segmentation, employing models such as 2D, 2.5D, and 3D CNN [22,23]. In relation to these applications of DL, Figure 1 outlines the spectrum of segmentation methods reviewed in this paper. The figure aims to provide a structured visual overview, placing the core segmentation models discussed in detail throughout the text within a unified framework.

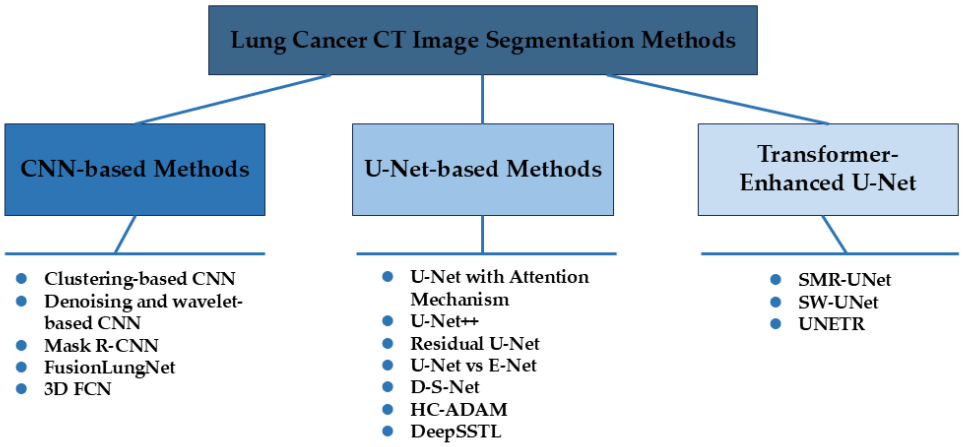


Figure 1. The segmentation methods reviewed in this paper. Abbreviations: Mask region-based CNN with the K-means kernel (Mask R-CNN); 3D Fully convolutional network (3D FCN); U-shaped network (U-Net); Efficient neural network (E-Net); Dual-stage network (D-S-Net); Adaptive densely connected U-Net with attention mechanism (HC-ADAM); Deep Semi-Supervised Transfer Learning (DeepSSTL); U-Net Transformer (UNETR).

In normal lungs, distinct differences in image attenuation exist between pulmonary and non-pulmonary regions, making the differentiation between the two relatively straightforward in CT scans. In early studies of lung segmentation, simple and direct approaches, such as region-based growing methods, edge detection-based methods, and atlas-based registration methods, were employed to separate pulmonary tissue from non-pulmonary tissue.

4.1. CNN-Based Segmentation Methods

CNN methods have found extensive applications in diverse areas, including natural and medical image analysis. Within medical imaging, lung nodule segmentation emerged as a significant early research focus. A simple CNN architecture for lung segmentation was introduced by Xu M et al. [24]. CT slices were grouped into two categories via k-means clustering, utilizing the mean and minimum voxel intensities; the final dataset was then constructed by employing cross-validation, volume intersection, connected component analysis, and image block expansion. The proposed CNN architecture comprises a six-layer convolutional network, which was complemented by a max-pooling layer. In addition, two fully connected layers were incorporated to enhance the model's ability to learn complex patterns. Researchers developed an automated lung segmentation algorithm based on the image decomposition filtering method introduced in Ref. [25]. This approach first employs image decomposition filtering for denoising without altering lung contours, followed by wavelet transformation and multiple morphological techniques to complete lung segmentation. At the end, the segmentation results are refined and smoothed through contour correction.

Another automated lung segmentation technique, described in Ref. [26], was developed using the mask region-based convolutional neural networks (Mask R-CNN) approach, which combines both supervised and unsupervised ML techniques. Rezvani et al. [18] proposed a hybrid network architecture called FusionLungNet. By integrating a residual network with 50 layers (ResNet-50) encoder, channel aggregation attention modules, multi-scale feature fusion blocks, self-optimization modules, and multiple decoders, it addresses the shortcomings of traditional lung segmentation methods in modeling long-range dependencies, fusing semantic features, and ensuring the integrity of information transmission. Experiments demonstrated that after comprehensive optimization using Structural Similarity Index Measure (SSIM), IoU, and focus loss functions, the method

achieved an IoU score of 98.04%, significantly outperforming existing approaches and substantially improving segmentation accuracy. Unlike two-dimensional approaches, the study in Ref. [27] first trained a 3D CNN on LIDC-derived regions of interest for nodule detection, which was then extended to a 3D fully convolutional network (FCN) architecture, enabling rapid, single-pass score map generation across full volumetric data. The proposed architecture, based on a FCN, demonstrated that the discriminative CNN could efficiently and rapidly generate candidate regions of interest (ROI).

4.2. Segmentation Methods Using the U-Net Network

U-Net is a deep learning-based encoder–decoder architecture, first introduced by Ronneberger et al. in 2015 [28]. It was specifically designed for semantic segmentation tasks, with particular success in biomedical image analysis. Building upon FCN, it features a symmetrical "shrink-expand" path design. The left encoder captures contextual semantic information through downsampling, while the right decoder progressively upscales to restore spatial details and achieve precise localization. This model fuses feature maps from each encoder layer with corresponding decoder layers through skip connections, effectively integrating deep semantic information with shallow positional information. U-Net leverages multi-scale feature propagation and retains multiple feature channels, enabling high-precision segmentation even with limited training data. This makes it especially effective for medical image analysis tasks.

Since the introduction of the U-Net model, it has demonstrated outstanding performance in segmenting well-defined pulmonary nodules. However, its segmentation accuracy remains limited when dealing with complex or irregularly shaped nodules. To address this limitation, Balachandran et al. [29] proposed a U-Net architecture incorporating a context-aware attention mechanism. This model achieves more precise segmentation of various nodules from lung CT images, followed by benign/malignant classification using a CNN. Experiments on Lung Nodule Analysis Challenge 2016 (LUNA 16) and LIDC-IDRI datasets demonstrated significant improvements in key metrics, including the Dice coefficient, sensitivity, and specificity. This approach effectively addresses the limitation of traditional methods in capturing irregularly shaped nodules, demonstrating significant clinical value. However, there is still potential for improvement in the segmentation accuracy of such nodules. Due to the heterogeneous nature of pulmonary nodules and their low contrast with surrounding tissues, accurately distinguishing nodular from non-nodular regions remains a key challenge in CT image segmentation. By integrating a bivariate logit distribution into the U-Net framework, Albalawi et al. [30] introduced a nested U-Net (U-Net++) model to boost segmentation accuracy and address this limitation.

For lung segmentation, Khanna et al. [31] developed a residual U-Net. The proposed model consists of a more complex network that incorporates residual units, facilitating the extraction of discriminative features required for lung segmentation. On the other hand, a comparative analysis of two deep learning models, U-Net and efficient neural network (E-Net), was conducted by Comelli et al. [32] to evaluate their relative performance. The results demonstrated that these models can effectively segment the lung parenchyma in pulmonary fibrosis patients. In Ref. [33], the authors proposed a deep learning-based method for enhancing and segmenting pulmonary X-ray images. Through a systematic comparison of four mainstream segmentation models, they found that U-Net++ achieved optimal segmentation performance due to its enhanced skip connections, dense connections, and deep supervision mechanism, significantly outperforming other models. Yi et al. [34] proposes a dual-stage network called D-S-Net that employs a detection-then-segmentation strategy, integrates a spatial attention mechanism and a combined loss function, and significantly improves the accuracy and computational efficiency of automatic gross tumor volume

(GTV) segmentation in lung cancer CT images. Deepa et al. [35] achieved lung cancer segmentation by developing the hybrid convolution (2D/3D)-based adaptive densely connected U-Net (DenseUNet) with attention mechanism (HC-ADAM). Kevin H. Leung et al. [36] developed the Deep Semi-Supervised Transfer Learning (DeepSSTL) method using nnU-Net as the backbone network for fully automated whole-body tumor segmentation and cancer prognosis assessment in PET/CT.

4.3. Transformer-Enhanced U-Net Network for Segmentation

In 2017, researchers proposed a Transformer model based on attention mechanisms [37], which was originally applied in the field of natural language processing and was subsequently successfully introduced into medical image analysis tasks.

Hou et al. [38] proposed an innovative U-Net enhancement model called SMR-UNet that significantly improves performance by organically integrating residual structures, Transformer modules, and multi-scale feature fusion strategies. This research primarily encompasses three key technical improvements: First, replacing the standard convolutional structure of traditional U-Net with residual modules to better suppress gradient explosion during deep network training Second, the Transformer module is inserted between the encoder and decoder to remodel the high-level features extracted by the encoder before performing upsampling. Finally, the approach enables an efficient combination of multi-scale features. To address the challenge of capturing long-range feature dependencies in lung nodule image segmentation, Ma et al. [39] proposed the SW-UNet model, which employs a hybrid architecture combining CNN with Visual Transformers (ViT). By incorporating the self-attention mechanism and sliding window design from Transformers, this model effectively overcomes the receptive field limitations inherent in traditional U-Nets due to convolutions, enabling the modeling of global feature correlations. Lei et al. [40] proposed a multi-task analysis model based on U-Net Transformer (UNETR) for lung cancer CT images. By leveraging UNETR for precise tumor segmentation, they achieved significant improvements over traditional U-Net and no new U-Net (nnU-Net): the Dice coefficient reached 0.96, the IoU ratio reached 0.95, and the boundary error was reduced to 2.2 mm.

The integration of Transformers with U-Net, despite enhancing medical image segmentation via superior global context modeling and multi-scale fusion, faces critical challenges. Its high computational cost, significant hardware requirements, and architectural complexity substantially limit practical deployment in many real-world settings.

Based on the review of the methods above, and to more clearly illustrate the differences in datasets and methodologies among the reviewed literature, Table 1 shows the datasets employed in the cited literature, while Table 2 summarizes the research methodologies and evaluation metrics adopted in these studies. Table 1 systematically organizes the main datasets used in the reviewed literature, including key attributes such as their names, imaging modalities, sources, and sample sizes. Table 2, on the other hand, takes a methodological perspective by summarizing the specific model architectures employed in each study, along with the performance metrics achieved on their corresponding datasets. It must be emphasized that the results presented in all tables stem from diverse experimental settings, including differences in data sources, annotation standards, and evaluation protocols. Consequently, the metrics shown in the tables should not be interpreted as directly comparable rankings across studies. Instead, they are intended to illustrate the performance potential and relative trends of different methodological approaches within their respective research contexts. In all tables, "-" indicates that the information was not provided in the Reference.

Table 1. Segmentation datasets from the reviewed literature. LUNA16 dataset includes 888 CT images from the LIDC-IDRI dataset. Different studies have used varying scales of the LUNA16 dataset. The "Sample Number" in this table refers to the dataset scale used in the Reference.

Reference	Year	Imaging	Datasets	Sample Number
[24]	2019	CT	Shengjing Hospital of China Medical University	19,967 CT scans from 201 patients
[25]	2020	CT	The multimedia database of interstitial lung diseases (ILDs)	128 patients
[26]	2020	CT	Private	13,000 CT images
[18]	2025	CT	LIDC-IDRI and Chest CT Cancer Images from Kaggle datasets	1800 CT images from LIDC-IDRI and 700 CT images from the Chest CT Cancer Images
[27]	2018	CT	Evaluation of Methods for Pulmonary Image REgistration 2010 (EMPIRE10) and VESsel SEgmentation in the Lung 2012 (VESSEL12)	83 3D CT scans
[29]	2022	CT	Lung Nodule Analysis 2016(LUNA16) and LIDC-IDRI	5000 CT images
[30]	2024	CT	LUNA16	5000 CT slices from LIDC-IDRI dataset
[31]	2020	CT	LUNA16, VESSEL12 and Hôpitaux Universitaires de Genève-Interstitial Lung Diseases database (HUG-ILD)	50 CT images from LIDC-IDRI dataset, 8050 CT images from VESSEL12 and 3000 CT images from HUG-ILD
[32]	2020	CT	Policlinico-Vittorio Emanuele Hospital	42 patients
[33]	2022	X-ray	Montgomery County X-ray Set and Shenzhen Hospital X-ray Set	138 X-ray from Montgomery County X-rays Set and 662 X-rays from Shenzhen Hospital X-ray Set
[34]	2025	CT	-	122 patients
[35]	2025	CT	LIDC-IDRI, Iraqi National Center for Cancer Diseases/Oncology Teaching Hospital (IQ-OTH/NCCD) and University of Torino Chest CT Dataset (UniToChest)	-
[36]	2024	PET/CT	TCIA and Private	168 patients
[38]	2023	CT	LIDC-IDRI	1018 CT images from LIDC-IDRI
[39]	2023	CT	LUNA16	888 CT images from LIDC-IDRI dataset
[40]	2025	CT	Private	678 CT images

Abbreviations: The Lung Image Database Consortium and Image Database Resource Initiative (LIDC-IDRI); The Cancer Imaging Archive (TCIA).

Table 2. Lung cancer segmentation approaches. IoU, DSC, Sensitivity, and Specificity are dimensionless scalars with no units. "±" represents the mean plus or minus the standard deviation. High specificity indicates that the model rarely produces false positives, reliably excludes healthy tissue; whereas a relatively lower IoU reflects that, in the context of severe class imbalance with small targets, the model still has room for improvement in pixel-level localization accuracy at tumor boundaries. Ref. [34] did not elaborate on the reasons for the lower IoU; the present study strictly adheres to the principle of objectivity in academic discourse, providing only a description and presentation of the experimental data. If the "Dataset" column lists multiple independent datasets, it indicates that the study performed internal validation only. If the "Dataset" column shows a merged entry, it indicates that the study employed cross-dataset or external validation. IoU and Dice are primary segmentation metrics. Sensitivity and Specificity are more typical for classification but are included here as reported in the original segmentation studies for cancer vs. background discrimination.

Reference	Method	Datasets	IoU	DSC	Sensitivity	Specificity
[24]	Clustering-based CNN	Private	-	96.80%	90.75%	99.90%
[25]	Denoising and wavelet-based CNN	ILDs	98.68%	98.04%	-	-
[26]	Mask R-CNN with the K-means kernel	Private	-	97.33% ± 3.24	96.58% ± 8.58	97.11% ± 3.65
[18]	FusionLungNet	LIDC-IDRI	98.04%	99.02%	99.41%	-
		Chest CT Cancer Images	98.12%	99.01%	99.20%	-
[27]	FCN	EMPIRE10	-	98.30%	-	-
		VESSEL12		99.00%		
[29]	U-Net with Attention Mechanism	LUNA16	92.47%	97.18%	95.42%	98.81%
		LIDC-IDRI	93.47%	99.88%	95.42%	98.81%
[30]	U-Net++	LUNA16	87.90%	91.76% ± 26.67	89.54% ± 3.65	85.98% ± 25.98
[31]	Residual U-Net	LUNA16	97.32% ± 0.10	98.63%	-	-
		VESSEL12	99.24%	99.62%		
		HUG-ILD	97.39% ± 0.06	98.68%		
[32]	U-Net E-Net	Private	-	95.61% ± 1.82	92.40% ± 4.20	-
				95.90% ± 1.56	93.56% ± 3.41	
[33]	U-Net++	Montgomery County X-ray Set and Shenzhen Hospital X-ray Set	95.98%	97.96%	98.38%	99.32%
[34]	D-S-Net	-	68.48%	78.52%	79.78%	99.98%
[35]	HC-ADAM	LIDC-IDRI	-	-	96.34%	96.39%
		IQ-OTH/NCCD	-	-	95.97%	95.85%

Table 2. *Cont.*

Reference	Method	Datasets	IoU	DSC	Sensitivity	Specificity
[36]	DeepSSTL	TCIA and Private	-	81.00%	-	-
[38]	SMR-UNet	LUNA16	86.88%	91.87%	92.78%	-
[39]	SW-UNet	LUNA16	-	84.00%	82.00%	99.00%

Abbreviations: Mask region-based CNN with the K-means kernel (Mask R-CNN); Fully convolutional network (FCN); U-shaped network (U-Net); Efficient neural network (E-Net); Dual-stage network (D-S-Net); Adaptive densely connected U-Net with attention mechanism (HC-ADAM); Deep Semi-Supervised Transfer Learning (DeepSSTL); The multimedia database of interstitial lung diseases (ILDs); Evaluation of Methods for Pulmonary Image REgistration 2010 (EMPIRE10);VESsel SEgmentation in the Lung 2012 (VESSEL12); Lung Nodule Analysis 2016 (LUNA16); Hôpitaux Universitaires de Genève-Interstitial Lung Diseases database (HUG-ILD); Iraqi National Center for Cancer Diseases/Oncology Teaching Hospital (IQ-OTH/NCCD); The Lung Image Database Consortium and Image Database Resource Initiative (LIDC-IDRI); The Cancer Imaging Archive (TCIA).

4.4. Comprehensive Analysis of Segmentation Methods

CNN-based segmentation methods offer architectural simplicity, training stability, and relatively low computational demands, making them suitable for initial lung parenchyma segmentation or the detection of well-defined nodules in scenarios with limited annotated data. However, their inherent local receptive field constrains the modeling of long-range spatial dependencies, leading to suboptimal performance when segmenting complex, ill-defined, or disseminated lesions.

U-Net and its variants achieve a balance between fine detail and high-level semantics through their encoder–decoder structure and skip connections, establishing themselves as the de facto standard for medical image segmentation, particularly for precise pulmonary nodule delineation. Yet, their performance gains often come at the cost of increased network depth and complexity, and the foundational U-Net still exhibits limitations in capturing comprehensive global context. Transformer-enhanced U-Net models fundamentally augment global contextual modeling via self-attention mechanisms, demonstrating enhanced performance on challenging segmentation tasks. This superior performance, however, is accompanied by substantial computational costs, prolonged training times, and a strong dependency on large-scale, high-quality annotated datasets, which poses significant barriers to their direct deployment in resource-constrained or real-time clinical settings.

5. Classification Methods

DL methods have been shown to achieve better performance than conventional approaches in disease classification. Aiming to improve lung cancer diagnostic accuracy and image quality, the authors in Ref. [41] focused on reducing misdiagnosis rates. CT images were sourced from the TCIA dataset, with noise removed using a weighted average histogram equalization method. Furthermore, a novel segmentation algorithm based on enhanced dense clustering techniques was proposed to extract spectral features from diseased regions in the segmented images. These features were subsequently integrated into deep learning algorithms to facilitate lung cancer detection.

Computer-aided diagnosis (CAD) systems utilize advanced algorithms to assist radiologists by identifying subtle abnormalities in medical images. By providing additional insights and highlighting areas of concern, CAD systems enhance the radiologist's ability to make more reliable and timely diagnoses. Consequently, CAD has thus become an indispensable tool across major medical imaging disciplines, notably in radiology, oncology, and cardiology. With significant advancements in intelligent DL tools and methodologies, the accuracy of CAD systems across various medical imaging modalities has substantially improved. CXRs, CT, PET, MRI, and contrast-enhanced CT (CE-CT) are among the common

imaging modalities employed in the diagnosis of lung cancer. Among these, CT images are widely regarded as the preferred choice for staging and classification of lung cancer due to their ability to capture rich and informative features. The combined use of CT images and deep learning constitutes a widely adopted strategy in the literature for detecting malignant lung tumors, as noted in a relevant review [14]. Figure 2 provides a conceptual framework by presenting an overview of the classification methods reviewed in this paper, illustrating the technical evolution from CNNs to advanced Transformer-based and ensemble learning approaches, with the lung cancer image sourced from the TCIA dataset.

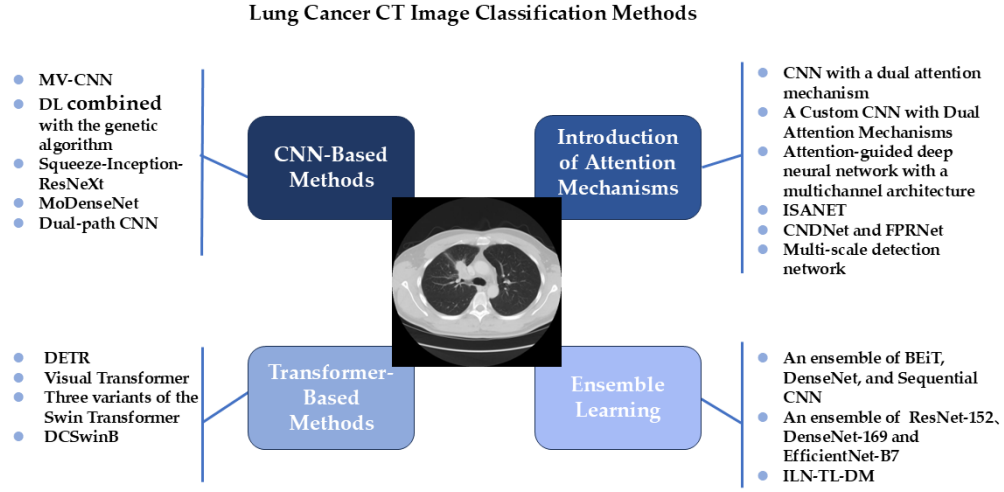


Figure 2. The classification methods reviewed in this paper. The lung cancer image is sourced from the TCIA dataset. Abbreviations: Multi-view CNN (MV-CNN); Candidate Nodule Detection Network (CNDNet); False Positive Reduction Network (FPRNet); Detection Transformer (DETR); Three variants of the Swin Transformer (Swin Transformer Base, Swin Transformer Tiny, Swin Transformer Small); Improved LeNet with Transfer Learning and DeepMaxout (ILN-TL-DM); Bidirectional Encoder Representation from Image Transformers (BEiT); Densely connected convolutional network (DenseNet).

5.1. CNN-Based Classification Methods

Historically, research on lung cancer detection has centered on distinguishing benign pulmonary nodules from malignant ones. The goal of these studies was to provide a clear differentiation between the two categories, helping healthcare professionals make quicker diagnostic decisions. In Ref. [42], a multi-view convolutional neural network (MV-CNN) was introduced by Liu et al. specifically tailored for lung cancer classification across binary and ternary tasks. Empirical results confirmed that the multi-view approach surpassed traditional single-view techniques across both binary and ternary classification tasks. In a later study, da Silva et al. [43] presented a lung nodule classifier merging deep learning with genetic algorithms for malignancy determination. For classification, a novel approach called Squeeze-Inception-ResNeXt is proposed. Nagaraj et al. [44] performed feature extraction using a CNN optimized by the slime mold algorithm (SMA). Dey et al. [19] developed a binary classifier that employs a four-channel CNN to convert 3D images into category labels.

In Ref. [45], the authors describe a CNN-based lung cancer classifier that utilizes two networks to extract and classify cancer features from FDG-PET and CT images. Convolutional operators generate low-level features, while rectified linear units (ReLU) activation and max-pooling downsampling of feature maps extract higher-level features. This study aims to distinguish T1-T2 and T3-T4 categories, achieving an accuracy rate of 90%.

The literature indicates several challenges inherent in current lung cancer classification algorithms: (a) binary segmentation of the lung frequently demonstrates limited

concordance with the standard diagnoses established by clinical specialists; (b) ROI-based methods are slow and heavily depend on expert manual input; (c) standard image processing algorithms cannot accurately segment lung nodules.

5.2. Introduction of Attention Mechanisms

During the feature fusion stage, the attention mechanism plays a pivotal role. It selectively emphasizes important features and enhances feature representation through dynamic weight adjustment, enabling nodule classification models to more accurately identify and distinguish different nodule types, thereby improving overall classification performance.

To address the challenges of low detection efficiency and high misclassification rates in lung nodule detection from lung cancer CT images, UrRehman et al. [46] proposed a custom CNN integrating dual attention mechanisms. This model precisely focuses on key regions of nodular lesions through the synergistic action of channel attention and spatial attention mechanisms, while employing global average pooling to integrate spatial feature information. Zheng et al. [47] adopted a method of standardizing CT image voxels to construct RGB images at three scales for each nodule. They then introduced three distinct attention models to extract the corresponding category features. By effectively highlighting the most critical regions within the images, this approach enables precise differentiation between nodules and surrounding tissues, thereby achieving accurate classification of various pulmonary nodule types. To address the classification challenges arising from overlapping features among non-small cell lung cancer subtypes in CT images, Xu et al. [48] proposed the ISANET classification model, which integrates dual attention mechanisms at both the channel and spatial levels. Built upon the InceptionV3 architecture, the model incorporates an embedded attention mechanism to precisely focus on lesion regions, enabling three-way classification of squamous cell carcinoma, adenocarcinoma, and normal tissue. Experimental results showed that ISANET achieved classification accuracies of 95.24% and 98.14%, respectively, significantly outperforming traditional models such as AlexNet and VGG16. These findings validate the effectiveness and clinical value of attention mechanisms in enhancing lung cancer image classification accuracy.

To overcome the challenges posed by the multi-scale characteristics and high false positive rates in lung nodule detection, Li et al. [49] introduced a two-stage detection framework. This framework integrates a Global Channel Spatial Attention Mechanism (GCSAM) to enhance the model's ability to focus on relevant features and suppress irrelevant ones. The model employs the Candidate Nodule Detection Network (CNDNet) architecture with a residual resolution network (Res2Net) as the backbone to extract multi-scale nodule features. GCSAM is then used to fuse global contextual information and adaptively adjust feature weights, while the Hierarchical Progressive Feature Fusion (HPFF) module effectively integrates deep semantic information with shallow positional cues to achieve high sensitivity across nodules of varying sizes. In the second stage, the False Positive Reduction Network (FPRNet) module further distinguishes true nodules from nodule-like structures, substantially reducing false positives. Cao et al. [50] proposed a multi-scale detection network that integrates an attention mechanism. The network enhances feature extraction through a specially designed ResSCBlock module and effectively mitigates missed detections of lesions of various sizes, particularly small nodules, by employing multi-scale prediction.

The attention mechanism facilitates the effective transmission of feature information by computing attention weights, enabling more comprehensive integration of features across different modalities. In some cases, it can also perform weighted fusion of features

across different levels, allowing for enhanced feature representation and thereby improving the model's overall performance.

5.3. Transformer-Based Classification

The Transformer model, introduced in 2017 and originally developed for natural language processing, has been increasingly applied to medical image analysis owing to its powerful contextual modeling enabled by a global receptive field. This global receptive field substantially strengthens the model's ability to capture long-range dependencies and contextual relationships.

The Detection Transformer (DETR), a Transformer-based deep neural network, was proposed by Barbouchi et al. [51] to leverage PET/CT images for lung cancer classification. After data preprocessing, the DETR model, which integrates a ResNet-50 backbone for feature extraction with a Transformer encoder–decoder architecture, was employed to simultaneously perform tumor localization and histological subtype classification. Experimental results show that the model achieves a mean Intersection over Union (mIoU) of 0.83 for tumor localization and a histological classification accuracy of 0.96, both surpassing traditional CNN-based models such as MobileNet and ResNet-50. These findings suggest that Transformer architectures offer clear advantages in multi-task medical image analysis, providing a promising direction for automated lung cancer diagnosis. Future work will focus on integrating additional Transformer variants and incorporating advanced segmentation techniques to further enhance overall performance.

To differentiate benign from malignant tissues, Nishigaki et al. [52] investigated the application of ViT on ^{18}F-FDG PET/CT images. Following preprocessing, the ViT model was trained and compared against baseline CNN architectures such as EfficientNet and densely connected convolutional network (DenseNet). The results showed that ViT achieved an AUC of 0.90, markedly outperforming the CNN counterparts. Even in challenging cases characterized by low FDG uptake, the ViT maintained strong performance with an AUC of 0.81, substantially higher than DenseNet's 0.65. Visual analysis shows that ViT can attend to critical regions across the entire image, effectively identifying low-uptake lesions that CNNs often fail to detect, as CNNs tend to rely more heavily on localized high-uptake features. The study further found that PET/CT fusion input yielded the best performance, whereas CT-only input produced the poorest results, highlighting the importance of integrating functional and anatomical information. Despite the limitation of single-center data, this work provides the first evidence supporting the potential of ViT in automated PET/CT diagnosis and demonstrates clear clinical value, particularly in identifying low-uptake lesions. These findings open new avenues for reducing missed diagnoses and advancing precision medicine.

Sun et al. [53] proposed a lung cancer image classification and segmentation framework based on an enhanced Swin Transformer, aiming to improve the accuracy of automated medical image diagnosis. For classification, 3D CT slices were converted into 2D images with data augmentation applied. Three variants of the Swin Transformer architecture, namely Swin Transformer Tiny (Swin-T), Swin Transformer Small (Swin-S), and Swin Transformer Base (Swin-B), were trained. Among these, the pre-trained Swin-B model achieved a top-1 accuracy of 82.26%. In segmentation tasks, a model based on the Unified Perceptual Parsing Network (UPerNet) architecture combined with Swin Transformer achieved a substantially higher mean IoU compared to ResNet-101 and Data-efficient image Transformers-Small (DeiT-S). Faizi et al. [54] proposed DCSwinB, a lung cancer nodule classification model for CT images, based on Swin-Tiny ViT. The model employs a dual-branch architecture, where CNNs capture local features and Swin Transformers extract global features, while also incorporating convolutional multilayer perceptron (Conv-

MLP) modules. After pretraining, DCSwinB outperformed models such as ResNet50 in 10-fold cross-validation, while offering advantages in parameter efficiency, computational complexity, and inference speed, making it well-suited for supporting early lung cancer diagnosis.

5.4. Ensemble Learning

Ensemble learning is an advanced feature processing strategy that improves classification performance by combining multiple models to fully exploit their complementary features. By training diverse models to capture distinct patterns in image data and then fusing their predictions, this approach achieves a more comprehensive feature representation and higher classification accuracy. In the field of medical image analysis, this study demonstrates that ensemble learning-based methods for classifying lung cancer CT images can substantially enhance diagnostic performance.

Quasar et al. [55] developed a novel lung cancer detection and classification framework by integrating heterogeneous models such as Bidirectional Encoder Representation from Image Transformers (BEiT), DenseNet, and sequential CNNs, combined with ensemble strategies including weighted frame fusion and boosting. Gautam et al. [56] innovatively combined three advanced convolutional networks (ResNet-152, DenseNet-169, and EfficientNet-B7) with a dual-metric evaluation framework, which integrates the area under the receiver operating characteristic curve (ROC-AUC) and the F1 score, to determine optimal weights for each base model. This optimized weighting strategy effectively reduced the risk of missed diagnoses, providing a reliable auxiliary decision-support solution for early lung cancer detection. CM et al. [57] addressed the challenges of accurate lung cancer CT classification caused by variable lesion characteristics by proposing the improved LeNet with Transfer Learning and DeepMaxout (ILN-TL-DM) hybrid deep architecture, which incorporates ensemble learning principles. Its core strategy combines "multi-base learner construction with soft voting fusion" to enhance the classification performance of lung cancer CT images. Based on the summary of the methods presented above, and in order to more clearly illustrate the differences in datasets and methodologies across the reviewed literature and to compare classification outcomes among different studies, Table 3 presents the datasets and their scales used in these publications, providing readers with an intuitive comparison of data foundations. Meanwhile, Table 4 outlines the respective research methods, classification tasks, and evaluation metrics for each study, facilitating a horizontal comparison of model performance.

Table 3. Classification datasets from the reviewed literature.

Reference	Year	Imaging	Datasets	Sample Number
[42]	2017	CT	LIDC-IDRI	96 patients
[43]	2017	CT	LIDC-IDRI	3243 nodules from 833 patients
[44]	2025	CT	Chest CT-Scan Images Dataset	967 CT images
[19]	2018	CT	LIDC-IDRI and Private Dataset	1010 CT images from LIDC-IDRI and 147 lung nodule samples from Private Dataset
[45]	2018	PET/CT	Humanitas Clinical and Research Center	472 patients
[46]	2024	CT	LUNA16	888 CT images from LIDC-IDRI dataset

Table 3. *Cont.*

Reference	Year	Imaging	Datasets	Sample Number
[47]	2024	CT	LIDC-IDRI	1018 patients
[48]	2022	CT	Private, Kaggle datasets and TCIA	619 CT images from private 737 CT images from Kaggle 674 CT images from TCIA
[49]	2025	CT	LUNA16	888 CT images from LIDC-IDRI dataset
[50]	2023	CT	LUNA16	888 CT images from LIDC-IDRI dataset
[51]	2023	PET/CT	Lung-PET-CT-Dx	25 patients
[52]	2024	PET/CT	PET/CT	18,301 PET/CT images from 207 patients
[53]	2023	CT	LUNA16	888 CT images from LIDC-IDRI dataset
[54]	2025	CT	LIDC-IDRI	1004 nodules
[55]	2024	CT	Chest CT-Scan Images Dataset from Kaggle datasets	-
[56]	2024	CT	LIDC-IDRI	8106 CT images
[57]	2025	CT	LUNA16	888 CT images from LIDC-IDRI dataset

Abbreviations: The Lung Image Database Consortium and Image Database Resource Initiative (LIDC-IDRI); The Cancer Imaging Archive (TCIA); Lung Nodule Analysis 2016 (LUNA16).

Table 4. Lung cancer classification approaches. Swin-S is a model with a large number of parameters, which requires a substantial amount of data to fully leverage its performance. However, the dataset constructed by the authors [53] was too small in scale, resulting in a very low classification accuracy when using Swin-S.

Reference	Method	Datasets	Class	Accuracy	Sensitivity	Specificity	AUC
[42]	MV-CNN	LIDC-IDRI	Benign vs. Malignant	94.59%	89.13%	99.91%	0.98
[43]	DL combined with the genetic algorithm	LIDC-IDRI	Benign vs. Malignant	94.50%	98.00%	91.00%	0.95
[44]	Squeeze-Inception-ResNeXt	Chest CT-Scan Images Dataset	ADC vs. LCC vs. SQC	97.70%	98.10%	97.40%	-
[19]	MoDenseNet	LIDC-IDRI	Benign vs. Malignant	90.40%	90.47%	90.33%	0.95
		Private		86.84%	-	-	0.90
[45]	Dual-path CNN	Private	T1-T2 vs. T3-T4	69.10%	70.00%	66.70%	0.68
[46]	CNN with a dual attention mechanism	LUNA16	Benign vs. Malignant	94.40%	94.69%	93.17%	0.98
[47]	Attention-guided deep neural network with a multichannel architecture	LIDC-IDRI	Benign vs. Malignant	90.11% ± 0.24	-	-	0.96

Table 4. *Cont.*

Reference	Method	Datasets	Class	Accuracy	Sensitivity	Specificity	AUC
[48]	ISANET	Private	SQC vs. ADC vs. normal	99.60%	ADC: 92.48% SQC: 91.35% Normal: 98.47%	-	-
		Kaggle		95.24%	ADC: 82.41% SQC: 88.20% Normal: 94.21%	-	-
		TCIA		98.14%	-	-	-
[49]	CNDNet and FPRNet	LUNA16	True Nodule vs. False Positive Nodule	-	97.70%	-	-
[50]	Multi-scale detection network	LUNA16	Pulmonary Nodule vs. Non-nodule	-	45.65%	-	-
[51]	DETR	Lung-PET-CT-Dx	ADC vs. SCC vs. LCC	96.00%	ADC: 100.00%	-	0.98
		Lung-PET-CT-Dx			SCC: 99.00%		
		Lung-PET-CT-Dx			SQC: 88.00%		
[52]	ViT	-	Benign vs. Malignant	90.00%	-	-	0.90
[53]	Swin-B	LUNA16	Pulmonary Nodule vs. Non-nodule	82.26%	-	-	-
	Swin-T			82.26%	-	-	-
	Swin-S			19.76%			
[54]	DCSwinB	LUNA16	Benign vs. Malignant	87.94%	85.56%	85.65	0.94
[55]	An ensemble of BEiT, DenseNet, and Sequential CNN	Chest CT-Scan Images Dataset	ADC vs. SCC vs. LCC vs. normal	98.00%	98.70%	97.30%	-
[56]	An ensemble of ResNet-152, DenseNet-169 and EfficientNet-B7	LIDC-IDRI	Benign vs. Malignant	97.23%	98.07%	-	0.95
[57]	ILN-TL-DM	LUNA16	cancer vs. non-cancer	96.20%	96.70%	95.50%	0.99

Abbreviations: Multi-view CNN (MV-CNN); Candidate Nodule Detection Network (CNDNet); False Positive Reduction Network (FPRNet); Detection Transformer (DETR); Visual Transformers (ViT); Swin Transformer Base (Swin-B); Swin Transformer Tiny (Swin-T); Swin Transformer Small (Swin-S); Bidirectional Encoder Representation from Image Transformers (BEiT); Densely connected convolutional network (DenseNet); Improved LeNet with Transfer Learning and DeepMaxout (ILN-TL-DM); The Lung Image Database Consortium and Image Database Resource Initiative (LIDC-IDRI); The Cancer Imaging Archive (TCIA); Lung Nodule Analysis 2016 (LUNA16); Adenocarcinoma (ADC); Large Cell Carcinoma (LCC); Squamous Cell Carcinoma (SQC); Small cell carcinoma (SCC).

Ensemble learning leverages the strengths of multiple models to fully exploit their complementary capabilities in feature learning, thereby improving overall nodule prediction accuracy. By extracting and fusing features from different perspectives, multiple models produce a more comprehensive and precise final feature representation [58]. The success of ensemble methods hinges on two critical factors: the prudent selection of a diverse set of base learners and careful parameter tuning, to ensure optimal performance while mitigating overfitting or underfitting. Therefore, when implementing ensemble

learning, it is crucial not only to choose suitable base models for feature extraction and fusion but also to fine-tune each model to ensure that the combined framework achieves optimal performance.

5.5. Comprehensive Analysis of Classification Methods

A comprehensive synthesis of classification methodologies reveals distinct profiles suited to different diagnostic goals. CNN-based classification methods, serving as the foundational architecture, deliver robust and efficient performance on binary classification tasks with moderate data volumes, thanks to their efficient local feature extraction and mature training protocols. Their limitation lies in modeling complex global relationships between lesions and surrounding tissues, which hampers performance in distinguishing NSCLC subtypes with high visual similarity. The incorporation of attention mechanisms enables dynamic focus on critical image regions, effectively suppressing irrelevant background noise. This leads to marked improvements in fine-grained classification accuracy without a proportional surge in parameters and provides a degree of visual interpretability for diagnosis. However, designing and optimizing these attention modules requires domain expertise, and there is a risk of over-focusing on local patterns at the expense of holistic semantic context.

Transformer-based classification models achieve genuine long-range dependency modeling through global self-attention, demonstrating superior potential over traditional CNNs in challenging cases characterized by low contrast or low metabolic activity, especially within multimodal fusion frameworks. A significant limitation is that these models typically require substantial computational resources, and their considerable computational footprint restricts deployment on edge devices. Ensemble learning strategies combine predictions from multiple heterogeneous models, effectively leveraging complementary features to significantly enhance final classification accuracy, stability, and generalization, often attaining top-tier performance. The trade-offs are multiplied model complexity, training costs, inference time, and a further reduction in interpretability due to the compounded "black-box" nature of the decision process.

5.6. Comparative Analysis Between 2D/3D and Single/Multi-Modal Methods

With the widespread application of deep learning in lung cancer image analysis, various data representation methods and modality fusion strategies have been adopted, primarily including 2D, 2.5D, and 3D convolutional neural networks, as well as single-modal and multi-modal fusion approaches. These methods differ significantly in model architecture, computational requirements, data needs, and performance, directly influencing their suitability for in real and complex contexts clinical settings. This section systematically compares these approaches and provides practical guidance for selection under different research and application scenarios.

2D CNNs extract features from individual slices or projected views, making them suitable for processing 3D medical image slices. Their advantages include simple architecture, fewer parameters, fast training, and ease of transfer learning from pre-trained models. However, 2D models cannot capture spatial context across consecutive slices, which may limit accurate modeling of nodule volume, 3D morphology, and spatial relationships.

3D CNNs process 3D volumetric data directly, extracting spatial features through 3D convolutional kernels. They better preserve the 3D shape and spatial continuity of lesions, often achieving higher accuracy in segmentation and classification tasks, especially for complex or poorly defined nodules. The main drawbacks are high computational complexity, large memory consumption, longer training times, and limited availability of large-scale pre-trained models.

2.5D/multi-view CNNs offer a compromise by combining multiple adjacent slices or multi-planar reconstructions as input. This approach retains partial spatial information while controlling computational cost. For example, multi-view CNNs extract features from axial, coronal, and sagittal views and fuse them to approximate 3D context awareness.

Single-modal methods are simpler, easier to train and deploy, and form the basis of most studies. However, they lack functional information, which can limit their ability to differentiate between benign and malignant lesions with similar anatomical appearance.

Multi-modal fusion methods combine CT with functional imaging such as PET, integrating anatomical and metabolic information to improve diagnostic accuracy. Fusion strategies include early fusion, late fusion, and intermediate fusion. While multi-modal approaches often achieve higher AUC and sensitivity, they face challenges such as data registration difficulties, higher annotation costs, increased computational demands, and limited clinical availability of multi-modal data.

Practical selection guidance:

- Data availability: For limited CT data, start with 2D/2.5D models and use transfer learning. For complete 3D CT volumes with sufficient resources, consider 3D CNNs or Transformers. With well-aligned multi-modal data, explore attention-based fusion networks.
- Hardware constraints: 2D models are preferable under limited memory; 3D/multi-modal models require high-performance GPUs and optimized training strategies.
- Task requirements: 3D models excel in volumetric segmentation; 2D/2.5D methods can be effective for slice-based classification; multi-modal fusion is recommended for challenging diagnoses requiring metabolic information.

Future directions include lightweight 3D networks, cross-modal self-supervised pre-training, and federated learning for multi-center multi-modal collaboration, which will help balance performance and resource efficiency in clinical deployment.

6. Challenges Toward Clinical Translation

Current deep learning-based methods for pulmonary nodule analysis have demonstrated excellent performance in laboratory settings, but their true value depends on whether they can be translated into trustworthy, reliable, and usable clinical tools. This chapter systematically outlines the key challenges and feasible pathways that must be addressed during this translation process from a clinical practice perspective.

6.1. Clinical Adaptation and Integration

To facilitate clinical adoption, lung cancer segmentation and classification systems should be integrated into existing hospital infrastructures, including the Picture Archiving and Communication System (PACS) and the Radiology Information System (RIS). From a workflow perspective, automated inference processes should be initiated upon image acquisition or study archival, with the resulting outputs returned to PACS in the form of structured reports, visual overlays, or quantitative annotations that can be directly reviewed within the radiologist's routine reading environment. For large-scale screening or triage applications, near–real-time performance is desirable, with end-to-end processing times typically expected to remain within tens of seconds to a few minutes per study. In contrast, for diagnostic confirmation or treatment planning, slightly longer processing times may be tolerated, provided that accuracy and reliability are prioritized. In terms of performance requirements, screening-oriented systems should emphasize high sensitivity to minimize missed malignant lesions, even at the expense of a moderate increase in false positives. Conversely, diagnostic or decision-support systems intended for clinical confirmation should aim for a balanced trade-off between sensitivity and specificity, ensuring both

accurate detection and acceptable false alarm rates. Effective clinical deployment also requires clearly defined protocols for handling false positive and false negative predictions. False positives should be presented as low-confidence alerts or secondary findings. False negatives, which pose greater clinical risk, should be mitigated through uncertainty-aware model outputs, confidence scoring, or human-in-the-loop review mechanisms.

6.2. Clinical Validation Paradigm

To ensure the effectiveness of deep learning models in real and complex contexts, their performance validation must extend beyond evaluation on single, idealized datasets. A comprehensive clinical validation should follow a rigorous paradigm similar to that established by Cui et al. [59] in their multicenter study, comprising three key validation tiers: First, preliminary validation on large-scale, multi-source internal datasets; followed by external testing on completely independent international public benchmark datasets to assess cross-domain robustness; and finally, systematic comparison of model performance against the diagnostic outcomes of multiple radiologists. Research indicates that constructing this complete chain of evidence—encompassing internal validation, external testing, and human–machine comparison—enables systematic evaluation of deep learning models' reliability and practical value within real clinical workflows.

6.3. Interpretability and Trustworthiness

The inherent opacity of deep learning models remains a major obstacle to their broad clinical adoption. Enhancing model interpretability is therefore essential for establishing clinical trust. This includes: using visualization techniques to intuitively display the image regions the model relied on for classification or segmentation; exploring multimodal fusion to provide clinical rationale-aligned corroboration for model predictions; and attempting to incorporate medical prior knowledge or constraints into model design to make its outputs more consistent with anatomical and pathophysiological principles. Furthermore, models deployed clinically must possess capabilities for auditability and traceability, and be able to undergo continuous learning and safe updates while protecting privacy.

6.4. FEM–AI Synergistic Integration

Finite Element Method-Artificial Intelligence (FEM-AI) integration transcends simple descriptive juxtaposition through three core synergistic approaches: leveraging FEM as a physics-informed foundation to guide AI training with thermally meaningful features, establishing a closed-loop calibration where FEM predictions refine AI models and AI-derived insights optimize FEM parameters, and designing AI architectures tailored to capture the physical phenomena modeled by FEM. This integration outperforms disjointed simulation or data-driven methods by ensuring AI learns physically interpretable patterns, dynamically enhancing the accuracy of both FEM and AI, and aligning outputs with in real and complex contexts physics—bolstering accuracy, interpretability, and generalizability in relevant monitoring applications—with methodological maturity most pronounced in AI model design and results validation sections.

7. Summary and Outlook

7.1. Summary

We summarize and analyzes the research progress, key challenges, and future trends in the field of deep learning-based segmentation and classification of lung cancer images. By synthesizing a large body of literature, we have constructed a comprehensive analytical framework encompassing datasets, evaluation systems, methodological evolution, and

clinical translation, revealing the core dynamics and underlying bottlenecks in the transition from fundamental algorithm research to clinical practicality.

The core findings and insights are summarized as follows:

1. Data serves as the cornerstone of performance, yet its quality and diversity constitute a major bottleneck: Public datasets provide standardized benchmarks for research, promoting algorithm comparability and reproducibility. These approaches effectively utilize large-scale datasets to enhance model performance, allowing for more accurate and efficient analysis of lung cancer [11]. However, these datasets often originate from specific acquisition protocols, leading to domain shift issues that limit model generalizability in real and complex contexts, diverse clinical environments. While private datasets offer greater clinical representativeness, their small scale, difficulty of acquisition, and high annotation costs hinder broader research. For the future, constructing large-scale, multi-center, multimodal benchmark datasets with unified annotation quality is a primary task for advancing the field.
2. Methodological evolution follows a clear trajectory from "local to global" and "single to fusion":
    - Segmentation Task: Methods have evolved from CNN-based architectures relying on local receptive fields, to U-Net and its variants that achieve multi-scale feature fusion through encoder–decoder structures and skip connections, and further to Transformer-enhanced networks that establish global contextual modeling using self-attention mechanisms. This path reflects the progressive and deeper resolution of inherent challenges such as irregular nodule morphology, blurred boundaries, and variable sizes.
    - Classification Task: Evolution has progressed from robust feature extraction using CNNs, to the introduction of attention mechanisms for dynamic focus on key lesion regions and noise suppression, to leveraging Transformers for capturing complex long-range dependencies between lesions and surrounding tissues, and finally to ensemble learning strategies that combine the complementary strengths of heterogeneous models to maximize performance. This progression aims to enhance model discriminative power for visually similar subtypes. These DL-based methods, mainly relying on CNNs, have made significant advancements in lung nodule diagnosis [60,61].
3. The evaluation system requires deeper alignment with clinical objectives: Although metrics such as Dice coefficient, IoU, accuracy, and AUC are widely adopted, they have limitations in reflecting the clinical utility of models. Future evaluation should place greater emphasis on:
    - Robustness under extreme class imbalance.
    - Strict control of false negatives, given their higher clinical risk.
    - The introduction of uncertainty estimation and model calibration to provide clinicians with decision confidence references.
    - Conducting rigorous external validation and human–machine comparison experiments, moving beyond internal performance reporting on single datasets.
4. The performance of models in controlled lab settings often does not translate well to in real and complex contexts clinical practice: While most current studies report excellent performance on controlled datasets, issues such as poor model interpretability, high computational resource demands, difficulties in integration with existing hospital workflows, and a lack of cross-institutional generalization validation severely hinder clinical deployment. Successful clinical translation requires not only high-performance algorithms but also a comprehensive solution encompassing system integration,

real-time requirements, human–computer interaction design, continuous learning, and ethical considerations.

7.2. Outlook

Notwithstanding the considerable progress made, barriers persist to the full clinical translation of these research advances. Building on this, future research can focus on breakthroughs in the following key areas:

1. Advancing automation and standardization in data annotation: Addressing the challenges of high cost, lengthy cycles, and subjective variability in high-quality medical image annotation requires a focus on developing automated techniques based on semi-supervised and self-supervised learning. These methods utilize pre-trained high-performance models for preliminary annotation. For example, Xu et al. [24] utilized clustering-generated datasets to train CNNs for lung segmentation, demonstrating the potential of leveraging unlabeled data for annotation generation. In future research, automated annotation technologies have the potential to address issues related to poor annotation quality and limited annotation resources [62].
2. Deepening multimodal data fusion and federated learning applications: Accurate lung cancer diagnosis depends on integrating multi-source information. Studies such as Leung et al. [36] employed semi-supervised transfer learning on PET/CT data, showcasing the value of multi-modal integration for tumor segmentation. Future efforts should focus on creating unified multimodal datasets that include CT, PET/CT, pathological images, and clinical texts, while also developing network architectures capable of deeply integrating these heterogeneous data. A For both protecting patient privacy and leveraging multicenter data, it is imperative to actively explore federated learning frameworks. Such frameworks facilitate collaborative model training without sharing raw data, thereby significantly enhancing the generalization capability of algorithms.
3. Exploring lightweight models and multi-model fusion approaches: To meet the clinical demands for computational efficiency, lightweight architectures such as E-Net [30] have shown competitive performance with reduced complexity. Meanwhile, ensemble methods like those proposed by Quasar et al. [55] illustrate how multi-model fusion can boost classification accuracy. The advancement of lightweight models is a key driver for enabling real-time algorithm deployment on edge devices. For challenging lung cancer lesion segmentation tasks, future research may explore integrating multi-model fusion technologies, such as Generative Adversarial Networks (GANs), CNNs, Transformer models, and large visual models [63], to further enhance lung cancer segmentation performance.
4. Enhanced Model Interpretability: Generally, DL models lack a comprehensive theoretical framework to explain the relationship between inputs and outputs through the hidden layers. Model interpretability helps clinicians understand the basis of the model's decisions, thereby increasing trust in its predictive outcomes. Attention-based models such as ISANET [48] and dual-attention CNNs [46] have begun to provide visual cues for model decisions, offering a step toward interpretability. Future research may focus on developing interactive interfaces for medical diagnostic systems, enabling human–machine interaction to help healthcare professionals better understand the model's decision-making process. This would improve physicians' responsiveness to varying inputs, while enhancing the transparency and clarity of diagnostic systems [58].
5. Enhancing Model Generalization and Cross-Scenario Adaptability: Future research should focus on improving model generalization across different imaging devices,

protocols, and patient groups. Methods from other fields, such as industrial inspection, can be referenced to enhance model robustness. For instance, the hybrid FEM–AI approach proposed by Prattico et al., which integrates finite element modeling with infrared thermography, combines physical simulations with real data to improve the understanding of thermal anomaly patterns under varying conditions, offering valuable insights for addressing domain shifts in medical image analysis [64]. Further work by the same team demonstrated how FEM simulations can generate diverse thermal datasets, and when combined with U-Net and MLP, achieve high-accuracy segmentation and classification of thermal anomalies, validating the potential of hybrid modeling in improving model adaptability [65].

Author Contributions: Conceptualization, A.D., X.Y. and X.L.; methodology, A.D. and C.W.; writing—original draft preparation, A.D. and Z.J.; writing—review and editing, X.Y., A.D. and J.W.; project administration, X.Y., J.Z., X.L. and C.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: No new data were created or analyzed in this study.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Leiter, A.; Veluswamy, R.R.; Wisnivesky, J.P. The global burden of lung cancer: Current status and future trends. *Nat. Rev. Clin. Oncol.* **2023**, *20*, 624–639. [CrossRef] [PubMed]
2. Jiang, H.; Ma, H.; Qian, W.; Gao, M.; Li, Y. An Automatic Detection System of Lung Nodule Based on Multigroup Patch-Based Deep Learning Network. *IEEE J. Biomed. Health Inform.* **2018**, *22*, 1227–1237. [CrossRef] [PubMed]
3. Ait Skourt, B.; El Hassani, A.; Majda, A. Lung CT Image Segmentation Using Deep Neural Networks. *Procedia Comput. Sci.* **2018**, *127*, 109–113. [CrossRef]
4. Sung, H.; Ferlay, J.; Siegel, R.L.; Laversanne, M.; Soerjomataram, I.; Jemal, A.; Bray, F. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J. Clin.* **2021**, *71*, 209–249. [CrossRef]
5. Song, Q.; Zhao, L.; Luo, X.; Dou, X. Using Deep Learning for Classification of Lung Nodules on Computed Tomography Images. *J. Healthc. Eng.* **2017**, *2017*, 8314740. [CrossRef]
6. Ettinger, D.S.; Wood, D.E.; Aisner, D.L.; Akerley, W.; Bauman, J.; Chirieac, L.R.; D'Amico, T.A.; DeCamp, M.M.; Dilling, T.J.; Dobelbower, M.; et al. Non–Small Cell Lung Cancer, Version 5.2017, NCCN Clinical Practice Guidelines in Oncology. *J. Natl. Compr. Cancer Netw.* **2017**, *15*, 504–535. [CrossRef]
7. Sun, W.; Zheng, B.; Qian, W. Computer aided lung cancer diagnosis with deep learning algorithms. In Proceedings of the SPIE Medical Imaging, San Diego, CA, USA, 27 February–3 March 2016; p. 97850Z.
8. Gordienko, Y.; Gang, P.; Hui, J.; Zeng, W.; Kochura, Y.; Alienin, O.; Rokovyi, O.; Stirenko, S. Deep Learning with Lung Segmentation and Bone Shadow Exclusion Techniques for Chest X-Ray Analysis of Lung Cancer. In Proceedings of the ICCSEEA 2018, Sydney, NSW, Australia, 22–23 December 2018; pp. 638–647.
9. NarİN, D.; Onur, T.Ö. The Effect of Hyper Parameters on the Classification of Lung Cancer Images Using Deep Learning Methods. *Erzincan Üniv. Fen Bilim. Enst. Derg.* **2022**, *15*, 258–268. [CrossRef]
10. Lyu, L. Lung Cancer Diagnosis Based on Convolutional Neural Networks Ensemble Model. In Proceedings of the 2021 2nd International Seminar on Artificial Intelligence, Networking and Information Technology (AINIT), Shanghai, China, 15–17 October 2021; pp. 360–367.
11. Ciompi, F.; Chung, K.; van Riel, S.J.; Setio, A.A.A.; Gerke, P.K.; Jacobs, C.; Scholten, E.T.; Schaefer-Prokop, C.; Wille, M.M.W.; Marchianò, A.; et al. Towards automatic pulmonary nodule management in lung cancer screening with deep learning. *Sci. Rep.* **2017**, *7*, 46479. [CrossRef]
12. Begum, S.; Sarkar, R.; Chakraborty, D.; Maulik, U. Identification of Biomarker on Biological and Gene Expression data using Fuzzy Preference Based Rough Set. *J. Intell. Syst.* **2021**, *30*, 130–141. [CrossRef]

13. Razzak, M.I.; Naz, S.; Zaib, A. Deep Learning for Medical Image Processing: Overview, Challenges and the Future. In *Classification in BioApps: Automation of Decision Making*; Dey, N., Ashour, A.S., Borra, S., Eds.; Springer International Publishing: Cham, Switzerland, 2018; pp. 323–350.
14. Naik, A.; Edla, D.R. Lung Nodule Classification on Computed Tomography Images Using Deep Learning. *Wirel. Pers. Commun.* **2021**, *116*, 655–690. [CrossRef]
15. Tang, H.; Kim, D.R.; Xie, X. Automated pulmonary nodule detection using 3D deep convolutional neural networks. In Proceedings of the 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018), Washington, DC, USA, 4–7 April 2018; pp. 523–526.
16. Prior, F.; Smith, K.; Sharma, A.; Kirby, J.; Tarbox, L.; Clark, K.; Bennett, W.; Nolan, T.; Freymann, J. The public cancer radiology imaging collections of The Cancer Imaging Archive. *Sci. Data* **2017**, *4*, 170124. [CrossRef] [PubMed]
17. Zhang, G.; Yang, Z.; Jiang, S. Automatic lung tumor segmentation from CT images using improved 3D densely connected UNet. *Med. Biol. Eng. Comput.* **2022**, *60*, 3311–3323. [CrossRef] [PubMed]
18. Rezvani, S.; Fateh, M.; Jalali, Y.; Fateh, A. FusionLungNet: Multi-scale fusion convolution with refinement network for lung CT image segmentation. *Biomed. Signal Process. Control.* **2025**, *107*, 107858. [CrossRef]
19. Dey, R.; Lu, Z.; Hong, Y. Diagnostic classification of lung nodules using 3D neural networks. In Proceedings of the 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018), Washington, DC, USA, 4–7 April 2018; pp. 774–778.
20. Abdani, S.R.; Zulkifley, M.A.; Shahrimin, M.I.; Zulkifley, N.H. Computer-Assisted Pterygium Screening System: A Review. *Diagnostics* **2022**, *12*, 639. [CrossRef]
21. Zulkifley, M.A.; Moubark, A.M.; Saputro, A.H.; Abdani, S.R. Automated Apple Recognition System Using Semantic Segmentation Networks with Group and Shuffle Operators. *Agriculture* **2022**, *12*, 756. [CrossRef]
22. Stofa, M.M.; Zulkifley, M.A.; Zainuri, M.A.A.M. Skin Lesions Classification and Segmentation: A Review. *Int. J. Adv. Comput. Sci. Appl.* **2021**, *12*, 532–540. [CrossRef]
23. Stofa, M.M.; Zulkifley, M.A.; Zainuri, M.A.A.M.; Ibrahim, A.A. U-Net with Atrous Spatial Pyramid Pooling for Skin Lesion Segmentation. In Proceedings of the 6th International Conference on Electrical, Control and Computer Engineering, Kuantan, Malaysia, 23 August 2021; pp. 1025–1033.
24. Xu, M.; Qi, S.; Yue, Y.; Teng, Y.; Xu, L.; Yao, Y.; Qian, W. Segmentation of lung parenchyma in CT images using CNN trained with the clustering algorithm generated dataset. *Biomed. Eng. OnLine* **2019**, *18*, 2. [CrossRef]
25. Liu, C.; Pang, M. Automatic lung segmentation based on image decomposition and wavelet transform. *Biomed. Signal Process. Control.* **2020**, *61*, 102032. [CrossRef]
26. Hu, Q.; Souza, L.F.d.F.; Holanda, G.B.; Alves, S.S.A.; dos S. Silva, F.H.; Han, T.; Rebouças Filho, P.P. An effective approach for CT lung segmentation using mask region-based convolutional neural networks. *Artif. Intell. Med.* **2020**, *103*, 101792. [CrossRef]
27. Mohammadreza, N.; David, B.; Tanveer, S.-M. Automated volumetric lung segmentation of thoracic CT images using fully convolutional neural network. In Proceedings of the Medical Imaging 2018: Computer-Aided Diagnosis, Houston, TX, USA, 10–15 February 2018; p. 105751J.
28. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In Proceedings of the Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015, Munich, Germany, 5–9 October 2015; pp. 234–241.
29. Balachandran, S.; Ranganathan, V. Semantic context-aware attention UNET for lung cancer segmentation and classification. *Int. J. Imaging Syst. Technol.* **2022**, *33*, 822–836. [CrossRef]
30. Albalawi, E.; Neal Joshua, E.S.; Joys, N.M.; Bhatia Khan, S.; Shaiba, H.; Ahmad, S.; Nazeer, J. Hybrid healthcare unit recommendation system using computational techniques with lung cancer segmentation. *Front. Med.* **2024**, *11*, 1429291. [CrossRef] [PubMed]
31. Khanna, A.; Londhe, N.D.; Gupta, S.; Semwal, A. A deep Residual U-Net convolutional neural network for automated lung segmentation in computed tomography images. *Biocybern. Biomed. Eng.* **2020**, *40*, 1314–1327. [CrossRef]
32. Comelli, A.; Coronnello, C.; Dahiya, N.; Benfante, V.; Palmucci, S.; Basile, A.; Vancheri, C.; Russo, G.; Yezzi, A.; Stefano, A. Lung Segmentation on High-Resolution Computerized Tomography Images Using Deep Learning: A Preliminary Step for Radiomics Studies. *J. Imaging* **2020**, *6*, 125. [CrossRef] [PubMed]
33. Gite, S.; Mishra, A.; Kotecha, K. Enhanced lung image segmentation using deep learning. *Neural Comput. Appl.* **2022**, *35*, 22839–22853. [CrossRef]
34. Yi, C.; Jiang, S.; Xiong, L.; Yang, J.; Shi, H.; Xiong, Q.; Hu, B.; Zhang, H. D-S-Net: An efficient dual-stage strategy for high-precision segmentation of gross tumor volumes in lung cancer CT images. *BMC Cancer* **2025**, *25*, 1387. [CrossRef]
35. Deepa, J.; Badhu Sasikala, L.; Indumathy, P.; Jerrin Simla, A. A novel lung cancer diagnosis model using hybrid convolution (2D/3D)-based adaptive DenseUnet with attention mechanism. *Netw. Comput. Neural Syst.* **2025**, 1–58. [CrossRef]

36. Leung, K.H.; Rowe, S.P.; Sadaghiani, M.S.; Leal, J.P.; Mena, E.; Choyke, P.L.; Du, Y.; Pomper, M.G. Deep Semisupervised Transfer Learning for Fully Automated Whole-Body Tumor Quantification and Prognosis of Cancer on PET/CT. *J. Nucl. Med.* **2024**, *65*, 643. [CrossRef]
37. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention Is All You Need. In Proceedings of the Neural Information Processing Systems 2017, Long Beach, CA, USA, 4–9 December 2017.
38. Hou, J.; Yan, C.; Li, R.; Huang, Q.; Fan, X.; Lin, F. Lung Nodule Segmentation Algorithm With SMR-UNet. *IEEE Access* **2023**, *11*, 34319–34331. [CrossRef]
39. Ma, J.; Yuan, G.; Guo, C.; Gang, X.; Zheng, M. SW-UNet: A U-Net fusing sliding window transformer block with CNN for segmentation of lung nodules. *Front. Med.* **2023**, *10*, 1273441. [CrossRef]
40. Lei, L.; Li, W. Transformer-based multi-task model for lung tumor segmentation and classification in CT images. *J. Radiat. Res. Appl. Sci.* **2025**, *18*, 101657. [CrossRef]
41. Shakeel, P.M.; Burhanuddin, M.A.; Desa, M.I. Lung cancer detection from CT image using improved profuse clustering and deep learning instantaneously trained neural networks. *Measurement* **2019**, *145*, 702–712. [CrossRef]
42. Liu, K.; Kang, G. Multiview convolutional neural networks for lung nodule classification. *Int. J. Imaging Syst. Technol.* **2017**, *27*, 12–22. [CrossRef]
43. da Silva, G.L.F.; da Silva Neto, O.P.; Silva, A.C.; de Paiva, A.C.; Gattass, M. Lung nodules diagnosis based on evolutionary convolutional neural network. *Multimed. Tools Appl.* **2017**, *76*, 19039–19055. [CrossRef]
44. Lakshmi G, G.; Nagaraj, P. Lung cancer detection and classification using optimized CNN features and Squeeze-Inception-ResNeXt model. *Comput. Biol. Chem.* **2025**, *117*, 108437. [CrossRef]
45. Kirienko, M.; Sollini, M.; Silvestri, G.; Mognetti, S.; Voulaz, E.; Antunovic, L.; Rossi, A.; Antiga, L.; Chiti, A. Convolutional Neural Networks Promising in Lung Cancer T-Parameter Assessment on Baseline FDG-PET/CT. *Contrast Media Mol. Imaging* **2018**, *2018*, 1382309. [CrossRef]
46. UrRehman, Z.; Qiang, Y.; Wang, L.; Shi, Y.; Yang, Q.; Khattak, S.U.; Aftab, R.; Zhao, J. Effective lung nodule detection using deep CNN with dual attention mechanisms. *Sci. Rep.* **2024**, *14*, 3934. [CrossRef]
47. Zheng, R.; Wen, H.; Zhu, F.; Lan, W. Attention-guided deep neural network with a multichannel architecture for lung nodule classification. *Heliyon* **2024**, *10*, e23508. [CrossRef]
48. Xu, Z.; Ren, H.; Zhou, W.; Liu, Z. ISANET: Non-small cell lung cancer classification and detection based on CNN and attention mechanism. *Biomed. Signal Process. Control.* **2022**, *77*, 103773. [CrossRef]
49. Li, Y.; Hui, L.; Wang, X.; Zou, L.; Chua, S. Lung nodule detection using a multi-scale convolutional neural network and global channel spatial attention mechanisms. *Sci. Rep.* **2025**, *15*, 12313. [CrossRef]
50. Cao, Z.; Li, R.; Yang, X.; Fang, L.; Li, Z.; Li, J. Multi-scale detection of pulmonary nodules by integrating attention mechanism. *Sci. Rep.* **2023**, *13*, 5517. [CrossRef]
51. Barbouchi, K.; El Hamdi, D.; Elouedi, I.; Aïcha, T.B.; Echi, A.K.; Slim, I. A transformer-based deep neural network for detection and classification of lung cancer via PET/CT images. *Int. J. Imaging Syst. Technol.* **2023**, *33*, 1383–1395. [CrossRef]
52. Nishigaki, D.; Suzuki, Y.; Watabe, T.; Katayama, D.; Kato, H.; Wataya, T.; Kita, K.; Sato, J.; Tomiyama, N.; Kido, S. Vision transformer to differentiate between benign and malignant slices in ^{18}F-FDG PET/CT. *Sci. Rep.* **2024**, *14*, 8334. [CrossRef] [PubMed]
53. Sun, R.; Pang, Y.; Li, W. Efficient Lung Cancer Image Classification and Segmentation Algorithm Based on an Improved Swin Transformer. *Electronics* **2023**, *12*, 1024. [CrossRef]
54. Faizi, M.K.; Qiang, Y.; Wei, Y.; Qiao, Y.; Zhao, J.; Aftab, R.; Urrehman, Z. Deep learning-based lung cancer classification of CT images. *BMC Cancer* **2025**, *25*, 1056. [CrossRef]
55. Quasar, S.R.; Sharma, R.; Mittal, A.; Sharma, M.; Agarwal, D.; de La Torre Díez, I. Ensemble methods for computed tomography scan images to improve lung cancer detection and classification. *Multimed. Tools Appl.* **2024**, *83*, 52867–52897. [CrossRef]
56. Gautam, N.; Basu, A.; Sarkar, R. Lung cancer detection from thoracic CT scans using an ensemble of deep learning models. *Neural Comput. Appl.* **2024**, *36*, 2459–2477. [CrossRef]
57. R, N.; C.M, V. Transfer learning based deep architecture for lung cancer classification using CT image with pattern and entropy based feature set. *Sci. Rep.* **2025**, *15*, 28283. [CrossRef]
58. Awudan, G.; Junxiang, Y.; Abudoukelimu, M.; Mengfei, W.; Abudoukelimu, H.; Abulizi, A. A Review of Deep Learning-Based Segmentation and Classification of Pulmonary Nodules in CT Images. *Comput. Eng. Appl.* **2025**, *61*, 14–35.
59. Cui, S.; Ming, S.; Lin, Y.; Chen, F.; Shen, Q.; Li, H.; Chen, G.; Gong, X.; Wang, H. Development and clinical application of deep learning model for lung nodules screening on CT images. *Sci. Rep.* **2020**, *10*, 13657. [CrossRef]
60. Wang, S.; Zhou, M.; Liu, Z.; Liu, Z.; Gu, D.; Zang, Y.; Dong, D.; Gevaert, O.; Tian, J. Central focused convolutional neural networks: Developing a data-driven model for lung nodule segmentation. *Med. Image Anal.* **2017**, *40*, 172–183. [CrossRef]
61. Shen, W.; Zhou, M.; Yang, F.; Yu, D.; Dong, D.; Yang, C.; Zang, Y.; Tian, J. Multi-crop Convolutional Neural Networks for lung nodule malignancy suspiciousness classification. *Pattern Recognit.* **2017**, *61*, 663–673. [CrossRef]

62. Qiangqiang, B.; Siyuan, T.; Yu, G. A Review of Challenging Issues in Deep Learning-Based Pulmonary Nodule Detection. *Comput. Eng. Appl.* **2024**, *60*, 18–31.
63. Yin, J.; Su, D. Research Progress on the Role of Deep Learning Technology in Breast and Breast Cancer Image Segmentation. *China J. Cancer Prev. Treat.* **2023**, *15*, 587–592.
64. Pratticò, D.; Laganà, F. Infrared Thermographic Signal Analysis of Bioactive Edible Oils Using CNNs for Quality Assessment. *Signals* **2025**, *6*, 38. [CrossRef]
65. Pratticò, D.; Carlo, D.D.; Silipo, G.; Laganà, F. Hybrid FEM-AI Approach for Thermographic Monitoring of Biomedical Electronic Devices. *Computers* **2025**, *14*, 344. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.