



Landscape of 2D Deep Learning Segmentation Networks Applied to CT Scan from Lung Cancer Patients: A Systematic Review

Somayeh Sadat Mehrnia^{1,2} · Zhino Safahi^{2,3} · Amin Mousavi² · Fatemeh Panahandeh² · Arezoo Farmani² · Ren Yuan^{4,5} · Arman Rahmim^{4,6,7} · Mohammad R. Salmanpour^{2,4,6} 

Received: 7 December 2024 / Revised: 6 February 2025 / Accepted: 16 February 2025 / Published online: 4 March 2025
© The Author(s) under exclusive licence to Society for Imaging Informatics in Medicine 2025

Abstract

Background The increasing rates of lung cancer emphasize the need for early detection through computed tomography (CT) scans, enhanced by deep learning (DL) to improve diagnosis, treatment, and patient survival. This review examines current and prospective applications of 2D- DL networks in lung cancer CT segmentation, summarizing research, highlighting essential concepts and gaps; **Methods:** Following the Preferred Reporting Items for Systematic Reviews and Meta-Analysis guidelines, a systematic search of peer-reviewed studies from 01/2020 to 12/2024 on data-driven population segmentation using structured data was conducted across databases like Google Scholar, PubMed, Science Direct, IEEE (Institute of Electrical and Electronics Engineers) and ACM (Association for Computing Machinery) library. 124 studies met the inclusion criteria and were analyzed. **Results:** The LIDC-LIDR dataset was the most frequently used; The finding particularly relies on supervised learning with labeled data. The UNet model and its variants were the most frequently used models in medical image segmentation, achieving Dice Similarity Coefficients (DSC) of up to 0.9999. The reviewed studies primarily exhibit significant gaps in addressing class imbalances (67%), underuse of cross-validation (21%), and poor model stability evaluations (3%). Additionally, 88% failed to address the missing data, and generalizability concerns were only discussed in 34% of cases. **Conclusions:** The review emphasizes the importance of Convolutional Neural Networks, particularly UNet, in lung CT analysis and advocates for a combined 2D/3D modeling approach. It also highlights the need for larger, diverse datasets and the exploration of semi-supervised and unsupervised learning to enhance automated lung cancer diagnosis and early detection.

Keywords 2D Segmentation · Deep Learning · Computed Tomography · Lung Cancer · Review Article

Introduction

Lung cancer is a major public health concern due to its high mortality rate [1]. In 2024, the American Cancer Society [2] projects 235,000 new lung cancer cases and 125,000 deaths in the U.S. As the leading cause of cancer death, lung cancer underscores the urgent need for advanced research. Continued studies are crucial for developing more effective treatments, improving early detection methods, and ultimately saving lives in this critical area of public health. Despite declining mortality rates, lung cancer remains the leading cause of cancer death in the U.S. A recent study [3] highlights that mortality declines have outpaced incidence declines in men (5.0% vs. 2.6% annually) and women (4.3% vs. 1.1% annually), driven by advances in treatment, earlier diagnoses, and expanded access to care. However, disparities persist, with Native Americans experiencing the

✉ Mohammad R. Salmanpour
msalman@bccrc.ca

¹ Department of Integrative Oncology, Breast Cancer Research Center, Motamed Cancer Institute, ACECR, Tehran, Iran

² Technological Virtual Collaboration (TECVICO Corp.), Vancouver, BC, Canada

³ Department of Computer Engineering, University of Kurdistan, Sanandaj, Iran

⁴ Department of Radiology, University of British Columbia, Vancouver, BC, Canada

⁵ BC Cancer, Vancouver Center, Vancouver, BC, Canada

⁶ Department of Integrative Oncology, BC Cancer Research Institute, Vancouver, BC V5Z 1L3, Canada

⁷ Department of Physics & Astronomy, University of British Columbia, Vancouver, BC, Canada

slowest decline and states like Mississippi and Kentucky facing mortality rates two to three times higher than western states, underscoring the need for targeted tobacco control, expanded screening, and equitable access to advanced treatments.

Another study [4] using 2015–2019 data found that lung cancer incidence in women aged 35–54 years was equal to or higher than in men in 40 of 51 U.S. states, with significant differences in 20 states ($p < 0.05$). This trend was not explained by historical smoking prevalence, which was lower or similar in women in 33–34 states, suggesting other contributing factors. Another study on the worldwide lung cancer population [5] projects that in 2035, incidence rates (ASRs) will decrease by 23% in males (35.8 to 27.6 per 100,000) but increase by 2% in females (16.8 to 17.1). Female ASRs will rise sharply in many countries, nearing male rates in Ireland, Norway, the UK, Canada, and the U.S. The highest ASRs are expected in Belarus (49.3, males) and Denmark (36.8, females). In 2035, new cases in 40 countries are projected to rise by 65%, reaching 2.17 million, with China bearing the largest burden, highlighting the urgent need for strengthened global lung cancer control efforts. Early lung cancer detection ultimately contributes to a successful treatment and prevention and may increase the survival rate in humans [6, 7]. Medical imaging provides 2D (dimensional) or 3D views aiding diagnosis, prognosis, and survival prediction [8–14]. While CT (Computed Tomography) scans provide detailed images used to detect cancer, visual interpretations by radiologists can be challenging due to the difficulty in distinguishing malignant and benign tumors.

visual interpretations by radiologists can be challenging due to difficulty distinguishing malignant and benign tumors. [15, 16]. Underreading errors occur in 42% of cases, and satisfaction of search errors in 22%. Cognitive errors contribute to 9% of diagnostic mistakes. These findings highlight the need for improved diagnostic processes, including artificial intelligence (AI) tools to reduce these errors [17]. Medical image segmentation methods have the potential to improve diagnostic outcomes by accurately delineating regions of interest (ROI) and reducing ambiguity in clinical images [9, 18, 19]. Computer-assisted tumor segmentation enables accurate, automated Evaluation Criteria in Solid Tumors (RECIST) and volumetric analyses, leading to more precise assessments of tumor response compared to manual measurements [20].

Volumetric analysis provides additional tumor burden and growth pattern information, useful for assessing treatment response and planning [21–24]. CT imaging is widely used for segmentation due to its accessibility and cost-effectiveness. AI, especially deep learning (DL), automates analysis, saving time and reducing variability and errors in manual measurements [11, 23, 25].

DL advancements have transformed medical image segmentation, achieving radiologist-level accuracy with minimal pre-processing. Models like Convolutional Neural Networks (CNNs), Residual Neural Networks (RNNs), and Generative Adversarial Networks (GANs) efficiently analyze raw images, achieving high accuracy and speed, which has led to widespread adoption [26–29], aiding in tumor assessment and treatment evaluation, and enhancing lung cancer screening by improving nodule detection and classification [30, 31]. Recent reviews on DL-based segmentation methods have mainly focused on either the evolution of deep network architectures or techniques specific to certain architectures. For instance, Zhang et al. [32] delved into sophisticated DL techniques for COVID-19 CT image segmentation, highlighting Transformer structures and categorizing segmentation approaches into encoder-decoder, attention-based, multi-scale/pyramid, and Transformer-based models. Qin et al. [33] explored the automated segmentation of retinal arteries, discussing both traditional and DL approaches with a focus on supervised learning, including CNN, GAN, and UNet-based methods, while offering insights into the strengths and weaknesses of these techniques. Han et al. [34] investigated semi-supervised medical image segmentation, organizing studies into pseudo-labeling, consistency regularization, GAN, contrastive learning, and hybrid methods, and tested several on popular datasets. Khajuria et al. [35] examined reinforcement learning in cancer diagnosis and treatment, reviewing various techniques and challenges. Wang et al. [36] conducted a systematic review on the efficacy of DL models, especially UNet, in identifying brain metastases in Magnetic Resonance Imaging (MRI) images, noting their high segmentation accuracy and sensitivity across different MRI systems. This demonstrates a growing diversity in methods and applications for DL in medical image segmentation.

A review [37] explored federated learning methods to enhance global segmentation performance while maintaining privacy, focusing on brain tumor segmentation and a client-based training strategy. Jiang et al. [38] reviewed deep neural network algorithms for segmenting pulmonary nodules, addressing nodule diversity, segmentation clarity, and environmental factors, and analyzed open-source models using LIDC (The Lung Image Database Consortium image collection) and LUNA 16 (Lung Nodule Analysis 2016) datasets. Ciceri et al. [39] concentrated on DL algorithms for fetal brain segmentation, categorizing algorithms by target structures and highlighting research gaps and future directions for managing fetal MR images. Chen et al. [40] surveyed DL approaches for cerebrovascular segmentation since 2015, covering models like sliding window, UNet, CNNs, and Transformers, and discussed their evolution, challenges, and research opportunities, aiming to guide future studies. Wang et al. [41] reviewed transformer-based

methods for brain tumor segmentation, dividing them into pure and hybrid transformer types, and discussed their innovations and benefits. Zhang et al. [29] focused on automating pancreas segmentation using DL for CT and MRI scans, aiming to aid medical image analysis and identify future research avenues. Hung et al. [42] analyzed DL methods for segmenting crucial features in carotid artery ultrasound images, vital for clinical diagnosis. Abbasi [43] examined the progress in DL for stroke lesion segmentation in MRI and CT scans, comparing the performance and noting the strengths and limitations of each modality. Zhang et al. [44] evaluated the effectiveness of DL in segmenting teeth, mandibular canals, and alveolar bones in dental CBCT images, highlighting issues such as dataset inconsistency and lack of method transparency. Xun et al. [45] explored GAN-based architectures for medical image segmentation, highlighting their potential to improve accuracy despite challenges like instability and interpretability issues, which are crucial for wider clinical acceptance. Gu et al. [46] reviewed 2D instance segmentation with DL, focusing on fully-, weakly-, and semi-supervised methods, especially two-stage methods for performance, and provided benchmarks with datasets and metrics for evaluating models.. Instance segmentation is a computer vision task that involves identifying and separating individual objects within an image, including detecting the boundaries of each object and assigning a unique label to each object instance, resulting in a pixel-wise segmentation map of the image [47]. Marques et al. [48] conducted a systematic review on automating Optic Nerve Head segmentation in Optical Coherence Tomography, noting inconsistencies and the need for standardized methodologies for clinical trust and integration. Kaur et al. [49] traced the evolution of medical image segmentation from traditional methods to DL, pointing out the trend towards two-step DL models for multi-organ segmentation and calling for more research on less-studied organs.

The above-mentioned studies show that much of the research in medical image segmentation, especially for CT, centers on 3D techniques. However, the importance of 2D segmentation, particularly in CT imaging, is often overshadowed by the growing interest in 3D methods, despite the critical role that 2D segmentation plays in lung cancer diagnosis. Thanoon et al. [50] and Carvalho et al. [51] emphasize that 2D segmentation offers significant clinical benefits by allowing the processing of individual CT slices, which clinicians use to identify disease patterns with high precision. 2D segmentation models are effective for CT image analysis, offering adaptability, cost-efficiency, and diagnostic accuracy [52]. Despite additional studies on 3D techniques, research [30, 53] has shown 2D segmentation techniques are effective for lung tissue analysis, especially as significant advanced research in 2D-image DL techniques (e.g. natural sceneries) can be utilized and fine-tuned for medical imaging

research. The results of the research indicated that employing multiple 2D models lead to quicker training speeds and enabled matched or excelling segmentation performances for 3D convolution models.

These systematic reviews distinguish itself by focusing specifically on the landscape of 2D DL segmentation networks applied to CT scans from lung cancer patients. Unlike previous reviews that may generalize across various DL techniques, we provide a detailed analysis of 2D segmentation networks, their architectures, and their specific applications in lung cancer imaging. Additionally, we highlight the strengths and limitations of existing models. This targeted approach allows us to identify gaps in the current literature and propose future research directions that are specifically relevant to 2D segmentation methods in lung cancer diagnosis. Addressing a gap in studies on 2D segmentation for lung cancer imaging, it explores recent 2D auto-segmentation techniques in DL architectures for CT scans, highlighting developments, advantages, and challenges from 2020 to 2024. Key questions include identifying DL methods for lung cancer segmentation, optimizing 2D segmentation with loss functions, assessing recent advances, and evaluating the benefits and limitations of 2D segmentation in lung cancer analysis. Through a systematic review of recent literature from 2020 to 2024, we focus on answering the following key research questions: (i) what are the most commonly used 2D segmentation models for lung cancer detection in CT images, and how do they perform in terms of accuracy and efficiency? (ii) What are the current gaps in research, and what are the prospects for improving 2D segmentation in lung cancer diagnosis? (iii) What are the advantages and limitations of 2D segmentation methods in lung cancer CT analysis? (iv) What are the strengths and weaknesses of commonly used 2D segmentation architectures, such as UNet and its variants, in segmenting lung cancer lesions from CT images?

Materials and Methods

This article undertakes a systematic review guided by the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) framework. The criteria for study selection and the strategies employed to gather these studies are defined and will be elaborated upon in the sections that follow.

Eligibility Criteria for Study Inclusion

Eligibility Criteria for Study Selection: For this systematic review, we set clear eligibility criteria to select relevant studies. This review included original research papers and conference proceedings on 2D DL segmenta-

tion methods in CT scans for lung cancer diagnosis. Filters were applied to English-language articles published between 2020 and the end of 2024.

Inclusion Criteria: The criteria for inclusion were: "Segmentation" in the title or abstract with at least two related search terms; Studies involving human subjects; CT images only; and A focus on 2D DL segmentation.

Exclusion Criteria: Rigorous exclusion criteria filtered out non-CT imaging studies, preclinical/animal studies, testing-only datasets, case reports, reviews, and articles related to radiomics, COVID, or in non-English languages.

Final Study Selection: Researchers carefully reviewed all articles for relevance, including only studies with 2D DL models for lung cancer segmentation in the final analysis. This stringent selection resulted in 124 studies for detailed review, contributing valuable insights into DL-driven 2D segmentation techniques for accurate, automated lung tumor segmentation in CT imaging.

Search Strategy for Information Source

To locate pertinent studies for a systematic review, an extensive search was carried out from January 1, 2020, to December 31, 2024. A variety of platforms, such as PubMed (pubmed.ncbi.nlm.nih.gov), Science Direct (sciencedirect.com), and Google Scholar (scholar.google.com), IEEE (Institute of Electrical and Electronics Engineers) (<https://ieeexplore.ieee.org>), and ACM (Association for Computing Machinery) Digital Library (<https://dl.acm.org>) were employed to access high-quality academic materials including articles from scientific journals and proceedings from conferences.

Study Selection and Data Extraction Process

The study selection process for this systematic review is outlined in the PRISMA diagram shown in Fig. 1 demonstrating the structured and methodical approach used. PubMed yielded 779 papers when searched with the keywords: "segmentation" AND "Lung" AND "CT" OR "computed tomography" OR "deep learning" NOT "COVID", as shown in Table 1. Science Direct returned 3304 results using the search terms "CT" AND "lung OR cancer OR tumor" NOT "COVID" OR "deep learning." In the ACM library search, 94 articles were found based on the search title: segmentation and all: (CT or lung or computed tomography) AND (E-publication Date: (01/01/2020 to 12/31/2024). IEEE produced 759 papers using the same search terms. Finally, Google Scholar returned 41,700 results for the search terms "CT" + "segmentation" + "lung". A more in-depth search of titles within Google Scholar identified 1006 papers that included all of these keywords in the title: "Lung

CT segmentation OR cancer OR tumor OR deep learning. To focus on segmentation techniques, "segmentation" was required in article titles. An initial title search and abstract review included the term "lung" to capture studies on lung-related topics like nodules and tumors, while excluding terms like "infection," "COVID," and "organ." Full texts were reviewed to confirm the use of deep learning models beyond pre-processing, and model descriptions were checked to ensure a focus on 2D segmentation for lung cancer CT diagnosis. Four independent researchers screened articles by titles and abstracts, using EndNote to organize references and remove duplicates, with a fifth reviewer resolving any discrepancies for an unbiased selection, we included studies that incorporated both 2D and 3D segmentation. Out of the 124 studies reviewed, 17 included both 2D and 3D models, with our primary focus on analyzing the 2D models. Only original research on 2D segmentation in lung cancer CT imaging meeting the inclusion criteria proceeded to full review.

Comprehensive Methodology for Evaluating Research Biases

We conducted a comprehensive review of studies, the most widely used in 2D segmentation of lung cancer in CT images, focusing on identifying research biases. Adapting the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD) guidelines specifically for DL, we developed targeted questions to rigorously evaluate each study's strengths, weaknesses, and potential biases.

Question Design for Bias Analysis

To systematically assess critical bias aspects of model development in DL-based studies, we designed Yes/No questions covering key areas where bias might happen. We applied these questions to 124 studies, using the responses to evaluate each study systematically. This helped us identify trends such as gaps in socioeconomic considerations, variability in cross-validation methods, and limited discussions on generalizability. Our findings highlight best practices and gaps, offering insights to enhance UNet model robustness and clinical applicability. These questions addressed key areas, as outlined below and in Table 3. i) Data Sources: Assessed whether datasets were clearly identified as multi-center, single-center, or public, essential for reproducibility and variability understanding. ii) Dataset Splitting: Focused on training, validation, and test set divisions to ensure fair model evaluation. iii) Class Distribution: Checked for class imbalance issues, especially relevant for distinguishing cancerous from non-cancerous areas. iv) Preprocessing Techniques: Evaluated consistency in preprocessing (normalization,

Fig. 1 Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) flowchart illustrating the selection process of relevant articles from prominent scientific databases. The symbol 'n' represents the number of articles at each stage of the process

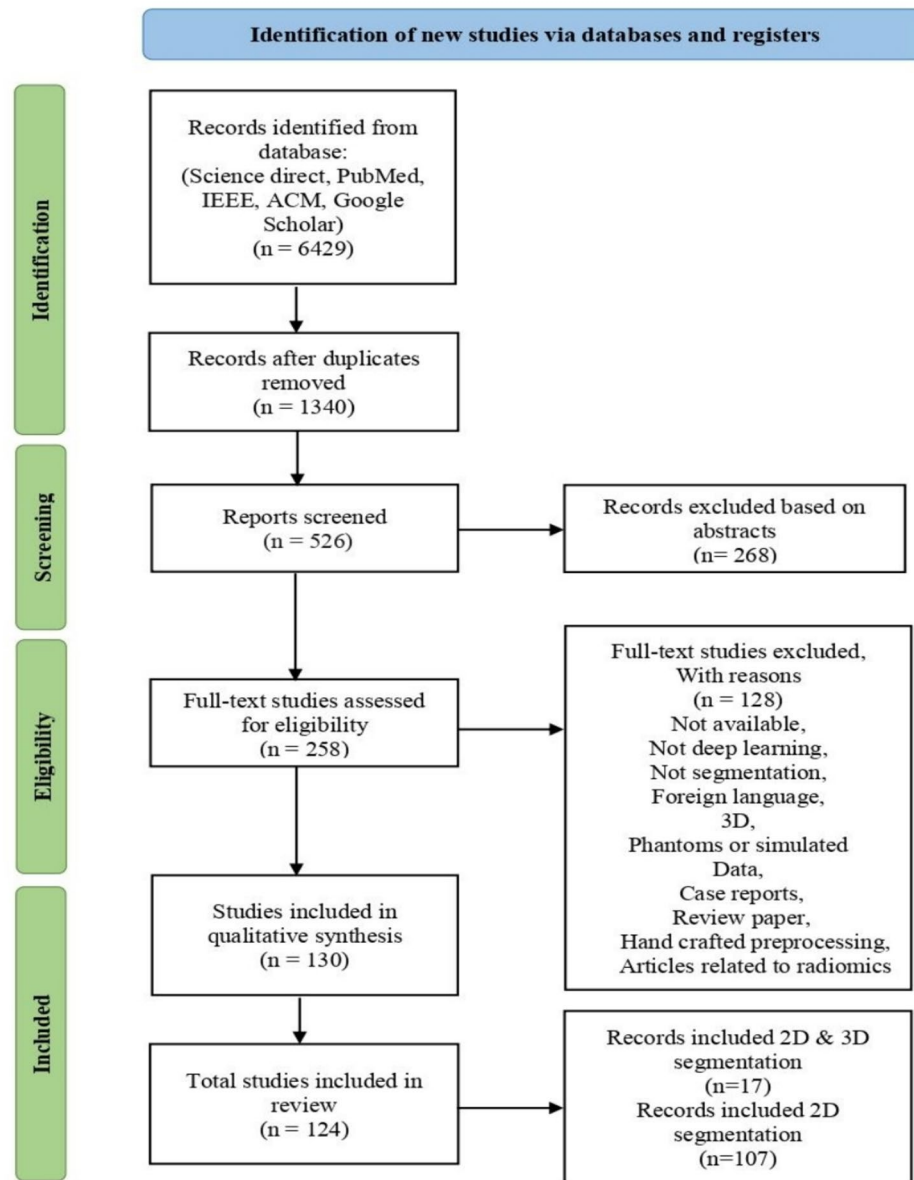


Table 1 Study Selection Process for Systematic Review on Lung CT Segmentation

Search Database	Search Terms Used (2020–2025)
PubMed	"segmentation" AND "Lung" AND "CT" OR "computed tomography" OR "deep learning" NOT "COVID" AND E-publication Date: (01/01/2020 to 01/01/2025)
Science Direct	"CT" OR "computed tomography" AND "lung OR cancer OR tumor" NOT "COVID" OR "deep learning" AND E-publication Date: (01/01/2020 to 01/01/2025)
ACM Library	segmentation AND all: (CT OR "computed tomography" OR lung) AND (E-publication Date: (01/01/2020 to 01/01/2025))
IEEE	"CT" OR "computed tomography" AND "lung OR cancer OR tumor" NOT "COVID" OR "deep learning" AND E-publication Date: (01/01/2020 to 01/01/2025)
Google Scholar	"CT" OR "computed tomography" + "segmentation" + "lung" AND E-publication Date: (01/01/2020 to 01/01/2025)
Google Scholar—Title Search	"Lung CT OR computed tomography segmentation OR cancer OR tumor OR deep learning" AND E-publication Date: (01/01/2020 to 01/01/2025)
Article Screening	Title and Abstract Review, Full Text Review AND E-publication Date: (01/01/2020 to 01/01/2025) Titles were screened to ensure relevance to lung CT segmentation using deep learning; full texts were reviewed to confirm the use of 2D segmentation techniques for diagnosis

cropping, resizing). v) Data Augmentation: Examined the use of augmentation to improve model generalization. vi) Comparison of Models: Reviewed studies comparing models for strengths and weaknesses. vii) Cross-Validation: Ensured proper validation methods, like leave-one-center-out, to prevent overfitting. viii) Hyperparameter Tuning: Reviewed tuning efforts for performance optimization. ix) Model Stability: Assessed stability across random seeds for performance consistency. x) Evaluation Metrics: Evaluated breadth in metrics beyond accuracy and Dice, such as sensitivity, specificity, and AUC. xi) Generalizability: Looked for validations on external datasets to gauge real-world applicability. xii) Case Diversity: Checked for a range of disease severity in cases to improve model versatility. xiii) Socioeconomic and Geographic Factors: Noted studies considering socioeconomic and geographic variability. xiv) Eligibility Criteria: Reviewed clarity in patient selection and treatment descriptions. xv) Handling of Missing Data: Assessed transparency in handling missing data. xvi) Training and Evaluation Datasets: Identified differences between training and evaluation sets to avoid overfitting. xvii) Ground Truth Labeling Consistency: Ensured consistency and minimized annotator bias in labels. xviii) Expert Involvement in Labeling: Checked for multiple annotators to reduce subjectivity. xix) Multi-Modal Data Training: Reviewed use of data from different imaging modalities for robustness. xx) Ablation Study Evaluation: Assessed whether the study performed an ablation study to evaluate the impact of removing certain components or features. xxi) External Dataset Validation:

Ensured that the study used external datasets for validation to improve model generalizability. xxii) Statistical Analysis of Metric Differences: Reviewed whether the study conducted statistical.

Results

The study also presents findings on the effectiveness of 2D segmentation in CT scans for lung cancer diagnosis, using visual aids for clarity and emphasizing the potential for practical applications and further research.

Distribution of Published DL Articles in 2D Lung Cancer Segmentation and Analysis of DL-Based Segmentation Models Across Different Publishers

Figure 2a features interactive pie charts from 2020 to 2024, showing research publication volumes by country. China's dominance in this area is evident with 52 publications, making up 31% of the total observed research output. China and India have contributed to over 50% of the publications in this dataset over the past few years, highlighting their substantial role in advancing medical imaging research. Figure 2b presents a pie chart illustrating the distribution of research papers on 2D segmentation across the top eight publishers. Each segment represents a specific publisher, with its size indicating the proportion of research papers they have published. IEEE leads with approximately 54% of total

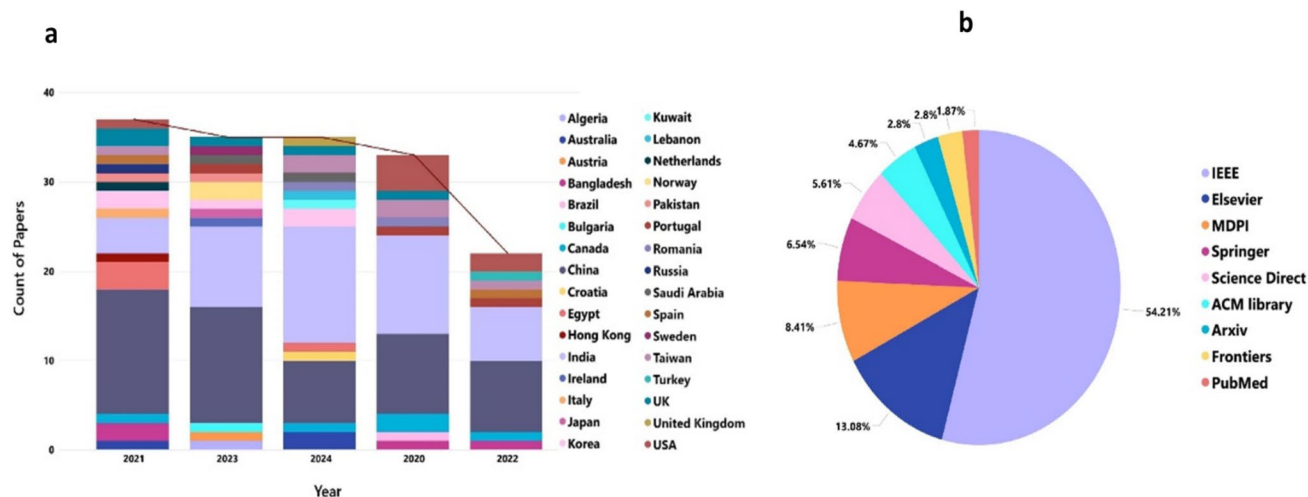


Fig. 2 **a** Distribution of published articles by year and country This bar chart represents the number of research contributions from different countries across the years 2020 to 2024. The stacked bars show the distribution of contributions from various countries, including Australia, Austria, Bangladesh, Brazil, Canada, and others. The line plot overlaid on the chart indicates a trend in the overall contributions, peaking in 2021 and showing a decline through to 2024. **b**

Publication distribution by publishers This pie chart displays the distribution of research publications across various publishers The chart shows that IEEE holds the largest share with ~54%, followed by Elsevier at ~13%, and other publishers such as MDPI, Springer, Science Direct, ACM Library, Arxiv, Frontiers, and PubMed making up smaller portions of the total distribution

publications, highlighting its key role in research dissemination in this field. Elsevier follows with around 13%, MDPI with roughly 8%, Springer with approximately 7%, Science-Direct with roughly 6%. This concentration of publications among a few major publishers underscores their substantial influence in the field.

Distribution of 2D Supervised, Semi-Supervised, and Unsupervised Methods in Lung Cancer Segmentation Tasks and Comparison of 2D vs. Combined 2D & 3D Segmentation Techniques in Lung Cancer Research

Figure 3a showcases a bar graph analysis of learning approaches in 2D segmentation studies for lung cancer, revealing a dominant use of supervised learning in 124 instances, accounting for 90% of all methods applied. In contrast, semi-supervised and unsupervised learning methods are minimally used, each found in only 9 studies. This stark differentiation highlights the field's heavy reliance on supervised learning, which prefers models trained on labeled data. The limited use of semi-supervised and unsupervised learning suggests challenges like scarce unlabeled data or the complexity of applying these approaches to medical imagery. Despite the dominance of supervised learning, the findings indicate room for growth in semi-supervised and unsupervised learning methods, potentially broadening data use to enhance lung cancer diagnosis technology. In

this review, we included studies that contained both 2D and 3D segmentation. From a total of 124 studies, 17 featured both 2D and 3D models, and we focused on the 2D models for further analysis. Figure 3b from our systematic review illustrates the distribution of 2D versus combined 2D and 3D segmentation approaches in the study of lung cancer. Out of the total, 124 studies employed 2D segmentation methods, which represents 86.0% of the total count. In contrast, 17 studies utilized a combination of 2D and 3D segmentation techniques.

Distribution of DL Model Usages in 2D Lung Cancer Segmentation

Figure 4 illustrates the frequency of top models used in 2D segmentation studies for lung cancer. The flow diagram emphasizes the dominance of the UNet model, utilized in 40% of studies, making it the preferred choice. CNN follows at 23%, with Attention mechanisms at 12% and Residual Network (ResNet) at 9%. Other models, including Fully Convolutional Network (FCN), Semantic Segmentation Network (SegNet), Mask Region-based CNN (Mask-RCNN), Region-Based CNN (RCNN), and Transformer, each have frequencies between 2 and 3%. This data highlights the research community's strong preference for UNet due to its effectiveness in medical image segmentation, particularly for lung cancer. The significant gap between UNet and other models underscores its

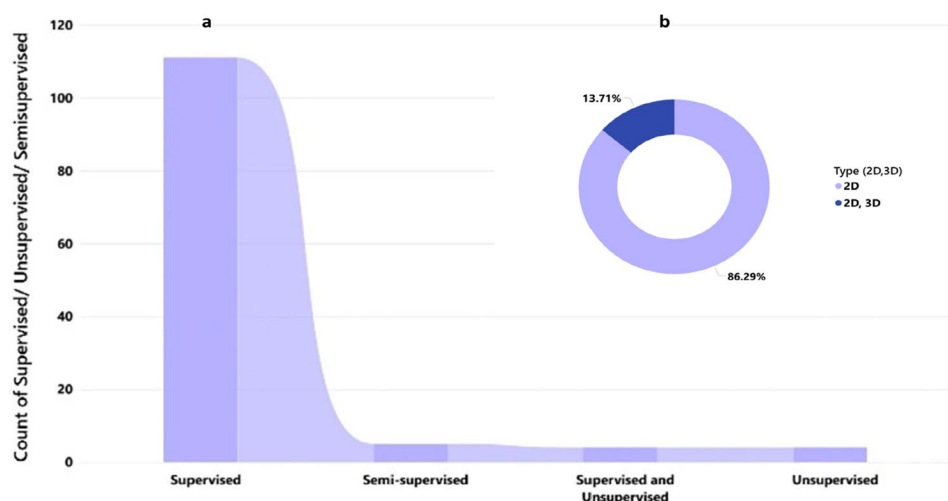


Fig. 3 **a** Comparative frequency of supervised, semi-supervised, and unsupervised methods. Bar chart illustrates the distribution of segmentation model types used in lung cancer diagnosis research. Supervised had the highest count of supervised/unsupervised/semi-supervised at 115, followed by semi-supervised, supervised and unsupervised, and unsupervised. Supervised learning had the highest count among supervised, unsupervised, and semi-supervised approaches at 102. This was followed by semi-supervised learning and then unsupervised learning. This emphasizes the dominant reli-

ance on supervised learning approaches in the field. **b** Proportion of studies utilizing 2D versus combined 2D and 3D segmentation techniques in lung cancer in this research. The pie chart shows the distribution of 2D versus combined 2D and 3D segmentation approaches in lung cancer studies. 2D segmentation methods are used in ~86% of the studies, while a combination of 2D and 3D segmentation techniques is employed in ~14%. This highlights the prevalent use of 2D approaches in current research

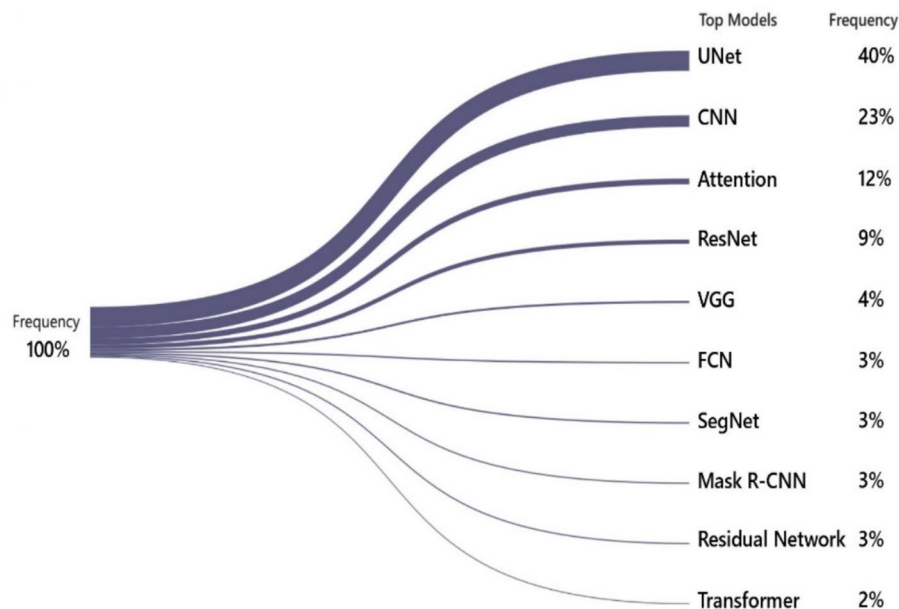


Fig. 4 Distribution of model usage in 2D segmentation research for lung cancer. This image illustrates the frequency distribution of popular deep learning (DL) models used in various tasks, with UNet being the most frequently utilized model at 40%, followed by Convolutional Neural Network (CNN) at 23%, Attention at 12%, and other models such as Residual Network (ResNet), Fully Convolutional Network

(FCN), Mask Region-based Convolutional Neural Network (Mask-RCNN), Semantic Segmentation Network (SegNet), Visual Geometry Group (VGG), Region-Based Convolutional Neural Networks (RCNN), and Transformer with lower frequencies. The chart highlights the dominant role of UNet and CNN in the selected domain

reliability, while the variety of models reflects ongoing exploration of DL architectures to improve segmentation accuracy and performance. The emerging interest in Transformer models, although less common, suggests a shift towards advanced architectures for complex medical imaging challenges.

In the domain of computer-aided diagnosis and image segmentation for lung cancer detection, this review delves into leading DL models that have revolutionized the field by improving the accuracy, precision, and efficiency of identifying malignant areas in medical images. Notably, the following models stand out as pivotal advancements in 2D segmentation for lung cancer in CT images. Additional descriptions and categorizations of these models, along with related studies on 2D DL segmentation methods, are provided in Supplemental File 1. Tables S1 and S2 in Online Resource Materials provide a detailed summary of the various 2D-DL networks employed for lung cancer segmentation in CT scans, including the segmentation approach utilized, their respective advantages and limitations, as well as the datasets leveraged for training and evaluation. Additionally, Table 2 provides a summary of Supplemental File 1 (Supplemental Table S2) in the Online Resource.

Usage of Evaluation Metrics in DL-Based 2D Lung Cancer Segmentation Tasks

Figure 5 uses a bar chart visualization to depict the range of evaluation metrics for 2D segmentation models in lung cancer diagnosis. This bar chart represents the frequency of usage of various evaluation metrics in medical imaging research. The most frequently used metric is Dice Similarity, followed by Accuracy, Sensitivity, and Intersection over Union (IoU). Other commonly used metrics include Precision, Specificity, and Jaccard Index, with fewer studies utilizing metrics like Hausdorff Distance, False Positive, and False Negative.

Figure 6 outlines the top six varieties and frequency of loss functions used in developing 2D segmentation models for lung cancer, as identified in this systematic review. This visualization presents the results of applying different loss functions to a 2D segmentation task for lung cancer images, showcasing the frequency of each function's use. The Dice Similarity Loss, Binary Cross-Entropy Loss, and Cross-Entropy Loss occupy the largest portions of the chart, indicating their prominent roles in driving the segmentation models to learn the spatial and structural relationships within the lung cancer images. Other loss functions, such

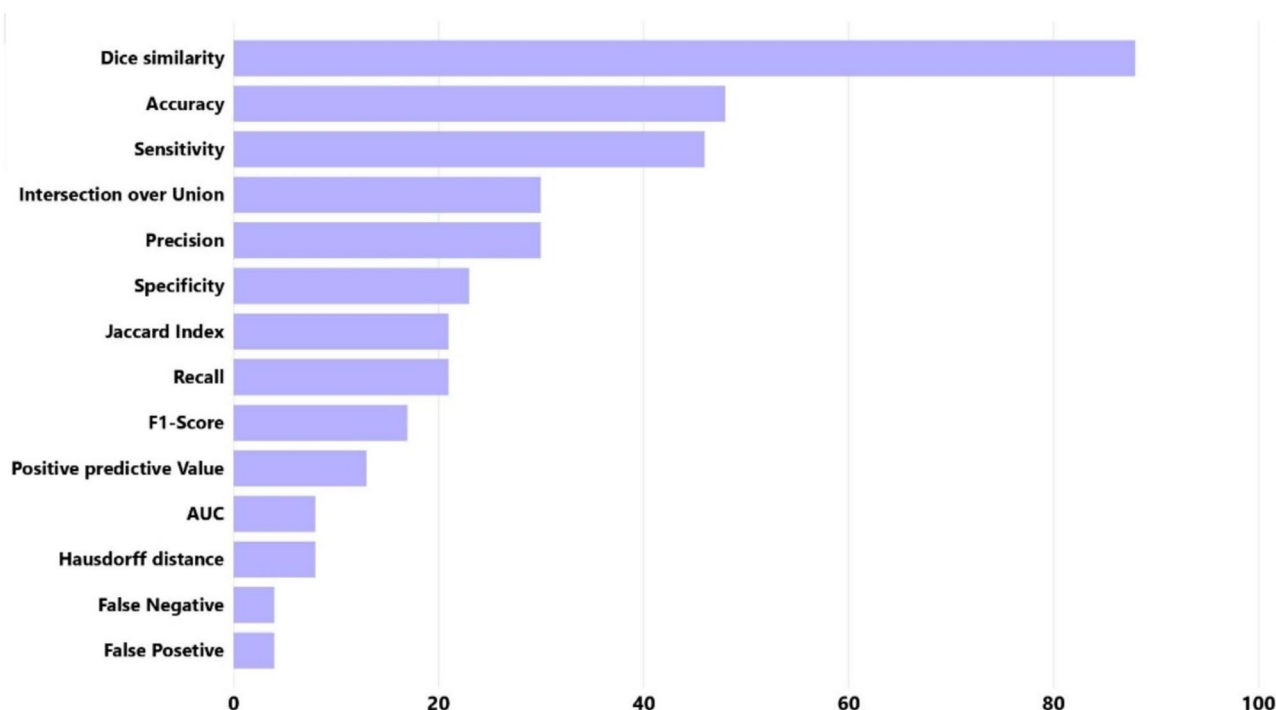


Fig. 5 Distribution of Evaluation Metric Usage in 2D DL segmentation in lung cancer for CT imaging This bar chart represents the frequency of usage of various evaluation metrics in medical imaging research. The most frequently used metric is Dice Similarity, fol-

lowed by Accuracy, Sensitivity, and Intersection over Union. Other commonly used metrics include Precision, Specificity, and Jaccard Index, with fewer studies utilizing metrics like Hausdorff Distance, False Positive, and False Negative

as Focal Loss, Weighted Cross-Entropy, and GAN Loss, are also present, albeit in smaller portions. They contribute by addressing specific challenges like hard-to-segment regions or rare class occurrences. The variation in the use of these loss functions reflects the complexity and multifaceted nature of lung cancer image segmentation, where different loss functions emphasize different aspects of model performance, from accurate tumor detection to minimizing false negatives.

Distribution of Datasets in DL-Based 2D Lung Cancer Segmentation

This paper section focuses on the usage frequency of top 10 different datasets in 2D lung cancer segmentation research, as shown in Fig. 7. A systematic review illustrated by a bar chart identifies LIDC-IDIR as the top-used dataset, accounted for 51% of Sum of Count of dataset.. The LIDC-IDRI dataset, created through collaboration between seven academic centers and eight medical imaging companies, contains 1,018 cases. Each case includes clinical thoracic CT images and a file documenting a two-phase annotation process by four experienced thoracic radiologists. In the blinded-read phase, radiologists independently identified lesions in three categories: "nodule ≥ 3 mm," "nodule < 3

mm," and "non-nodule ≥ 3 mm [50, 51, 53, 54] The second most used dataset for 2D segmentation after dataset LIDC-IDIR is LUNA16. The LUNA16 dataset is a dataset for lung segmentation. Derived from the LIDC-IDRI dataset, it contains 888 CT scans with annotations from four radiologists, categorizing nodules as solid, part-solid, or non-solid, focusing on nodules that are 3 mm or larger in diameter. It includes annotations confirmed by four experienced radiologists, categorizing lesions into nodules ≥ 3 mm, nodules < 3 mm, and non-nodules.

Analysis of the Most Commonly Used LIDC-IDRI and LUNA16 and Datasets Across DL Segmentation Models

In lung nodule segmentation using the LIDC-IDRI dataset, various UNet-based models have shown distinct performances. Li Dong et al. [55] proposed an improved U-Net++ model with a ResNeXt encoder and SCSE attention module for lung nodule segmentation in CT images. This approach enhanced feature extraction and achieved a sensitivity of 99.41%, specificity of 99.97%, a false positive rate of 0.02%, and a Dice Similarity Coefficient (DSC) of 0.98, indicating high accuracy in identifying cancerous nodules. The Deep Feature Decoupling Module (DFDM) improves

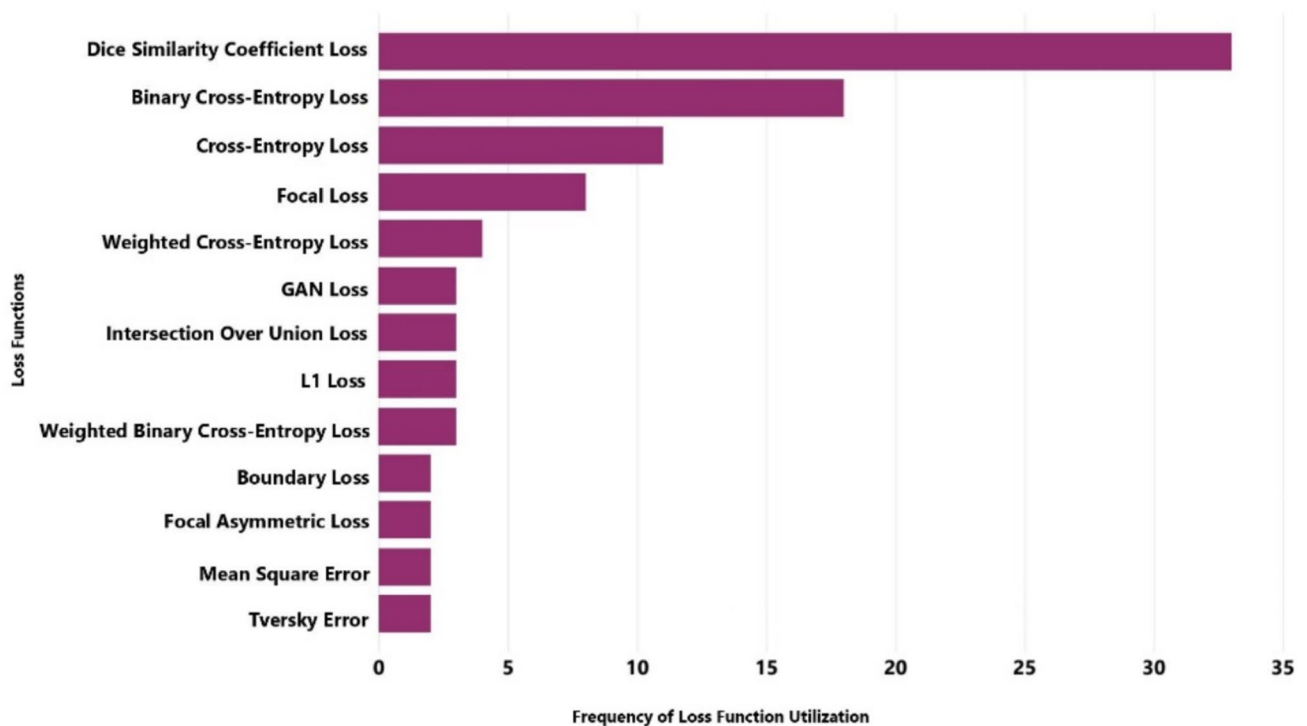


Fig. 6 The bar chart illustrates the frequency of utilization of different loss functions in medical imaging research. The loss functions are listed on the y-axis, and their corresponding usage frequency is represented by the length of the bars on the x-axis. The most frequently

used loss function is the Dice Similarity Coefficient Loss, followed by Binary Cross-Entropy Loss and other types of loss functions such as Cross-Entropy Loss, Focal Loss, and Weighted Cross-Entropy Loss

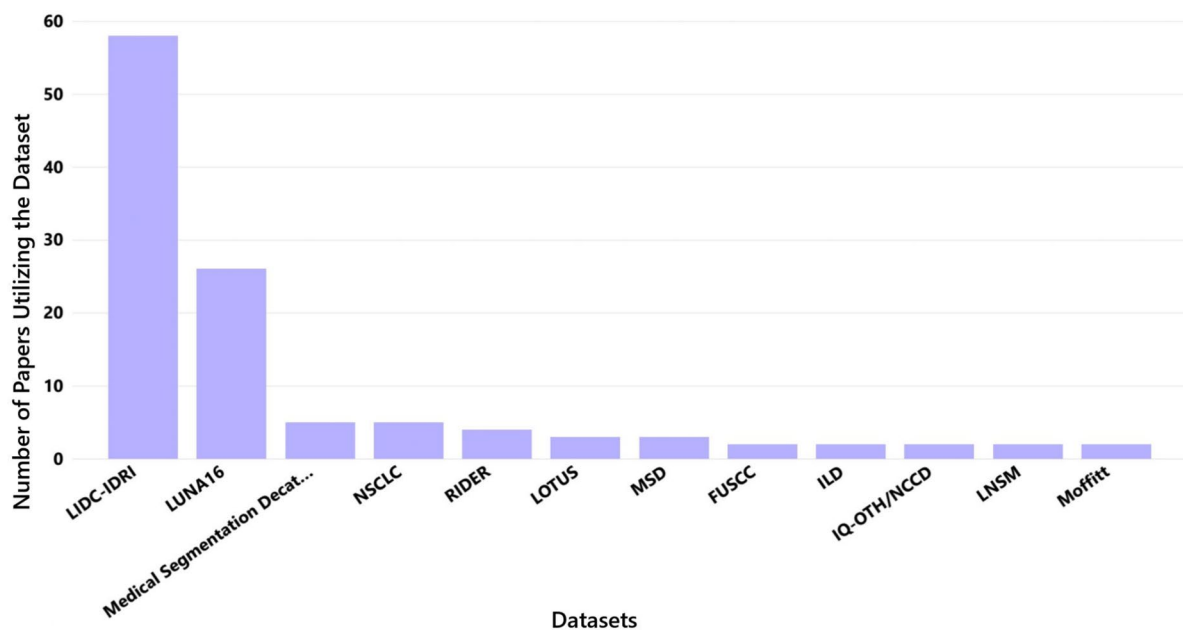


Fig. 7 Top 12 frequency distribution of datasets Bar chart displays the distribution of datasets used in lung cancer segmentation studies. The LIDC-IDRI and LUNA16 datasets are the most frequently used. The chart highlights the dominance of specific datasets in lung can-

cer 2D segmentation research, reflecting their importance and widespread application in the field. The number above each bar represents the number of subjects related to the dataset

lung nodule segmentation by separating boundary, noise, and texture information from CT images. Integrating DFDM into the UNet architecture enhanced segmentation performance, increasing IoU by 1.51% and Dice score by 1.02% [56]. Kumar et al. [57] propose an ensemble DL approach for lung cancer detection, combining UNet for segmentation and a convolutional neural network for classification. The model achieves a sensitivity of 90.83%, specificity of 94.28%, precision of 93.1%, and accuracy of 91.23%. The Incremental-MRRN and Dense-MRRN models enhance lung tumor detection accuracy while preserving edges, achieving a bit error rate of approximately 0.015%. These models outperform existing algorithms by effectively reducing noise and artifacts in CT images [58]. A Dual Encoding Fusion Network (DEF-Net) for segmenting atypical lung nodules in CT images, achieving an average Dice coefficient of 0.85. DEF-Net outperforms existing state-of-the-art methods, demonstrating its effectiveness in addressing segmentation challenges [59]. Box2Pseudo is a semi-supervised framework for pulmonary nodule segmentation, utilizing a small fully labeled dataset (pixel-level and bounding box annotations) and a larger weakly labeled dataset (bounding box labels only). It comprises three networks: the box-prompt network (BPN), pseudo-refine network (PRN), and main network (MAN), along with a background-filter layer (BFL) to improve pseudo label quality. On the LIDC-IDRI and HX-NODULE datasets, Box2Pseudo achieves an average Dice coefficient of 0.87, outperforming state-of-the-art methods and matching fully supervised performance [60].

The models were utilized for lung nodule segmentation on the LIDC-IDRI dataset, each exhibiting unique strengths and weaknesses. Advanced UNet with Weighted Binary Cross Entropy Loss achieved a high DSC of 0.98 but requires large datasets [61]. The standard UNet scored 0.82 in DSC, though it lacks malignancy classification [62]. More advanced variants like AWE-UNet (DSC: 0.91) [63] and RAD-UNet (mIoU: 0.88, F1: 0.94 [64] improved segmentation, particularly for small nodules. AMST achieved high sensitivity (0.97 for LIDC, 0.94 for IDRI) but may miss some nodules [65]. Dilated Convolutional FCN performed well with a DSC of 0.970 but is limited by 2D kernels [66]. LDDNet showed excellent recall and sensitivity (0.99), though it is time-consuming [67]. MAN-SVM reached a sensitivity of 0.980 and F1-score of 0.98 but requires validation on larger datasets [68]. DB-ResNet reported a DSC of 0.83 but demands significant data and computational power [69]. Lung_PAYNet achieved a DSC of 0.96 using attention mechanisms, while Global-Edge Dual-Path improved segmentation with a DSC of 0.90 [70].

Advanced UNet with Weighted Binary Cross Entropy Loss achieved [61] the highest DSC of 0.98, demonstrating robust segmentation accuracy. AWE-UNet [63] (DSC: 0.90) and RAD-UNet [64] (F1-score: 0.94) also performed

well, particularly in attention mechanisms and small nodule segmentation. TransUNet (DSC: 0.887) and Wavelet UNet + + [71] (DSC: 0.94) improved feature extraction using transformer-based architectures and wavelet pooling, respectively. Other models, like Dilated Convolutional FCN (DSC: 0.97) and LDDNet [67] achieved strong performances but are computationally expensive (Recall: 0.97). AMST [65] and Lung_PAYNet [70] offered high accuracy but are constrained by reliance on accurate segmentation and large datasets. Each model brings unique contributions, with some excelling in accuracy, while others focus on efficiency or handling challenging nodules. Thakur et al. [72] investigate uncertainty estimation for switch learning based on active learning (TLAL) in clinical image segmentation, utilizing a semi-supervised recurrent convolutional neural network (RCNN) to enhance model accuracy with minimal manual annotation. Their method introduces a novel uncertainty score and demonstrates significant improvements on lung CT datasets, outperforming traditional techniques by over 10% in Dice scores. Various models for lung nodule segmentation using the LIDC-IDRI dataset exhibit unique strengths and limitations. Overall, these models contribute to improved accuracy and efficiency in lung nodule detection, each targeting specific challenges within the field.

Table 2 presents a summarized version of Table S2 (Supplemental File 1), as shown below. The various studies on 2D segmentation methods for lung cancer, primarily utilized the LIDC-IDRI dataset. Key approaches include Fuzzy C-Means, UNet variants, CNNs, and semi-supervised methods, yielding notable metrics such as high sensitivity (99.41%) and accuracy (up to 98.71%). Many studies emphasize the advantages of enhanced segmentation accuracy and efficiency, along with effective multi-scale feature extraction. However, challenges such as high computational complexity, potential false positives, and limitations in distinguishing benign from malignant nodules are common. Notable papers include those detailing innovative frameworks and DL models, with links provided for further exploration. Overall, these studies contribute significantly to improving lung cancer detection through advanced segmentation techniques.

As shown in Table 2, in lung cancer segmentation using the LUNA16 dataset, a wide range of models based on the UNet architecture and its derivatives have been explored, demonstrating strong segmentation performances. R2UNet [73], with its densely connected layers, reported a DSC of 0.98, reflecting its exceptional accuracy in feature reuse and lung nodule segmentation. Similarly, DC-UNet [74], employing dilated convolutions, achieved a DSC of 0.745, indicating moderate segmentation capabilities but facing challenges with irregular nodule shapes. The basic UNet [75] model has been a foundational architecture, achieving consistently high accuracy metrics, including an impressive

F1-score of 0.99 and accuracy of 1.00. Enhancing the UNet structure, Preprocessing-UNet-ResNet50 [76] integrated ResNet50 as a backbone for feature extraction, resulting in superior performance with a DSC of 0.99, making it one of the top-performing models.

On the other hand, AWE-UNet [63], another UNet-based architecture, reported a DSC of 0.90, balancing segmentation accuracy and computational cost. The Improved UNet with Alpha-hull correction [77] yielded a DSC of 0.86, focusing on refining boundary segmentation, though it faced generalization challenges due to the limited dataset size. Combination of UNet and ResNet34 [78] and UNet-ResNet34 [79] further improved performance with residual learning, reporting DSCs above 0.93 for multiple datasets, demonstrating their robustness across various lung pathologies. SW-UNet [80] achieved a DSC of 0.84, offering high accuracy (0.99) but requiring substantial computational resources. U-Det [81] achieved a DSC of 0.83, emphasizing its efficacy in segmenting small and complex nodules, while Modified UNet-based [82] achieved a DSC of 0.88, reflecting strong performance in lobe segmentation. VL-SM-Net [83], combining DL with level set refinement, produced a DSC of 0.82 for LUNA16, excelling at detecting blurred boundaries, and UNet# [84] demonstrated high accuracy with a DSC of 0.98 on the Lits17 dataset, highlighting its capability to reduce false positives. Lastly, SquExUNet

[85] reported a DSC of 0.80, leveraging cascaded CNNs but struggling with irregular nodule shapes.

In contrast, other models employed for LUNA16, such as Mask R-CNN with K-means [86], focused on unsupervised detection, though its DSC was not reported, and it struggled to differentiate between cancerous and non-cancerous nodules. Faster R-CNN and RetinaNet [87] achieved moderate precision (0.357 and 0.34, respectively) in lung nodule detection, offering limited segmentation accuracy compared to the UNet-based models. The Cascaded Multi-stage UNet [88] performed well with an accuracy of 0.95, though it missed low-contrast nodules. DB-NET [89] offered a high DSC of 0.89, positioning it as a reliable segmentation model. FCN [90] and T-Net [91] demonstrated high accuracy and sensitivity, with T-Net reporting an accuracy of 0.992 and a F1-score of 0.99, though it was limited to 2D images. Finally, MDFN [91] achieved a DSC of 0.89 and strong edge detection, though it required further validation against existing segmentation models. Tang et al.'s SM-RNet [92], with innovative scale-aware multi-attention and reverse erasure, achieves 89.29% Dice on FUSCC and 86.50% on LUNA16, outperforming state-of-the-art model. The Improved UNet model [93], achieved a DSC of 84.48, showcasing enhanced accuracy and processing speed. However, it faced challenges related to its complexity, which could lead to overfitting if not managed with a sufficiently large and diverse dataset.

Table 2 Papers Reviewed for 2D Lung Cancer Segmentation Using deep learning (DL) Models. The table summarizes the key aspects of each study, including the DL model architecture employed, the datasets used for training and evaluation, and the reported performance metrics. Sen: Sensitivity, Specificity: Spe, Accuracy: ACC, JAC: Jaccard, F1-scores: F1-s, Precision = Prec, Recall: Rec, Spearman correlation coefficient: SCC, SVM: Support Vector Machine, interstitial lung diseases: ILDs, Deep Convolutional Neural Network based Segmentation with Two-Pass Contour Refinement: DCNN-TPCR, two-stage Region Proposal Network: RPN, Medical Segmentation Decathlon: MSD, Hausdorff distance (pixels): HD, mean Sensitivity: MSen, mean Precision: MPrec, Fully Convolutional Network: FCN, Binary Cross Entropy: BCE, Fast Multi-Crop Guided Attention: FMGA, Deep Residual Separable Convolutional Neural Network: DRS-CNN, The National Lung Screening Study: NLSS. Peak signal-to-noise ratio: PSNR, Structural similarity index: SMMI, False Positive Rate: FPR, Positive Predictive Value: PPV. The National Lung Screening Study: NLSS. CapsNet: Capsule Network, SC-BLSTM:

Spatial-Context Bidirectional Long Short-Term Memory, AC-UNET: Attention Constrained UNet, CSC-Net: Cancer Sensitive Cascaded Network, AEC-Net: Attention and Edge Constraint Network, BPNN: Back Propagation Neural Network MRR: Multi-Resolution Representation, DU-Ne: Dual UNet, ViFUNet: Vision Flash-Based UNet, SMR-UNE: Scale-Aware, Structural Measure: S-M, Enhanced Measure: E-M, Multiattention-Guided Reverse Network, DMC-UNT: Dilated Multi-scale Convolutional UNet, Interstitial Lung Diseases: ILD, Jaccard Similarity: JS, ResDSda_UNet: Residual UNet with Deep Supervision and Dense Connections, SDC: Sorensen-Dice Coefficient Rows with green-colored studies are related to the LUNA16 dataset, while those with yellow-colored studies are related to the LIDC-IDRI dataset. The blue-colored rows correspond to the LUNA 16 and LIDC-IDRI datasets. The maximum values of the evaluation metrics reported in the table. Bias score (Total score of 28) is the sum of "Yes" answers for each study. This Total is provided for each study in Supplemental File 2, with the corresponding details

Reference/ year	Network, model Type	Dataset	Performance	Bias score (Total core of 28)
[99]/2020	FRFCM	TCIA	Sen=0.92, Spe=0.933, Acc=0.927	12
[100]/2020	Mask R-CNN with K-means kernel	COMEPE	DSC=0.973	12
[101]/2020	DDRN	LIDC-IDRI	Jac=0.883, DSC= 0.946	14
[65]/2020	AMST	LIDC-IDRI	LIDC: Sen=0.972, Spe=0.960, Acc=0.968 IDRI: Sen=0.937, Spe=0.915, Acc=0.931	17
[102]/2020	UNet, VGGNet,	LUNA16	SCC=0.39	13

Table 2 (continued)

	ResNet			
[103]/2020	Two-stage convolutional neural network framework	LUNA16	Acc=0.980, Sen=0.990, Spe=0.968	19
[66]/2020	Inception Block based Residual Auto Encoder (Inception-Res Autoencoder).	LUNA16	F1-scores=0.984, Sen=0.995, Spe=0.989, Acc=0.994, AUC=0.992, Jac=0.994	13
[104]/2020	Dilated Convolutional Based Weighted FCN	LIDC-IDRI	DSC=0.970, Pixel Acc=0.983	23
[68]/2020	MAN-SVM	LIDC-IDRI	F1-scores=0.979, Sen=0.980, Spe=0.956, Acc=0.972, Prec=0.978	9
[67]/2020	LDDNet	LIDC-IDRI	Rec=0.996, Sen=0.996, Spe=0.981, Acc=0.994, Prec=0.996	16
[105]/2020	UNet and Seg-UNet	LIDC-IDRI	UNet : DSC=0.830, Prec=0.792, Rec=0.898	18
[87]/2020	Faster RCNN	LUNA16	Faster RCNN : Prec=0.357	12
[106]/2020	BPNN with SIFT and Watershed Segmentation	LIDC-IDRI	Sen=0.865, Spe=0.805, Acc=0.97	10
[69]/2020	Dual-Branch Residual Network	LIDC-IDRI	DSC=0.827, Sen=0.893, PPV=0.796	21
[107]/2020	ConvNets	LIDC-IDRI, Tri-Service General Hospital data	Tri-Service General Hospital : 92.5, Average JI: 0.91	20
[61]/2020	Advanced UNet	LIDC-IDRI	DSC=0.98	9
[88]/2020	Cascaded Multi-stage UNet	LUNA16 , ALIBABA TianChi	TianChi : Acc=0.947, Sen=0.903, Spe=0.965, AUC=0.972	19
[108]/2020	DCNN-TPCR	ILDs	Sen=0.97, Rec=0.97, DSC=0.984, Jac=0.96	19
[109]/2020	RPN	LIDC-IDRI	DSC=0.927, Rec=0.923, Prec=0.931	17
[93]/2020	UNet	LUNA16	DSC=84.48, Average Prec=85.35	11
[110]/2020	Fuzzy C-Means	LIDC-IDRI	PSNR, MSE, SSIM	9
[111]/2020	NSCLC	LIDC-IDRI	Sen=99.41, Spe=99.97, FPR=0.01, DSC=0.99, Acc=99.97	15
[112]/2020	DenseNet	RIDER	Sen=0.88, Spe=0.98, DSC=0.94	15
[113]/2020	Programmable algorithm	LIDC-IDRI	Acc= 1.00	3
[54]/2020	Capsule Network	LUNA16	JS: 98.471	9
[114]/2020	AEC-Net	LUNA16 SIC2017:	LUNA16 ::Acc: 0.9949, Sen: 0.9898, Spe: 0.9954 J S: 0.99 DSC: 0.88	18
[115]/2021	2D & 3D hybrid CNN	Private	DSC=0.75, Jac=0.62	18
[116]/2021	2D & 3D hybrid CNN	Private	DSC of all lung lobes contours exceeded 0.75, Ave DSC: 0.941, Ave Acc:99.63, Ave Sen:95.38, Ave Spe:99.77	18
[117]/2021	semi-sequential detection with synthetic CT images	LIDC-IDRI	DSC=0.931, Sen=0.917	14
[85]/2021	SquExUNet	LUNA16, LNDB	DSC=0.80	19
[97]/2021	Dual-Branch UNet with Local Lesion Information Enhancement	LIDC-IDRI , SHCH	LIDC-IDRI :DSC=0.809, Sen:0.885,PPV:0.7898	20
[118]/2021	MultiResUNet	LOTUS	DSC=0.847	18
[119]/2021	Fast Multi-Crop Guided Attention (FMGA)	LIDC-IDRI	DSC=0.813, Sen=0.923, PPV=0.747	18
[73]/2021	Dense R2UNet	LUNA	DSC=0.981, Jac=0.961, Prec=0.982, Rec =0.988, Sen=0.994, Spe=0.988, Acc=0.991, AUC=0.989	20
[74]/2021	DC-UNet	LUNA16	IOU= 0.786, DSC=0.745, Prec=0.718, Rec=0.745	17
[120]/2021	SegNet	private	Sen=0.983, Spe=0.866, Acc=0.925	10

Table 2 (continued)

[75]/2021	UNet	LUNA16	Acc=0.999, AUC=0.998, Sen=0.990, Prec=0.998, F1-score=0.987	19
[121]/2021	HDA-ResUNet	Lung dataset	DSC=0.979, JS=0.960, Acc=0.993, Pre=0.974, AUC=0.997	21
[122]/2021	Improved UNet with alpha-hull correction	LUNA16,NLSS	DSC= 0.862	18
[123]/2021	UNet	LIDC-IDRI	Acc= 0.987 DSC= 0.820	11
[124]/2021	UNet	LUNA16	Acc= 0.999, AUC=0.998,Sen= 0.990, Prec=0.998 F1-score=0.987	10
[62]/2021	UNet	LIDC-IDRI	Acc=0.987, DSC=0.820, IoU=0.753	16
[125]/2021	UNet	LIDC-IDRI	DSC=0.984, Sen=0.987, PPV=0.981, HD=0.076	20
[126]/2021	Labeling and Clustering-based	LIDC-IDRI	Acc=0.975, Spe=0.974, Sen=0.916, Prec=0.987	10
[127]/2021	2D-UNet	COSMOS	DSC=0.75, (MSD:1.34.mm, IoU:0.63	19
[128]/2021	DA-Net.	LIDC-IDRI	DSC=0.81, IOU=0.716	16
[75]/2021	UNet	TMUH	DSC=0.622, Acc=0.998, Sen=0.956, Spe=0.998	16
[76]/2021	Preprocessing-UNet-ResNet50	LUNA16, NLST	LUNA16:ACC=0.989, AUC=0.986, Sen=0.989, Prec=0.984, F1-score=0.988	17
[63]/2021	AWE-UNet	LIDC-IDRI, LUNA16	LIDC-IDRI: ACC=0. 946, Sen=0.908, Spe=0.964, DSC=0.903, IoU=0. 832	19
[129]/2021	Deeplabv2-VGG16,FCN 32s,FCN 16s,FCN 8s,Deeplab-largeFOV,D-FCN 16s,D-FCN 8s,D-FCN 4s,UNet	Private	D-FCN 4s: DSC=0.871	18
[130]/2022	Ccombination of UNet and ResNet34	LUNA16, LCTSC, HUG-ILD, VESSEL12, CHUSJ	VESSEL12:0.9767	15
[131]/2021	UNet	NSCLC	DSC=53.34, Acc=78	12
[55]/2021	UNet++, ResNeXt,SCSE attention module	LIDC-IDRI	DSC=83.43, IoU=72.51	11
[132]/2021	CNN	ACDC	DSC=0.83, Acc, Prec, Sen, Spe	19
[56]/2021	UNet	LIDC-IDRI	Iou=76.71, DSC=85.98	11
[57]/2021	2D UNet	LIDC-IDRI	Acc=91.23, Sen=90.83, Spe=94.28, Prec=93.10	9
[133, 134]/2021	CNN	LUNA16	IoU=0.85, DSC=0.87, Prec=0.84, Rec=0.86	14
[135]/2022	DRS-CNN	Medical Segmentation Decathlon (MSD) and StructSeg 2019	DSC=0.64, HD95=18.26, mean Sen=0.737, mean Prec=0.765	13
[136]/2022	DWT	LOTUS Dataset	DSC=0.847	15
[137]/2022	dilated convolutions and residual connections	LOTUS	AveDSC: 0.029, AveMSD: 0.666, Ave% HD(95%): 0.1247	17
[138]/2022	UNet-ResNet34	LCTSC, LUNA16,HUG-ILD, VESSEL12, CHUSJ	Dise > 0.93 for the 4 different cohorts tested.	13
[134]/2022	MediNet	QIN	Acc: 0.97	13
[58]/2022	Dense Residual Network	LIDC-IDRI	MSE:14.64, Bit Error Rate=0.01	6
[59]/2022	CNN	LIDC-IDRI	DSC=77.00 (best) 75.58 (average), IoU=77.00 (best) 75.58 (average)	14
[94]/2022	TransUNet	LUNA16	DSC=86.03, IoU=78.19, Positive Predictive Value=87.65, Sen=88.51, Average Surface Distance=0.4536mm	18
[139]/2022	MRF	LIDC-IDRI	PSNR=17.27, Acc=99.96, Jac=99.93	8
[140]/2022	Softgan	LNSM,LIDC-IDRI	LNSM: DSC=91.63, Sen=89.76, PPV=94.12	16
[141]/2022	UAA-UNet	LIDC-IDRI	DSC: 86.89%, Pre: 89.47% Rec: 85.11% ,IoU:	16

Table 2 (continued)

			77.45%	
[134]/2022	MedNet	Private	Accuracy: 97%	12
[142]/2023	SegChan-et	Private	Acc with CAM=0.984, Acc without CAM=0.968	16
[143]/2022	Mask-RCNN, UNet, SegNet, DeepLabv3+	LUNA16	DSC=0.829	16
[89]/2022	DB-NET	LUNA16	DSC=0.888	16
[90]/2020	Res2Net50	LUNA16, FUSCC	FUSCC: DSC: 0.868, JI: 0.767, Spe: 0.987, Sen: 0.865, S-M: 0.921, E-M: 0.914, MAE: 0.026	18
[70]/2020	Lung_PAYNet	LIDC-IDRI	DSC=0.957, mIoU=0.917, Sen=0.925, Prec 0.967	18
[144]/2020	UNet with Attention Mechanism	IELCAP	Sen=0.91, Spe=0.88	18
[145]/2020	Global-Edge Dual-Path	LIDC-IDRI	DSC=0.898, Sen=0.905	19
[80]/2023	SW-UNet	LUNA16, LiTS 2017, KiTS 2019	LUNA16: DSC=0.84, Acc=0.99, Sen=0.83, Spe=0.99, H95=7.41	19
[64]/2023	RAD-UNet	LIDC, AHAMU-LC	AHAMU-LC: mIoU=0.881, Rec=0.923, Prec=0.948, F1-score=0.937	19
[146]/2023	U-shaped Network Transformer (UNT)	public dataset available in the Kaggle and GitHub repositories	Acc=0.969	16
[147]/2023	VGG-SegNet	TCIA	Jac=0.901, DSC=0.940, Acc=0.973, Prec=0.935, Sen=0.937, Spe=0.991	18
[81]/2023	U-Det	LIDC-IDRI, LUNA16	DSC=0.828, Sen=0.922, PPV=0.789	18
[148]/2023	UNet, SegNet, GCN, FCN, DeepLabV3+, PspNet TransUNet, SwinUNet	LIDC-IDRI	TransUNet: DSC=0.887	18
[149]/2023	MobileNetV2-UNet	(MSD) 2018, TCIA	DSC=0.879, Rec: 0.8602, Pre: 0.93	18
[150]/2023	DEHA-Net	LIDC-IDRI	DSC=87.91, Sen=90.84, PPV=89.56	14
[71]/2023	Wavelet UNet++	LIDC-IDRI	IOU=0.878, DSC=0.937, Acc=0.989	10
[151]/2023	ESSegILD	First Affiliated Hospital of Guangzhou Medical University from 2015 to present.	DSC= 0.9616 (Cyst), IoU: 0.63, MSD: 0.70	14
[82]/2023	Modified UNet based	LUNA16	DSC=0.883, Sen=0.855, Prec=0.912	16
[152]/2023	Deep Fused Features-Based Cat-Optimized Networks (DFF-CON)	LIDC-IDRI	Acc=0.999, Prec=0.999, Sen=0.999, Spe=0.999, F1-score=0.999	18
[60]/2023	Unet	LIDC-IDRI, HX-NODULE	HX-NODULE: DSC=0.79, HD=2.92 mm	14
[153]/2023	U-Net, ResNet50 encoder, attention mechanism	MSD	CBAM Res50Unet++: DSC=0.74	10
[154]/2023	CNN	LIDC-IDRI	DSC: 62.00, Acc=82.00, F1-Score=78.5	15
[155]/2023	Swin Transformer	ILD, ANODE09, LIDC-IDRI, ELCAP, LUNA16	DSC=96.72, IoU=91.22, F1-Score=89.37, Prec=90.34, Rec=93.36	17
[98]/2023	U-Net, CNN, PCA	LUNA16	CNN+PCA: Acc=95.92, Sen=91.91, Spe=96.35, F1-Score=0.92	16
[156]/2023	Residual UNet	LIDC-IDRI	DSC=87.23, IoU=77.35, Sen=81.08, Prec=89.09	9
[157]/2023	UNet	LIDC-IDRI	DSC=71.99, Acc=76.67, AUC=84.49	9
[158]/2023	Attention-based Residual U-Net (AttResiU-Net)	IQ-OTH/NCCD	Acc=96.09, DSC=80.20, IoU=81.67, Prec=91.56, Rec=80.97	13
[159]/2023	DSNet	LNSM LIDC-IDRI	LNSM: DSC=89.45, PPV=0.12, Sen=88.67	15
[160]/2023	UNet	LIDC-IDRI	Jac :0.758, DSC:0.856	13
[161]/2023	DMC-UNet,U-Net	FAHGMU	IoU=65.52, F1-Score=76.02	17
[162]/2023	SMR-UNet	LIDC-IDRI, FAHGMU	LIDC: DSC: 0.91, IoU: 0.86	18
[163]/2023	ResDSda_UNet	LIDC-IDRI	DSC:86, IoU: 76.73, Sen: 86.30%, Prec: 87.22%	17

Table 2 (continued)

[92]/2023	SM-RNet	FUSCC LUNA16	FUSCC: DSC=89.29, Jac=80.65, HD=5.53, Rec, Spe=99.53, Prec=89.77	21
[164]/2023	ResNet-34 (used in the Res BCDU-Net)	LIDC-IDRI	Acc: 98.09%, Pre: 99.92, Rec: 97.77 F1-S: 98.83, DSCt: 97.23	14
[62]/2024	MHD-ATRes-UNet++	UnitoChest	Precision= 0.39	16
[84]/2024	UNet#	LUNA16, LIDC-IDRI, Lits17, BraTs19, Dsb2018	Lits17: IOU=0.953, DSC=0.982	14
[165]/2024	MDFN	LUNA16	Sen=0.888, DSC=0.891, Jac=0.807	12
[96]/2024	UNet	LUNA16	IoU=0.95, Acc=0.98	7
[166]/2024	CNN	LIDC-IDRI	Acc=98.71, DSC=0.83, IoU=0.74	13
[167]/2024	VGG16	IQ-OTH/NCCD	Acc=0.89, Prec=0.76, Rec=0.97, F1-Score=0.92, F1-Score=0.85	18
[168]/2024	Mask R-CNN	Not explicitly mentioned	DSC=94.87, IoU=95.02, F1-Score=97.44	16
[95]/2024	CNN	LIDC-IDRI, LUNA16	LIDC-IDRI: DSC=0.85 (deep learning)	17
[72]/2024	RCNN	LIDC-IDRI	Acc=97.0	17
[169]/2024	parallel ResUNet	LIDC-IDRI, LMTC	LIDC-IDRI : DSC: 90.19% ,PPV: 92.47%, Sen: 89.92% ,IoU: 87.03	12
[170]/2024	UNet	FEMH, LUNA16	DSC= 0.926, IoU= 0.862	19
[171]/2024	UNet	NPC-170	GED: 0.23, SDS: 80.43, DS: 79.78	16
[172]/2024	EDTNet	DS1, DS2	DS2: Prec: 97.67, IoU: 96.06, SDC: 97.85	14
[173]/2024	enhanced U-Net++	LUNA16	DSI : 91.76, IoU: 89.78, Sen: 89.54, Spec: 85.98	16
[174]/2024	Pix2Pix, U-Net	Finding Lungs in CT Data	Acc: 89.64, SD: 7.37, Overlap Rate: 93.03 ,Std Dev: 11.83, F-M: 95.85	8
[175]/2024	U-Net, Attention U-Net, FMC-Net, WDUNet, MultiResUNet, MRRN, UNet++, DMDF-Net	NSCLC Radiomics, NSCLC Radiogenomics, Decathlon, Moffitt	NSCLC Radiogenomics Dataset: DSC: 0.79, IoU: 0.73	15
[176]/2024	ResNet50, Inception ResNetV2, InceptionV3, VGG16, VGG19	MSD	VGG16: DSC: 0.4020	10
[91]/2024	T-Net, CenterNet, NASNet	LUNA-16, LIDC-IDRI	Acc= 0.992, Precision= 0.991, Sen= 0.992, Spe= 0.992 ,F1-score= 0.992	16
[177]/2024	GPD-Nodule	LIDC-IDRI	ACC: 97.99, AUC: 0.98	17
[178]/2024	LA-ResUNet	LIDC-IDRI	IoU: 60.62, DSC: 73.11	10
[179]/2024	Modified UNet	Medical Segmentation Decathlon	DSC: 0.80, Sens: 0.64, Spe: 0.9999, Acc: 0.9999	14
[180]/2024	UNet	Lung-PET-CT-Dx	Acc: 99.53, IoU: 77.27, DSC: 48.60	11
[181]/2024	WDUNet	NSCLC Radiomics, NSCLC Radiogenomics, Decathlon, LIDC-IDRI, Moffitt, and Rider datasets	LIDC-IDRI: DSC: 0.91, HD95: 5.38, AUC: 0.94	17

Following this, the ViFUNet (Vision Flash-based UNet), model [94], proposed in 2022, outperformed traditional UNet architectures by reaching a DSC of 86.03. This model benefits from a faster convergence speed and a reduced parameter count, making it efficient for processing large nodule datasets. Nonetheless, it may struggle with very small

nodules, which can limit its clinical applicability. Another study in 2024 [95] highlighted a classic UNet model, achieving an impressive Intersection over Union (IoU) of 0.95. This model demonstrated high accuracy in segmenting lung structures, although it also required substantial annotated datasets and computational resources for training.

In summary, UNet-based models, particularly Preprocessing-UNet-ResNet50, R2UNet, and UNet#, demonstrated the highest DSC values, offering robust and accurate lung nodule segmentation on the LUNA16 dataset. Models like SquExUNet and DC-UNet showed moderate performance, while others like DB-NET and MDFN performed well in specific tasks but required further validation for broader application. Non-UNet-based models such as Faster R-CNN and Mask R-CNN delivered moderate results but generally underperformed compared to UNet and its derivatives. A Semantic Segmentation of CT Scans study employed a UNet model, achieving an IoU of 0.95, indicating high accuracy and efficiency in lung structure segmentation, but it also highlighted the need for large annotated datasets for effective training [96].

Beyond the UNet family, other models have contributed to the landscape of lung cancer segmentation. The Deep Neural Networks for Semantic Segmentation study from 2021 employed a CNN-based approach, reporting an IoU of 0.85. While it achieved high segmentation accuracy, it also demanded considerable computational resources, which can be a barrier for implementation in resource-constrained settings [97]. The Detecting Lung Nodules Based on DL model utilized a combination of UNet, CNN, and PCA in 2023, achieving an accuracy of 95.92% for the CNN + PCA approach. This model offered a reduced computational cost and a simpler architecture; however, it was criticized for potentially losing spatial context in the segmentation process [98] (Table 2).

Bias Analysis in Lung Cancer Segmentation Studies Using CT Images: A Focus on UNet Models

Our analysis of 124 studies on lung cancer segmentation, assessed using tailored TRIPOD questions, revealed critical insights summarized in Table 3. Regarding Data Handling, while all studies (100%) report data sources (Q1) and 92% detail dataset splitting (Q2), only 67% address class imbalance between cancerous and non-cancerous regions (Q3). In Preprocessing and Augmentation, standardized preprocessing is universally applied (94%, Q4), but only 60% enhance dataset diversity through data augmentation (Q5). For Model Optimization, models' comparisons are present in 83% (Q6), but only 21% use cross-validation, risking overfitting (Q7). Hyperparameter tuning is mentioned in 68% (Q8), while model stability across initializations is rarely assessed (3%, Q9). Evaluation Metrics reveal 100% report at least one metric (Q10–Q11), yet only 11% use six metrics (Q16), reflecting limited evaluation depth. On Generalizability and Diversity, just 53% address external dataset generalizability (Q17), 71% include diverse cancer stages (Q18), and only 4% consider socioeconomic or geographic factors (Q19). Finally, under Annotation and Protocols, 74%

reduce subjectivity with multiple annotators (Q24), and 76% incorporate varied CT scanners and protocols (Q25) for robustness. However, key gaps persist, with only 10% discussing eligibility and treatments (Q20), 12% addressing missing data (Q21), 60% acknowledging training-evaluation dataset differences (Q22), and none ensuring ground truth consistency (Q23). 0.74% use multiple annotators (Q24) and 62% used data from Varied CT scanners and protocols (Q25). Only 17% of the reviewed studies used non-public datasets for external validation, indicating a need for broader validation to ensure generalizability (Q26). A significant 89% of reviewed studies performed ablation studies, highlighting a commitment to understanding the impact of model components (Q27). Additionally, 25% of studies highlighted top metrics without statistical analysis, underscoring a need for more rigorous statistical validation in reported results (Q28).

We have added a bias analysis specifically for UNet models, along with a question analysis, as these are the most frequently reported models in 2D segmentation of lung cancer CT imaging. Please refer to Table 3 for more details.

For all studies, similar strengths and weaknesses were observed. Data source identification remained high, and dataset splitting was also widely used. However, 67% of studies reported class imbalances, and 96% standardized preprocessing techniques, demonstrating strong methodological consistency. Cross-validation was underused (21%), and model stability was poorly evaluated (3%). While most studies reported at least one evaluation metric, a significant portion did not report enough metrics for thorough performance assessment. Generalizability was discussed in only 34% of studies, and socioeconomic factors or geographic variations were largely ignored. The handling of missing data was not addressed in most studies (12%), and external validation using non-public datasets was limited to 22%.

For UNet models, the studies demonstrate strong transparency in data source identification and dataset splitting, ensuring reproducibility and model validation. Preprocessing techniques were standardized in all studies, which is a significant strength. However, class imbalance was reported in 65% of studies, highlighting a common challenge. Data augmentation was used in 63% of studies, but its implementation could be expanded. Notably, only 23% of studies used cross-validation, and only 2% evaluated model stability across different initializations. While the majority of studies reported at least one evaluation metric, there was a gap in reporting a broader range, with only 18% reporting six or more metrics. Generalizability to other datasets was addressed in 53% of studies, but there was little focus on the impact of socioeconomic or geographic factors on model performance. Furthermore, missing data was handled in just 14% of studies, and only 65% ensured ground truth labels were reviewed by multiple experts to reduce bias. The bias score (total score of 28) is the sum of "Yes" answers for each

Table 3 Comprehensive Analysis of Bias Factors in 2D Segmentation Models for Lung Cancer Across All Studies and UNet Models. This table offers a detailed examination of the principal factors that contribute to bias in 2D segmentation models for lung cancer across all

studies and UNet models. It assesses 28 different criteria in various studies, emphasizing elements like dataset transparency, preprocessing standardization, model validation, evaluation metrics, generalizability, and data quality

Question No	Questions	Percent of Yes Answers (%) UNet Models	Percent of Yes Answers (%) All studies
1	Are the data sources clearly identified (e.g., multi-center, single-center, public datasets)?	100	99
2	was the dataset split (e.g., train, validation, and test sets)?	94	92
3	Are the class distributions (e.g., cancer vs. non-cancer regions) imbalanced?	65	67
4	Were the preprocessing techniques (e.g., normalization, cropping, resizing) standardized across all cases?	100	94
5	Is data augmentation used?	63	60
6	Are different models compared systematically?	90	83
7	Is cross-validation used properly (e.g., leave-one-center-out)?	23	21
8	Have articles referred to hyperparameter tuning methods?	65	68
9	Are models evaluated for stability across different initializations or random seeds?	2	3
10	Number of Evaluation Metrics Reported	-	-
11	At least 1 Evaluation Metric Reported	100	100
12	At least 2 Evaluation Metrics Reported	80	90
13	At least 3 Evaluation Metrics Reported	59	74
14	At least 4 Evaluation Metrics Reported	26	51
15	At least 5 Evaluation Metrics Reported	8	24
16	At least 6 Evaluation Metrics Reported	18	11
17	Does the paper specifically discuss the generalizability of the models to other datasets?	34	53
18	Does the dataset include a wide range of cases, not just severe or advanced cases?	63	71
19	Does model performance vary based on socioeconomic factors or geographic locations?	3	4
20	Are the eligibility criteria and any treatments received by participants described?	5	10
21	Is the handling of missing data described, and its impact on the results discussed?	14	12
22	Are differences between training and evaluation datasets identified?	68	60
23	Are the ground truth labels consistent and free from annotator bias?	0	0
24	Has the ground truth labeling been performed by multiple experts to reduce subjectivity?	65	74
25	Were the CT scans in the dataset obtained using different CT scanners and imaging protocols across various institutions?	70	62
26	Is data used non-public datasets especially for external validation purposes?	17	22
27	Is the study an ablation study ?	89	92
28	Is the study simply bolded or underlined the highest performing metric values without actually doing a statistical analysis of the data to test whether any differences were statistically significant?	25	20

study. This total is provided for each study in Supplemental File 2, with the corresponding details. Additionally, the sum of "Yes" answers for each study is included in Table 2.

Discussion

Lung cancer is the leading cause of cancer-related deaths worldwide with a five-year survival rate remaining alarmingly low due to late-stage diagnoses [1, 2]. Early CT screening has been proved to reduce mortality rates by enabling early detection and then intervention. Recent advancements

in AI, including radiomics, machine learning (ML), and DL, have significantly transformed imaging analysis, particularly in evaluating lung nodules. Automated segmentation has become a critical component of quantitative analysis, streamlining the detection, diagnosis, and monitoring of lung nodules by reducing time, labor, and variability. In medical imaging, 2D segmentation plays a vital role in accurately extracting ROIs. For instance, a recent study [182] indicated that 2D slice segmentation remains clinically important as it captures micro -environmental heterogeneity that may be overlooked in 3D volumetric analysis. Research has demonstrated that peritumoral radiomic features extracted from 2D

CT slices can independently predict treatment response in lung cancer patients, showing comparable prognostic value to 3D analyses. Additionally, 2D features enable a more granular, slice-by-slice assessment of lesion characteristics, providing statistical insights into texture variations that a single 3D value may miss. This approach is particularly valuable when full 3D segmentation is impractical in clinical workflows and supports future optimization studies using dominant lung lesions for prediction.

The rapid evolution of AI technologies, such as radiomics and DL algorithms, has enhanced quantitative imaging analysis for lung cancer, particularly in lung nodule imaging. Segmentation serves as the cornerstone of workflows ranging from nodule detection to diagnosis and ongoing monitoring. Automated nodule segmentation has garnered significant attention due to its ability to address challenges associated with time-intensive workflows, laborious manual effort, and susceptibility to human error and variability.

Automated 2D segmentation offers several distinct advantages. It aligns well with most medical imaging databases, which predominantly store 2D images. Unlike 3D models, it eliminates the need for resizing, ensuring seamless analysis. Additionally, it reduces processing time, enabling quicker review of CT slices, and offers flexibility for re-training across diverse 2D image datasets, including X-rays. These benefits make 2D segmentation a practical and scalable solution for clinical workflows [182, 183].

While 3D segmentation methods are increasingly favored for their ability to capture volumetric tumor characteristics and inter-slice relationships, 2D segmentation remains highly relevant and indispensable in medical imaging due to its accessibility, computational efficiency, and alignment with slice-based diagnostic workflows. It is particularly valuable and practical in resource-limited settings where 3D data acquisition or processing may be impractical, allowing for rapid analysis, easy adaptation across datasets, and serving as a foundation for advanced techniques like semi-supervised learning. One of its key advantages is feasibility with limited data, as annotated 3D datasets are costly and time-intensive to acquire, whereas 2D methods rely on slice-by-slice annotations, making them more accessible. Training 2D models on individual slices extracted from 3D volumes increases the number of training samples, and by integrating these slices, we can reconstruct volumetric segmentation, addressing data scarcity and enhancing generalizability. Although 3D CNNs can capture intricate volumetric features, they require substantial computational resources and extended training times. By contrast, analyzing volumetric images slice-by-slice using 2D CNNs, combined with data augmentation and tri-planar views, helps preserve essential 3D characteristics. This approach strikes a balance between efficiency and effectiveness, underscoring the importance of 2D segmentation in advancing lung cancer imaging

and treatment workflows. Additionally, 2D segmentation provides a steppingstone for semi-supervised and transfer learning strategies, bridging the gap between 2D and fully 3D segmentation tasks. Its continued clinical relevance in assisting radiologists with tumor identification and precise image analysis highlights its enduring importance alongside emerging 3D methodologies [93, 183].

Pang et al. [184] emphasized the effectiveness of 2D segmentation in addressing critical challenges such as detecting small ROIs and capturing subtle anatomical details, as demonstrated in liver tumor and retinal vessel segmentation tasks. Furthermore, 2D methods are more practical in scenarios where 3D data is unnecessary or difficult to obtain, requiring less computational power, simpler implementation, and benefiting from readily available pretrained models to accelerate development. This flexibility makes 2D segmentation an essential tool in many computer vision applications, particularly when balancing efficiency and effectiveness is a priority.

X-ray radiographs, among the most widely used imaging modalities in ML and DL studies, further demonstrate the importance of 2D segmentation. Radiographs are common in both developed and developing countries and serve as the basis for diagnosing numerous diseases. Publicly available datasets, such as the CheXpert and NIH Chest X-Ray datasets, have driven significant advancements in 2D segmentation techniques. Despite the initial limitations of 2D representation, such as reduced soft tissue contrast and the loss of volumetric context, 2D segmentation has laid the groundwork for more sophisticated, non-invasive imaging modalities and remains pivotal in advancing research and clinical practice in lung cancer diagnosis [185].

Additionally, while 2D segmentation techniques can extend to continuous video data by treating it as a series of 2D frames, this approach sacrifices the ability to leverage temporal continuity inherent in video. This trade-off highlights the need to balance computational efficiency with the incorporation of advanced temporal modeling for applications requiring sequential data processing [186].

Despite the shift towards 3D methods, the enduring value of 2D segmentation lies in its ability to address specific clinical needs effectively and efficiently. By refining 2D segmentation methods to mitigate limitations such as the lack of inter-slice correlations, their clinical relevance can be further enhanced, complementing 3D segmentation approaches while maintaining their unique advantages. This study underscores the foundational role of 2D segmentation in medical imaging, offering a bridge between accessible, efficient workflows and advanced diagnostic capabilities.

This study presented a review of 2D lung segmentation in CT images, answering the research question: “What are the 2D segmentation methods for the lung cancer, using CT images”? This analysis examined the latest developments

in 2D segmentation techniques for lung cancer CT scans, paying special attention to the growing influence of DL applications in this area. Findings highlight a strong shift towards DL, particularly supervised learning, driven by public datasets like LIDC-IDRI and LUNA16 and, which foster innovation. The analysis reveals satisfactory accuracy across 2D segmentation techniques, with UNet models often outperforming others due to its adaptability to complex shapes. The DSC is a popular metric for 2D segmentation, measuring overlap between the segmented image and ground truth to assess accuracy. It balances false positives and negatives, providing a robust evaluation. However, DSC can be sensitive to small variations, impacting robustness in cases requiring precise delineation [187, 188].

Although DL shows promising results, evaluating and comparing 2D segmentation approaches is challenging due to: (i) Differences in Study Methodologies: Variations in methods, sample sizes, inclusion criteria, and benchmarks lead to inconsistencies, complicating direct comparisons of performance metrics; and (ii) Absence of Standardized Metrics. The lack of uniform evaluation metrics and reporting protocols across studies further hinders the conclusive identification of the best approach. In lung cancer detection, performance metrics such as DSC, sensitivity, specificity, and accuracy are critical for evaluating diagnostic models. The DSC measures the overlap between predicted and actual tumor regions, with higher values indicating better segmentation accuracy. Sensitivity reflects the model's ability to correctly identify true positives, which is essential for early detection of malignant lesions, potentially improving treatment outcomes. Specificity assesses the model's capacity to correctly identify true negatives, thereby minimizing false positives and avoiding unnecessary diagnostic procedures. While accuracy provides an overall measure of correct predictions, it can be misleading in imbalanced datasets, as it may not adequately reflect true model performance. Müller et al. [187] highlight the importance of achieving an optimal balance between sensitivity and specificity to enhance patient management and ensure reliable diagnostic applications in clinical practice. However, as Liu et al. [189] point out, task-agnostic metrics like DSC, Jaccard Similarity Coefficient (JSC), and Hausdorff Distance (HD), though commonly used, do not always correlate with clinically relevant outcomes. Their study underscores the need for task-based evaluation of segmentation algorithms, particularly in assessing quantitative measures such as metabolic tumor volume (MTV) and total lesion glycolysis (TLG). They demonstrated that algorithms performing well on DSC may exhibit significant biases in estimating clinically important metrics. For example, deeper U-Net models improved MTV and TLG estimation accuracy despite showing similar performance on task-agnostic metrics compared to shallower networks. These findings emphasize the limitations of

traditional metrics in reflecting real-world clinical efficacy and the necessity of incorporating task-based metrics into model evaluation.

Our recent study [188] focused on evaluating metrics for segmentation and other tasks such as classification, regression, clustering, and image-to-image translation across Python libraries, R packages, and MATLAB functions. The findings revealed that Accuracy, Precision, and Recall for 2D segmentation, as well as Accuracy for 3D segmentation, are consistent across programming languages, underscoring their reliability for benchmarking segmentation tasks. Recognizing that metrics from various studies may not be directly comparable due to differences in study design, participant characteristics, and evaluation methods is essential.

Computational Complexity: When selecting an appropriate 2D segmentation technique for clinical use, it is important to consider computational complexity and user-friendliness. A comparison of eight DL models (UNet, SegNet, GCN, FCN, DeepLabV3+, PspNet, TransUNet, and SwinUNet) for lung nodule segmentation in CT scans found that TransUNet, with its transformer-based encoder, had the highest accuracy (DSC: 0.813) but was the slowest. The performance of all models improved with the extraction of the ROI over using entire CT scans, while adding lung masks had little to no benefit and sometimes hindered performance. Thus, for precise segmentation, transformer-based models like TransUNet with ROI extraction are recommended, but traditional CNNs such as UNet or SegNet are better for quicker tasks [32].

The review indicates the prevalence of models used for 2D lung cancer segmentation notably UNet. UNet is particularly effective in medical segmentation, such as lung nodule detection, due to its strong performance with limited data. Although it requires substantial annotated data for optimal results, UNet can deliver accurate segmentation with fewer labeled samples, making it ideal for medical applications. Derivatives like SegNet and Attention UNet further improve boundary refinement and feature capture, though UNet's complexity may limit its clinical use in some settings. [32, 190, 191]. CNNs excel in medical image analysis, providing accurate segmentation and versatility across modalities. However, challenges in generalizability and image-specific adjustments impact their clinical effectiveness [192]. The use of advanced CNN models for lung cancer detection and segmentation has significantly improved precision, efficiency, and clinical relevance, aiding diagnosis and treatment planning. However, CNNs may struggle with capturing long-range dependencies in complex structures, where transformers—especially in dual-path architectures—are emerging as a strong alternative for enhanced 2D medical image segmentation [145]. However, their complexity, data demands, potential for overfitting in small datasets [80], and high computational costs pose challenges [146]. Future research seeks

to enhance transformers' interpretability, reduce dependence on large datasets, and improve efficiency. Integrating transformers with CNNs and utilizing self-supervised learning could further advance medical image segmentation. DL models, especially UNet and its variations, are effective in lung nodule 2D segmentation, benefiting from their ability to learn from limited data and perform well across various datasets. However, their success is heavily dependent on the quality and amount of training data and requires significant computational power, posing challenges in resource-limited settings. Models like ResNet and DenseNet, with their DL, improve detection accuracy but also demand substantial computational resources, limiting their use in less equipped environments. Hybrid approaches like Mask R-CNN with K-means and Fuzzy C-Means segmentation offer alternatives that require fewer labeled images or provide computational efficiency, respectively, yet they compromise accuracy or computational demands compared to full DL methods. Despite the high accuracy and adaptability of DL models in lung cancer segmentation, their need for extensive computational resources, large, and annotated datasets, and concern over interpretability and transparency in clinical decisions highlight crucial areas for advancement to make these models more clinically viable and understandable.

This study identifies biases in 2D segmentation models for lung cancer, with a focus on UNet variants. UNet has rapidly emerged as a crucial model for medical image segmentation, thanks to its exceptional performance and distinctive architecture. UNet models are extensively utilized in 2D medical image segmentation tasks, such as identifying lung cancer in CT scans. Their encoder-decoder architecture, complemented by skip connections, adeptly captures both low-level and high-level image features, essential for precise segmentation of intricate regions like lung nodules. Despite their strengths, including straightforward implementation and suitability for processing 2D CT slices with lower computational demands compared to 3D models, UNet faces several challenges that require attention for enhanced clinical effectiveness. Key limitations include class imbalance in medical datasets, which affects model accuracy, especially in differentiating benign from malignant nodules. Although data augmentation is a common remedy, its adoption is not universal, with only 63% of studies employing this technique. This suggests a need for more sophisticated data balancing methods. Additionally, the models often fail to accurately segment smaller lesions due to the limited receptive field of the convolutional layers. Another concern is the stability of UNet models; only a small fraction of studies assess model performance consistency over different initializations, highlighting a gap in robustness evaluation. This is crucial for ensuring that models do not overfit and perform well across various training conditions. Furthermore, while UNet's adaptability is notable, its generalizability

across diverse datasets remains questionable. Many studies lack external validation, particularly on non-public datasets, which limits their applicability in real-world settings. To address these issues, future research should focus on more effective strategies for managing class imbalance and enhancing small lesion detection. Enhancing model stability through comprehensive evaluations and extending external validation efforts will improve reliability and adaptability in clinical environments. In the review of 2D segmentation for lung cancer CT images, several methodological insights emerge. Only 22% of UNet model studies use non-public datasets for external validation, potentially limiting the generalizability of the findings. Furthermore, while 89% of the studies include rigorous analyses such as ablation studies or comparisons with state-of-the-art models, 25% highlight top metrics without statistical validation. This lack of statistical rigor could lead to misleading conclusions, underscoring the need for more thorough validation in segmentation research.

Research Gaps and Limitations. Usability, Reusability, Reproducibility, and Accessibility (URRA) have become increasingly vital in the scientific community, especially within medical imaging research [66, 193–201]. Our review of studies from 2016 to 2024 highlights ongoing URRA-related challenges, including computational complexity, training time, and robustness to data variability. For example, a 2016 Nature survey [202] found that over 70% of researchers failed to reproduce others' experiments and more than half couldn't replicate their own. Subsequent studies [203–208] consistently reported low replication rates, with only 64–85% of results successfully replicated and many studies struggling with issues related to computational complexity and training time. Additional surveys [209] echoed these concerns, emphasizing how robustness to data variability and other factors contribute to a growing reproducibility crisis that hampers the timely translation of techniques into clinical applications [210]. However, this review did not encompass practical clinical considerations such as computational complexity, training time, and robustness to data variability. Furthermore, future studies should focus on evaluating reproducibility and reusability by re-implementing and comparing previous research using the same or various private and public datasets.

The paper provides a comprehensive review of the progress in computational methods for lung cancer segmentation, particularly emphasizing the superiority of DL models in terms of accuracy and flexibility. Despite these advances, the review identifies several unresolved challenges that need attention for further research. A critical issue highlighted is the trade-off between the speed and accuracy of these models, where models designed for quick processing compromise on detailed accuracy, which is crucial for clinical diagnosis. Conversely, highly accurate models require significant computational resources and time, potentially limiting

their practical use in clinical settings. The "black box" nature of DL models, due to their complexity and lack of transparency, poses another significant challenge. This opacity affects their acceptance among clinicians, who prefer a transparent decision-making process. Moreover, these models' performance varies across different datasets, raising concerns about their generalizability and effectiveness in diverse clinical environments. In computational lung cancer studies, developing and validating robust segmentation models using CT images faces several key challenges. Datasets often lack diversity in patient demographics, imaging protocols, and disease prevalence, primarily representing North American populations, which hampers model generalizability across varied clinical scenarios. Additionally, many datasets suffer from imbalanced class distributions, such as fewer positive cases and varying nodule sizes, increasing the risk of overfitting and biased performance. Inconsistent and inaccurate annotations, along with incomplete labeling and the absence of contextual clinical information, introduce noise and uncertainty, undermining model accuracy and effectiveness. The scarcity of sufficient external testing datasets with diverse and comprehensive data further restricts thorough validation in real-world settings. Lastly, relying solely on CT imaging neglects the potential benefits of integrating multi-modality data, which could enhance segmentation precision and overall model performance.

Data Variability: A major challenge in deep learning-based segmentation is the variability in datasets, which arises from differences in imaging protocols, patient demographics, tumor types, and imaging quality across different centers. This variability can lead to inconsistent model performance, limiting the model's applicability to diverse patient populations. Furthermore, class imbalance remains a critical issue in medical imaging, especially in the context of lung cancer, where benign lesions are far more common than malignant ones. Inconsistent preprocessing methods (e.g., variations in image normalization, cropping, and resizing) further exacerbate these issues. To address these concerns, it is essential to promote the use of multi-center, diverse datasets that capture a wide range of patient populations, imaging conditions, and disease severities. Data augmentation techniques, such as rotation, scaling, and elastic deformation, should be employed to increase dataset diversity and improve model robustness. Additionally, strategies to address class imbalance, such as oversampling minority classes or using focal loss, could help to balance the impact of benign and malignant lesions during training. To address these gaps, we propose several strategies to enhance dataset quality and diversity. Collecting data from diverse populations and global sources can improve model generalizability. Moreover, employing data augmentation and multi-center collaborations can increase dataset size and diversity, enhancing robustness and reducing overfitting. Furthermore,

developing semi-supervised or weakly supervised learning methods can leverage both labeled and unlabeled data, boosting performance with limited annotations. In addition, standardizing annotation quality through consistent protocols will ensure accurate training data. Expanding external testing datasets with diverse samples will facilitate thorough validation, while integrating clinical and radiological data into multi-modality datasets (e.g., PET, MRI) can provide a comprehensive view of lung cancer, enhancing segmentation precision and model performance.

In Supplemental File 1, we provide a detailed table evaluating the most frequently used models, including UNet and its variants, alongside other prominent approaches for 2D medical image segmentation. The table outlines the key advantages and disadvantages of each model. Among the advantages, these models generally offer a high DSC, demonstrating strong segmentation accuracy, particularly in complex medical imaging tasks. Their ease of implementation makes them accessible for a wide range of clinical applications, and their adaptability to various 2D segmentation tasks further enhances their versatility, enabling effective segmentation of different types of lesions, including lung cancer. However, these models are not without their challenges. One of the primary disadvantages is their computational complexity, especially in advanced variants, which can make them less suitable for resource-constrained environments. Additionally, these models are highly sensitive to data preprocessing steps, such as image normalization, cropping, and resizing, which can affect their performance. Another limitation is the difficulty in accurately segmenting small nodules, as traditional models often struggle with limited receptive fields and may fail to detect or precisely delineate smaller lesions. Despite these challenges, the table provides valuable insights into the strengths and weaknesses of the most commonly used segmentation models, helping guide future research and clinical adoption.

Innovation and Future Direction. The emergence of novel architectures, such as the DPBET [145] and SW-UNet [80], underscores the evolving landscape of DL models for medical image analysis. These models leverage dual-path architectures and hybrid CNN-Transformer mechanisms to capture both local details and global context, offering improved segmentation accuracy. For instance, Wang et al. [145] introduced the Dual-Path Boundary Enhancement Transformer (DPBET) model, a novel approach for lung nodule segmentation that combines global (CNN-Transformer) and edge (DASample) pathways. This dual-path network, incorporating a Cascade-Axial-Prune Transformer (CAP-Trans) for global feature extraction and a boundary enhancement mechanism for edge detection, improves both the recognition of broad contextual information and the precise delineation of nodule boundaries. This addresses the common challenge in traditional CNN-based models,

which struggle to balance capturing local details and global context, particularly due to their limited receptive fields. The integration of the DASample module further enhances boundary awareness, which is critical for accurate nodule segmentation in medical imaging. These recent advancements in lung nodule segmentation demonstrate the power of hybrid CNN-Transformer models and dual-path architecture. By effectively combining local detail extraction with global context modeling, these models address the limitations of traditional approaches and provide more accurate and efficient solutions for medical image segmentation. The application of transformers allows for better handling of large-scale contextual information, improving the precision of lung nodule detection and segmentation across various models and techniques.

Future work in lung cancer segmentation should focus on integrating multimodal data, such as CT, PET, MRI, and histopathology, to leverage complementary information for enhanced tumor characterization. Extending 2D methods to 3D segmentation can capture spatial context and volumetric features more effectively, while semi-supervised and unsupervised learning techniques, including generative adversarial networks and contrastive learning, can address the scarcity of annotated data. Federated learning frameworks can enable collaborative training across institutions while preserving data privacy, and attention mechanisms, such as transformers, can improve focus on critical regions within complex lung structures. Incorporating clinical data like patient history and genetic profiles can enhance personalized models, while explainable AI techniques, such as saliency maps, can make models more interpretable for clinical decision-making. Lightweight, real-time models should be developed for resource-constrained settings and intraoperative use, and longitudinal studies focusing on tumor evolution can aid early detection and monitoring. Domain adaptation techniques are essential to improve generalization across datasets, and synthetic data generation using GANs can address class imbalances and enrich training datasets. Additionally, integrating segmentation outputs with radiomics and post-segmentation analytics can support downstream tasks like tumor classification, risk prediction, and survival analysis, advancing both accuracy and clinical utility in lung cancer management.

Nonetheless, the increased model complexity and the need for extensive dataset validation pose challenges to their widespread adoption. Future research should focus on refining these models, optimizing computational resources, and enhancing their generalizability across diverse medical imaging tasks. To enhance the comparison of various 2D segmentation methods, it is crucial to standardize evaluation methodologies. Understanding how image quality and nodule characteristics influence segmentation performance is essential for improving accuracy. Additionally, exploring

the fusion of 2D techniques with other approaches such as 3D segmentation or ML can lead to increased robustness and precision. The integration of 2D and 3D segmentation methods is highly beneficial in medical imaging, enhancing accuracy, precision, and contextual awareness. While 2D segmentation provides detailed, slice-level analysis, it can miss broader spatial relationships, which 3D segmentation captures across entire volumes. Combining both methods reduces segmentation errors, improves efficiency, and handles complex anatomical structures more effectively.. The paper is optimistic about the potential of advanced DL models like Long Short-Term Memory (LSTM) networks and GANs in addressing current limitations and advancing segmentation performance. The continuous development of DL models, along with considerations for ethical and computational issues, is crucial for fully realizing the potential of these technologies in clinical settings [148].

In the context of 2D segmentation analysis, the pivotal insight gleaned from the literature review underscores the imperative to prioritize semi-supervised and unsupervised models. This strategic emphasis stems from the inherent complexities associated with image labeling, the limited availability of data, and the critical role played by radiologists in the segmentation process. The research findings advocate for a shift towards leveraging semi-supervised and unsupervised approaches as viable solutions to mitigate the challenges posed by labor-intensive image annotation, data scarcity, and the need for expert input in segmentation tasks. By embracing these advanced modeling techniques, researchers and practitioners can enhance the efficiency and effectiveness of 2D segmentation processes, ultimately advancing the field towards more robust and accurate segmentation outcomes.

Reusability and reproducibility are critical in scientific research, especially in medical imaging. Recent studies reveal challenges such as high computational demands, long training times, and the need for robustness across diverse datasets. Many researchers struggle to replicate experiments, leading to low replication success rates. Data variability and limited resources contribute to a reproducibility crisis, delaying the clinical adoption of new techniques. Future research should prioritize evaluating reproducibility and reusability by re-implementing and comparing previous studies using consistent or various private and public datasets.

In clinical settings, sensitivity and specificity directly influence diagnostic accuracy and patient outcomes, making their optimization critical. Meanwhile, relying solely on task-agnostic metrics may lead to overestimation of model performance and hinder clinical translation. Therefore, we advocate for a balanced approach, integrating standardized task-agnostic and task-based evaluations, as suggested by Liu et al. [189], to ensure robust and clinically relevant model assessments. These findings reinforce the importance

of designing evaluation frameworks that align with real-world clinical needs, promoting the development of reliable and impactful diagnostic tools in lung cancer detection. This dual approach not only addresses current gaps in model evaluation but also ensures better translation of research into clinical practice, ultimately benefiting patient care. Moreover, broadening the scope of evaluation metrics to include geometric and uncertainty-based measures, such as the Hausdorff Distance and Free-Response Receiver Operating Characteristics (FROC), will provide a more nuanced understanding of model performance and clinical utility.

Conclusions

This review has aimed to capture the current landscape of 2D segmentation techniques for lung cancer in CT images, underscoring the pivotal role of DL models in revolutionizing this domain. Our comprehensive analysis reveals a field at the nexus of rapid technological advancements and enduring challenges, where the efficiency and adaptability of 2D models shines, albeit not without limitations. The emergence of hybrid models and innovative architecture heralds a promising pathway to surmount these barriers, highlighting the imperative for sustained research and development. To harness the full spectrum of DL's capabilities in medical imaging, future endeavors must prioritize technological innovation and computational efficiency. Continued pursuit of these factors will be crucial in advancing the field towards more accurate, accessible, and reliable diagnostic tools for lung cancer, ultimately enhancing patient care and outcomes in oncology. Current research on 2D segmentation UNet models for lung cancer highlights progress in transparency and preprocessing but reveals gaps in stability, cross-validation, and evaluation metrics. Addressing these gaps and socioeconomic factors is crucial for developing robust, equitable models applicable across diverse clinical settings.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s10278-025-01458-x>.

Acknowledgements This study was the Technological Virtual Collaboration Corporation (TECVICO CORP.), Vancouver, BC, Canada.

Author Contributions Conceptualization, M.S and S.M; data analysis, Zh.S, S.M; data gathering and organizing, Zh.S,S.M, F.Ph, A.F and A.M; writing—original draft preparation, S.M, and M.S; writing—review and editing, M.S, A.R, I.H, R.Y; visualization, M.S, A.R, I.H, R.Y; supervision M.S; All authors have read and agreed to the published version of the manuscript.

Funding This study was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grant RGPIN-2019–06467, UBC Department of Radiology 2023–25 AI Fund.

Data Availability As this is a review paper, no primary data were collected or analyzed. All relevant studies have been cited appropriately in the main text.

Declarations

This article undertakes a systematic review guided by the Preferred Reporting Items for Systematic Reviews and Me-ta-Analysis (PRISMA) framework.

Competing Interests The authors have no relevant financial or non-financial interests to disclose.

References

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin*. 2021;71(3):209–49.
2. Cancer facts & figures 2024. American Cancer Society. 2024. Accessed June 18. Available from: <https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/annual-cancer-facts-and-figures/2024/2024-cancer-facts-and-figures-acsc.pdf>.
3. Kratzer TB, Bandi P, Freedman ND, Smith RA, Travis WD, Jemal A, et al. Lung cancer statistics, 2023. *Cancer*. 2024;130(8):1330–48.
4. Jemal A, Schafer EJ, Star J, Bandi P, Sung H, Islami F, et al. Lung cancer incidence rates in young women and men by state in the United States. *International Journal of Cancer*. 2025;156(3):499–504.
5. Luo G, Zhang Y, Etxeberria J, Arnold M, Cai X, Hao Y, et al. Projections of Lung Cancer Incidence by 2035 in 40 Countries Worldwide: Population-Based Study. *JMIR Public Health Surveill*. 2023;9:e43651.
6. Cruickshank A, Stieler G, Ameer F. Evaluation of the solitary pulmonary nodule. *Internal medicine journal*. 2019;49(3):306–15.
7. Team NLSTR. Reduced lung-cancer mortality with low-dose computed tomographic screening. *New England Journal of Medicine*. 2011;365(5):395–409.
8. Salmanpour MR, Shamsaei M, Saberi A, Setayeshi S, Taherinezhad E, Klyuzhin IS, et al. Machine Learning Methods for Optimal Prediction of Outcome in Parkinson's Disease. 2018 IEEE Nuclear Science Symposium and Medical Imaging Conference Proceedings (NSS/MIC). 2018;1–5.
9. Salmanpour MR, Hajianfar G, Rezaeijo SM, Ghaemi M, Rahmim A, editors. Advanced Automatic Segmentation of Tumors and Survival Prediction in Head and Neck Cancer. *Head and Neck Tumor Segmentation and Outcome Prediction; 2022 2022//; Cham: Springer International Publishing*.
10. Salmanpour M, Saberi A, Shamsaei M, Rahmim A. Optimal Feature Selection and Machine Learning for Prediction of Outcome in Parkinson's Disease. *Journal of Nuclear Medicine*. 2020;61(supplement 1):524.
11. Salmanpour MR, Rezaeijo SM, Hosseinzadeh M, Rahmim A. Deep versus Handcrafted Tensor Radiomics Features: Prediction of Survival in Head and Neck Cancer Using Machine Learning and Fusion Techniques. *Diagnostics*. 2023; 13(10).
12. Salmanpour M, Shamsaei M, Saberi A, Hajianfar G, Ashrafinia S, Davoodi-Bojd E, et al. Hybrid Machine Learning Methods for Robust Identification of Parkinson's Disease Subtypes. *Journal of Nuclear Medicine*. 2020;61(supplement 1):1429.
13. Javanmardi J, Hosseinzadeh M, Hajianfar G, Nabizadeh AH, Rezaeijo S, Rahmim A, et al., Multi-modality fusion coupled

- with deep learning for improved outcome prediction in head and neck cancer. *ProcSPIE*. 2022; 1203221.
14. Hajianfar G, Kalayinia S, Hosseinzadeh M, Samanian S, Maleki M, Rezaeijo SM, et al. Hybrid Machine Learning Systems for Prediction of Parkinson's Disease Pathogenic Variants using Clinical Information and Radiomics Features. *Journal of Nuclear Medicine*. 2022;63(supplement 2):2508.
 15. Schabath MB, Aberle DR. MILD trial, strong confirmation of lung cancer screening efficacy. *Nature Reviews Clinical Oncology*. 2019;16(9):529-30.
 16. Setio AAA, Traverso A, de Bel T, Berens MSN, Bogaard CVD, Cerello P, et al. Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images: The LUNA16 challenge. *Medical image analysis*. 2017;42:1-13.
 17. Pesapane F, Gnocchi G, Quarrella C, Sorce A, Nicosia L, Mariano L, et al. Errors in Radiology: A Standard Review. *Journal of Clinical Medicine*. 2024;13(15):40306.
 18. Fatan M, Hosseinzadeh M, Askari D, Sheikhi H, Rezaeijo SM, Salmanpour MR. Fusion-Based Head and Neck Tumor Segmentation and Survival Prediction Using Robust Deep Learning Techniques and Advanced Hybrid Machine Learning Systems. *Head and Neck Tumor Segmentation and Outcome Prediction*. Cham: Springer International Publishing.
 19. Fu KS, Mui JK. A survey on image segmentation. *Pattern Recognition*. 1981;13(1):3-16.
 20. Tang Y, Zhang N, Wang Y, He S, Han M, Xiao J, et al., editors. Accurate and Robust Lesion RECIST Diameter Prediction and Segmentation with Transformers. *Medical Image Computing and Computer Assisted Intervention – MICCAI 2022*; 2022 2022//; Cham: Springer Nature Switzerland.
 21. Salmanpour MR, Shamsaei M, Rahmim A. Feature selection and machine learning methods for optimal identification and prediction of subtypes in Parkinson's disease. *Computer Methods and Programs in Biomedicine*. 2021;206:106131.
 22. Salmanpour MR, Shamsaei M, Saberi A, Hajianfar G, Soltanian-Zadeh H, Rahmim A. Robust identification of Parkinson's disease subtypes using radiomics and hybrid machine learning. *Computers in Biology and Medicine*. 2021;129:104142.
 23. Rahmim A, Toosi A, Salmanpour MR, Dubljevic N, Janzen I, Shiri I, et al., Tensor Radiomics: Paradigm for Systematic Incorporation of Multi-Flavoured Radiomics Features. 2022 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC). 2022; 1–4.
 24. Baidya Kayal E, Kandasamy D, Yadav R, Bakhshi S, Sharma R, Mehndiratta A. Automatic segmentation and RECIST score evaluation in osteosarcoma using diffusion MRI: A computer aided system process. *European Journal of Radiology*. 2020;133:109359.
 25. Salmanpour MR, Hosseinzadeh M, Rezaeijo SM, Rahmim A. Fusion-based tensor radiomics using reproducible features: Application to survival prediction in head and neck cancer. *Computer Methods and Programs in Biomedicine*. 2023;240:107714.
 26. Ulku I, Akagündüz E. A Survey on Deep Learning-based Architectures for Semantic Segmentation on 2D Images. *Applied Artificial Intelligence*. 2022;36(1):2032924.
 27. Badrinarayanan V, Kendall A, Cipolla R. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE transactions on pattern analysis and machine intelligence*. 2017;39(12):2481-95.
 28. He K, Zhang X, Ren S, Sun J, editors. Deep residual learning for image recognition. *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016;770–778.
 29. Zhang Z, Yao L, Keles E, Velichko Y, Bagci U. Deep Learning Algorithms for Pancreas Segmentation from Radiology Scans: A Review. *Advances in Clinical Radiology*. 2023;5(1):31-52.
 30. Zhi L, Jiang W, Zhang S, Zhou T. Deep neural network pulmonary nodule segmentation methods for CT images: Literature review and experimental comparisons. *Computers in Biology and Medicine*. 2023;164:107321.
 31. Carmo D, Ribeiro J, Dertkigil S, Appenzeller S, Lotufo R, Rittner L. A Systematic Review of Automated Segmentation Methods and Public Datasets for the Lung and its Lobes and Findings on Computed Tomography Images. *Yearbook of medical informatics*. 2022;31(1):277-95.
 32. Zhang J, Ying C, Ye Z, Ma D, Wang B, Cheng Y. Recent developments in segmentation of COVID-19 CT images using deep-learning: An overview of models, techniques and challenges. *Biomedical Signal Processing and Control*. 2024;91:105970.
 33. Qin Q, Chen Y. A review of retinal vessel segmentation for fundus image analysis. *Engineering Applications of Artificial Intelligence*. 2024;128:107454.
 34. Han K, Sheng VS, Song Y, Liu Y, Qiu C, Ma S, et al. Deep semi-supervised learning for medical image segmentation: A review. *Expert Systems with Applications*. 2024;245:123052.
 35. Khajuria R, Sarwar A. Review of reinforcement learning applications in segmentation, chemotherapy, and radiotherapy of cancer. *Micron*. 2024;178:103583.
 36. Wang T-W, Hsu M-S, Lee W-K, Pan H-C, Yang H-C, Lee C-C, et al. Brain metastasis tumor segmentation and detection using deep learning algorithms: A systematic review and meta-analysis. *Radiotherapy and Oncology*. 2024;190:110007.
 37. Ahamed MF, Hossain MM, Nahiduzzaman M, Islam MR, Islam MR, Ahsan M, et al. A review on brain tumor segmentation based on deep learning methods with federated learning techniques. *Computerized Medical Imaging and Graphics*. 2023;110:102313.
 38. Jiang H, Diao Z, Yao Y-D. Deep learning techniques for tumor segmentation: a review. *The Journal of Supercomputing*. 2022;78(2):1807-51.
 39. Ciceri T, Squarcina L, Giubergia A, Bertoldo A, Brambilla P, Peruzzo D. Review on deep learning fetal brain segmentation from Magnetic Resonance images. *Artificial Intelligence in Medicine*. 2023;143:102608.
 40. Chen C, Zhou K, Wang Z, Zhang Q, Xiao R. All answers are in the images: A review of deep learning for cerebrovascular segmentation. *Computerized Medical Imaging and Graphics*. 2023;107:102229.
 41. Wang P, Yang Q, He Z, Yuan Y. Vision transformers in multi-modal brain tumor MRI segmentation: A review. *Meta-Radiology*. 2023;1(1):100004.
 42. Huang Q, Tian H, Jia L, Li Z, Zhou Z. A review of deep learning segmentation methods for carotid artery ultrasound images. *Neurocomputing*. 2023;545:126298.
 43. Abbasi H, Orouskhani M, Asgari S, Zadeh SS. Automatic brain ischemic stroke segmentation with deep learning: A review. *Neuroscience Informatics*. 2023;3(4):100145.
 44. Zhang L, Li W, Lv J, Xu J, Zhou H, Li G, et al. Advancements in oral and maxillofacial surgery medical images segmentation techniques: An overview. *Journal of Dentistry*. 2023;138:104727.
 45. Xun S, Li D, Zhu H, Chen M, Wang J, Li J, et al. Generative adversarial networks in medical image segmentation: A review. *Computers in Biology and Medicine*. 2022;140:105063.
 46. Gu W, Bai S, Kong L. A review on 2D instance segmentation based on deep neural networks. *Image and Vision Computing*. 2022;120:104401.
 47. Qurratulain S, Zheng Z, Xia J, Ma Y, Zhou F. Deep learning instance segmentation framework for burnt area instances characterization. *International Journal of Applied Earth Observation and Geoinformation*. 2023;116:103146.
 48. Marques R, Andrade De Jesus D, Barbosa-Breda J, Van Eijgen J, Stalmans I, van Walsum T, et al. Automatic Segmentation of the Optic Nerve Head Region in Optical Coherence Tomography:

- A Methodological Review. *Computer Methods and Programs in Biomedicine*. 2022;220:106801.
49. Kaur H, Kaur N, Neeru N. Evolution of multiorgan segmentation techniques from traditional to deep learning in abdominal CT images – A systematic review. *Displays*. 2022;73:102223.
 50. Thanoon MA, Zulkifley MA. A Review of Deep Learning Techniques for Lung Cancer Screening and Diagnosis Based on CT Images. *Diagnostics*. 2023;13(16):2617.
 51. Carvalho LE, Sobieranski AC, von Wangenheim A. 3D Segmentation Algorithms for Computerized Tomographic Imaging: a Systematic Literature Review. *Journal of digital imaging*. 2018;31(6):799–850.
 52. Srikrishna M, Heckemann RA, Pereira JB, Volpe G, Zettergren A, Kern S, et al. Comparison of Two-Dimensional- and Three-Dimensional-Based U-Net Architectures for Brain Tissue Classification in One-Dimensional Brain CT. *Frontiers in computational neuroscience*. 2021;15:785244.
 53. Nemoto T, Futakami N, Yagi M, Kumabe A, Takeda A, Kunieda E, et al. Efficacy evaluation of 2D, 3D U-Net semantic segmentation and atlas-based segmentation of normal lungs excluding the trachea and main bronchi. *J Radiat Res*. 2020;61(2):257–64.
 54. Yanqi S, Aiping J, Xiaojing L, Guotao W. Lung CT Image Segmentation Based on Capsule Network. *Proceedings of the 2020 4th International Conference on Electronic Information Technology and Computer Engineering*. 2021;386–90.
 55. Dong L, Liu H, editors. Segmentation of Pulmonary Nodules Based on Improved UNet++. 14th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI). 2021; 23–25.
 56. Chen W, Wang Q, Huang S, Zhang X, Li Y, Liu C, editors. DFDM: A Deep Feature Decoupling Module for Lung Nodule Segmentation. *ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*; 2021 6–11 June 2021.
 57. Kumar M, Kumar V, Sharma A. Segmentation and Prediction of Lung Cancer CT Scans through Nodules using Ensemble Deep Learning Approach. 2021 IEEE Mysore Sub Section International Conference (MysuruCon). 2021; 781–785.
 58. S G, D. B B, N. D K, Verma D, Selvam PP, Sharma DK, editors. Identifying Lung Nodules on MRR Connected Feature Streams for Tumor Segmentation. 2022 4th International Conference on Inventive Research in Computing Applications (ICIRCA); 2022 21–23 Sept. 2022.
 59. Xu W, Xing Y, Lu Y, Lin J, Zhang X, editors. Dual Encoding Fusion for Atypical Lung Nodule Segmentation. 2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI). 2022; 28–31.
 60. Zhang S, Yue J, Wang C, Liu X, Wang G. Box2Pseudo: A Semi-Supervised Learning Framework for Pulmonary Nodule Segmentation with Box-Prompt Pseudo Supervision. 2023 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). 2023;1696–1703.
 61. Pawar P, Badjate S, Gulhane S. Lung Nodule Segmentation using Deep Learning and Advanced UNet Model. *sms*. 2022;14(1 SPL):pp. 99–107.
 62. Kumar Illa P. An efficient lung nodule detection model using multiscale hybrid 2D-3D dilation assisted adaptive Trans-Res-UNet++ with hybrid heuristic mechanism. *Biomedical Signal Processing and Control*. 2024;90:105849.
 63. Banu SF, Sarker MM, Abdel-Nasser M, Puig D, Raswan HA. AWEU-Net: An Attention-Aware Weight Excitation U-Net for Lung Nodule Segmentation. *Applied Sciences*. 2021; 11(21):10132.
 64. Wu Z, Li X, Zuo J. RAD-UNet: Research on an improved lung nodule semantic segmentation algorithm based on deep learning. *Frontiers in oncology*. 2023;13:1084096.
 65. Halder A, Chatterjee S, Dey D, Kole S, Munshi S. An adaptive morphology based segmentation technique for lung nodule detection in thoracic CT image. *Computer Methods and Programs in Biomedicine*. 2020;197:105720.
 66. Adrash R, Gadipudi A, Pandeewari R, Deivalakshmi S. Inception block based Residual Auto Encoder For Lung Segmentation. 2020 4th International Conference on Computer, Communication and Signal Processing (ICCCSP). 2020; 28–29.
 67. Chen Y, Wang Y, Hu F, Wang D. A Lung Dense Deep Convolution Neural Network for Robust Lung Parenchyma Segmentation. *IEEE Access*. 2020;8:93527–47.
 68. Bhandary A, Prabhu GA, Rajinikanth V, Thanaraj KP, Satapathy SC, Robbins DE, et al. Deep-learning framework to detect lung abnormality – A study with chest X-Ray and lung CT scan images. *Pattern Recognition Letters*. 2020;129:271–8.
 69. Cao H, Liu H, Song E, Hung C-C, Ma G, Xu X, et al. Dual-branch residual network for lung nodule segmentation. *Applied Soft Computing*. 2020;86:105934.
 70. Bruntha PM, Pandian SIA, Sagayam KM, Bandopadhyay S, Pomplun M, Dang H. Lung_PAYNet: a pyramidal attention based deep learning network for lung nodule segmentation. *Scientific Reports*. 2022;12(1):20330.
 71. Akila Agnes S, Arun Solomon A, Karthick K. Wavelet U-Net++ for accurate lung nodule segmentation in CT scans: Improving early detection and diagnosis of lung cancer. *Biomedical Signal Processing and Control*. 2024;87:105509.
 72. Thakur G. Investigate Uncertainty Estimation in Transfer Learning Based Active Learning for Medical Image Segmentation. 2024 3rd International Conference for Innovation in Technology (INOCIN). 2024; 1–7.
 73. Dutta K. Densely connected recurrent residual (dense r2unet) convolutional neural network for segmentation of lung ct images. *arXiv*. 2021; 2102.00663v1.
 74. Chen K-b, Xuan Y, Lin A-j, Guo S-h. Lung computed tomography image segmentation based on U-Net network fused with dilated convolution. *Computer Methods and Programs in Biomedicine*. 2021;207:106170.
 75. Chiu T-W, Tsai Y-L, Su S-F. Automatic detect lung node with deep learning in segmentation and imbalance data labeling. *Scientific Reports*. 2021;11(1):11174.
 76. Salama WM, Aly MH, Elbagoury AM. Lung Images Segmentation and Classification Based on Deep Learning: A New Automated CNN Approach. *Journal of Physics: Conference Series*. 2021;2128(1):012011.
 77. Zhang X, Liu X, Zhang B, Dong J, Zhang B, Zhao S. Accurate segmentation for different types of lung nodules on CT images using improved U-Net convolutional network. *Medicine (Baltimore)*. 2021;100(40):e27491.
 78. Sousa J, Pereira T. The Influence of a Coherent Annotation and Synthetic Addition of Lung Nodules for Lung Segmentation in CT Scans. *Sensors*. 2022;22(9):3443.
 79. Rani V, Jawhar SJ. Automatic segmentation and classification of lung tumour using advance sequential minimal optimisation techniques. *IET Image Process*. 2020;14:3355–65.
 80. 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. *Frontiers in medicine*. 2023;10:1273441.
 81. Annavarapu CSR, Parisapogu SAB, Keetha NV, Donta PK, Rajita G. A Bi-FPN-Based Encoder–Decoder Model for Lung Nodule Image Segmentation. *Diagnostics*. 2023;13(8):1406.
 82. Naseer I, Akram S, Masood T, Rashid M, Jaffar A. Lung Cancer Classification Using Modified U-Net Based Lobe Segmentation and Nodule Detection. *IEEE Access*. 2023;11:60279–91.
 83. Gao Y, Guo J, Fu C, Wang Y, Cai S. VLSM-Net: A Fusion Architecture for CT Image Segmentation. *Applied Sciences*. 2023;13(7):4384.

84. Qian L, Wen C, Li Y, Hu Z, Zhou X, Xia X, et al. Multi-scale context UNet-like network with redesigned skip connections for medical image segmentation. *Computer Methods and Programs in Biomedicine*. 2024;243:107885.
85. Dutande P, Baid U, Talbar S. LNCDS: A 2D-3D cascaded CNN approach for lung nodule classification, detection and segmentation. *Biomedical Signal Processing and Control*. 2021;67:102527.
86. Kishore R, Satpathy M, Parida DK, Nussinov Z, Sahu KK, editors. *Detection of Lung Nodules Using Unsupervised Machine Learning Method*. Computational Vision and Bio-Inspired Computing; 2020 2020//; Cham: Springer International Publishing.
87. Traoré A, Ly AO, Akhloufi MA. Evaluating deep learning algorithms in pulmonary nodule detection. 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). 2020; 1335–1338.
88. Chi J, Zhang S, Yu X. A Novel Pulmonary Nodule Detection Model Based on Multi-Step Cascaded Networks. *Sensors* 2020; 20(15):4301.
89. Bhattacharyya D, Thirupathi Rao N, Joshua ESN, Hu YC. A bi-directional deep learning architecture for lung nodule semantic segmentation. *The Visual Computer*. 2022; 39(11):1–17.
90. Tang T, Li F, Jiang M, Xia X, Zhang R. Improved Complementary Pulmonary Nodule Segmentation Model Based on Multi-Feature Fusion. *Entropy*. 2022; 24(12):1755.
91. Thangavel C, Palanichamy J. Effective deep learning approach for segmentation of pulmonary cancer in thoracic CT image. *Biomedical Signal Processing and Control*. 2024;89:105804.
92. Tang T, Zhang R, Lin K, Li F, Xia X. SM-RNet: A Scale-Aware-Based Multiattention-Guided Reverse Network for Pulmonary Nodules Segmentation. *IEEE Transactions on Instrumentation and Measurement*. 2023;72:1–14.
93. Zhou Y, Chen M, Zhang M, Wang T, Yan F, Xie C. Automatic Segmentation of Lung Nodules using improved U-Net Network. 2020 Chinese Automation Congress (CAC). 2020;1609–13.
94. Chen T, Luo F, Zhou F, Liu F, Zha Y, editors. ViFUNet: a Vision Flash based UNet for lung nodules segmentation task. 2022 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). 2022;1481–1486.
95. Kongkham D, Beenarani BB, Nirmala P, Kumaresan SJ, Senthilkumar C. Deep Learning Approaches for Segmentation of Lung Nodules in CT Scans. 2024 2nd International Conference on Computer, Communication and Control (IC4). 2024;1–5.
96. Boodi D, Sudheer N, Bidargaddi AP, Shatagar S, Telkar M. Semantic Segmentation of Computed Tomography Scan of Lungs. 2024 5th International Conference for Emerging Technology (INCET). 2024;1–7.
97. Wu Z, Zhou Q, Wang F. Coarse-to-Fine Lung Nodule Segmentation in CT Images With Image Enhancement and Dual-Branch Network. *IEEE Access*. 2021;9:7255–62.
98. Sebastian K, Devi S. Detecting Lung Nodules Based on Deep Learning with Automatic Segmentation. 2023 First International Conference on Advances in Electrical, Electronics and Computational Intelligence (ICAEECI). 2023;1–7.
99. Bhakta R, Hossain AA. Lung tumor segmentation and staging from ct images using fast and robust fuzzy C-Means clustering. *International Journal of Image, Graphics and Signal Processing*. 2020;11(1):38.
100. Hu Q, de F. Souza LF, Holanda GB, Alves SSA, dos S. Silva FH, Han T, et al. An effective approach for CT lung segmentation using mask region-based convolutional neural networks. *Artificial Intelligence in Medicine*. 2020;103:101792.
101. Singadkar G, Mahajan A, Thakur M, Talbar S. Deep Deconvolutional Residual Network Based Automatic Lung Nodule Segmentation. *Journal of Digital Imaging*. 2020;33(3):678–84.
102. Yu K-H, Lee T-LM, Yen M-H, Kou SC, Rosen B, Chiang J-H, et al. Reproducible Machine Learning Methods for Lung Cancer Detection Using Computed Tomography Images: Algorithm Development and Validation. *J Med Internet Res*. 2020;22(8):e16709.
103. Sathish R, Sathish R, Sethuraman R, Sheet D. Lung Segmentation and Nodule Detection in Computed Tomography Scan using a Convolutional Neural Network Trained Adversarially using Turing Test Loss. 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). 2020;1331–1334.
104. Wu Y, Lin L. Automatic lung segmentation in CT images using dilated convolution based weighted fully convolutional network. *Journal of Physics: Conference Series*; 2020;1646:012032.
105. Rocha J, Cunha A, Mendonça AM. Conventional Filtering Versus U-Net Based Models for Pulmonary Nodule Segmentation in CT Images. *Journal of Medical Systems*. 2020;44(4):81.
106. Basha CZ, Pravalika BL, Vineela D, Prathyusha SL. An Effective and Robust Cancer Detection in the Lungs with BPNN and Watershed Segmentation. 2020 International Conference for Emerging Technology (INCET). 2020;1–6.
107. Kuo C-FJ, Huang C-C, Siao J-J, Hsieh C-W, Huy VQ, Ko K-H, et al. Automatic lung nodule detection system using image processing techniques in computed tomography. *Biomedical Signal Processing and Control*. 2020;56:101659.
108. Liu C, Pang M. Extracting Lungs from CT Images via Deep Convolutional Neural Network Based Segmentation and Two-Pass Contour Refinement. *Journal of digital imaging*. 2020; 33(6):1465–78.
109. Hesamian MHJ, Wenjing ; He, Xiangjian et al. Region proposal network for lung nodule detection and segmentation. *CEUR Workshop Proceedings*. 2020;2675.: 48-52.
110. Mathews AB, Jeyakumar MK. Analysis of Lung Tumor Detection using Various Segmentation Techniques. 2020 International Conference on Inventive Computation Technologies (ICICT). 2020;454–58.
111. Samundeeswari P, Gunasundari R. A Novel Multilevel Hybrid Segmentation and Refinement Method for Automatic Heterogeneous True NSCLC Nodules Extraction. 2020 5th International Conference on Devices, Circuits and Systems (ICDCS). 2020;226–35.
112. Hu H, Li Q, Zhao Y, Zhang Y. Parallel Deep Learning Algorithms With Hybrid Attention Mechanism for Image Segmentation of Lung Tumors. *IEEE Transactions on Industrial Informatics*. 2021;17(4):2880–9.
113. Valcan S, Găianu M. Lung Tumor Segmentation Accelerated by CUDA. 2020 22nd International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNAS). 2020;246–50.
114. Wang J, Zhao X, Ning Q, Qian D. AEC-Net: Attention and Edge Constraint Network for Medical Image Segmentation. 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). 2020;1616–19.
115. Gan W, Wang H, Gu H, Duan Y, Shao Y, Chen H, et al. Automatic segmentation of lung tumors on CT images based on a 2D & 3D hybrid convolutional neural network. *The British Journal of Radiology*. 2021;94(1126):20210038.
116. Gu H, Gan W, Zhang C, Feng A, Wang H, Huang Y, et al. A 2D–3D hybrid convolutional neural network for lung lobe auto-segmentation on standard slice thickness computed tomography of patients receiving radiotherapy. *BioMedical Engineering OnLine*. 2021;20(1):94.
117. Hesamian MH, Jia W, He X, Wang Q, Kennedy PJ. Synthetic CT images for semi-sequential detection and segmentation of lung nodules. *Applied Intelligence*. 2021;51(3):1616–28.

118. Farheen F, Shamil MS, Ibtehaaz N, Rahman MS. Segmentation of Lung Tumor from CT Images using Deep Supervision. *arXiv*.2021;211109262.
119. Chen Q, Xie W, Zhou P, Zheng C, Wu D. Multi-Crop Convolutional Neural Networks for Fast Lung Nodule Segmentation. *IEEE Transactions on Emerging Topics in Computational Intelligence*. 2022;6(5):1190–200.
120. Chen X, Duan Q, Wu R, Yang Z. Segmentation of lung computed tomography images based on SegNet in the diagnosis of lung cancer. *Journal of Radiation Research and Applied Sciences*. 2021;14(1):396–403.
121. Wang Z, Zou Y, Liu PX. Hybrid dilation and attention residual U-Net for medical image segmentation. *Computers in Biology and Medicine*. 2021;134:104449.
122. Chi J, Zhang S, Yu X, Wu C, Jiang Y. A Novel Pulmonary Nodule Detection Model Based on Multi-Step Cascaded Networks. *Sensors*. 2020;20(15):4301.
123. Kumar SN, Bruntha PM, Daniel SI, Kirubakar JA, Kiruba RE, Sam S, et al., Lung Nodule Segmentation Using UNet. 2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS). 2021;1:420–24.
124. Salama WM, Aly MH, editors. Lung CT Image Segmentation: A Generalized Framework Based on U-Net Architecture and Preprocessing Models. 2021 31st International Conference on Computer Theory and Applications (ICCTA); 2021 11–13 Dec. 2021.
125. Yoo SJ, Yoon SH. Automated Lung Segmentation on Chest Computed Tomography Images with Extensive Lung Parenchymal Abnormalities Using a Deep Neural Network. *Korean Journal of Radiology*. 2021;22(3):476–88.
126. Devi KY, Sasikala M. Labeling and clustering-based level set method for automated segmentation of lung tumor stages in CT images. *Journal of Ambient Intelligence and Humanized Computing*. 2021;12(2):2299–309.
127. Garau N, Paganelli C, Summers P, Bassis D, Lanza C, Minotti M, et al. A segmentation tool for pulmonary nodules in lung cancer screening: Testing and clinical usage. *Physica Medica*. 2021;90:23–9.
128. Maqsood M, Yasmin S, Mehmood I, Bukhari M, Kim M. An Efficient DA-Net Architecture for Lung Nodule Segmentation. *Mathematics*. 2021;9(13):1457.
129. Xie H, Zhang JF, Li Q. Application of Deep Convolution Network to Automated Image Segmentation of Chest CT for Patients With Tumor. *Frontiers in oncology*. 2021;11:719398.
130. Sousa J, Pereira T, Neves I, Silva F, Oliveira HP. The Influence of a Coherent Annotation and Synthetic Addition of Lung Nodules for Lung Segmentation in CT Scans. *Sensors*. 2022;22(9):3443.
131. Tuba HJ, Ullah MA, Badhon FK. Lung Tumor Segmentation and Detection using U-Net with Dilated Convolutions. 2021 5th International Conference on Electrical Information and Communication Technology (EICT); 2021;1–5.
132. Li Z, Zhang J, Tan T, Teng X, Sun X, Zhao H, et al. Deep Learning Methods for Lung Cancer Segmentation in Whole-Slide Histopathology Images—The ACDC@LungHP Challenge 2019. *IEEE Journal of Biomedical and Health Informatics*. 2021;25(2):429–40.
133. Wu L, Li C, Ma J, Lu X, Leiv H. Deep Neural Networks for Semantic Segmentation of Lung Nodules. 2021 40th Chinese Control Conference (CCC); 2021;8061–65.
134. Divya M, Sathya S. Novel Lung Cancer Classification and Segmentation Algorithm with Web UI Interface. 2022 6th International Conference on Electronics, Communication and Aerospace Technology. 2022;1–9.
135. Dutande P, Baid U, Talbar S. Deep Residual Separable Convolutional Neural Network for lung tumor segmentation. *Computers in Biology and Medicine*. 2022;141:105161.
136. Farheen F, Shamil MS, Ibtehaaz N, Rahman MS. Revisiting segmentation of lung tumors from CT images. *Computers in Biology and Medicine*. 2022;144:105385.
137. Afshar P, Mohammadi A, Plataniotis KN, Farahani K, Kirby J, Oikonomou A, et al. Lung-Originated Tumor Segmentation from Computed Tomography Scan (LOTUS) Benchmark. *arXiv preprint arXiv:220100458*. 2022.
138. Rani KV, Jawhar SJ. Automatic segmentation and classification of lung tumour using advance sequential minimal optimisation techniques. *IET Image Processing*. 2020; 14(14):3355–65.
139. Kumar A, Choudhry MS, editors. Optimum thresholding for nodule segmentation of lung CT images. 2022 IEEE World Conference on Applied Intelligence and Computing (AIC); 2022 17–19 June 2022.
140. Wang C, Xu R, Xu S, Meng W, Xiao J, Peng Q, et al. Softgan: Towards Accurate Lung Nodule Segmentation via Soft Mask Supervision. 2022 IEEE International Conference on Multimedia and Expo (ICME); 2022;1–6.
141. Liang G, Diao Z, Jiang H. Uncertainty Analysis Based Attention Network for Lung Nodule Segmentation from CT Images. *Proceedings of the 2022 6th International Conference on Virtual and Augmented Reality Simulations*. 2022;50–5.
142. Cifci MA. SegChaNet: A Novel Model for Lung Cancer Segmentation in CT Scans. *Applied Bionics and Biomechanics*. 2022;2022:1139587.
143. Wu L, Zhuang J, Chen W, Tang Y. Data augmentation based on multiple oversampling fusion for medical image segmentation. *PloS one*.2022;17(10):e0274522.
144. Zhang N, Lin J, Hui B, Qiao B, Yang W, Shang R, et al. Lung Nodule Segmentation and Recognition Algorithm Based on Multiposition U-Net. *Computational and Mathematical Methods in Medicine*. 2022;2022:5112867.
145. Wang S, Jiang A, Li X, Qiu Y, Li M, Li F. DPBET: A dual-path lung nodules segmentation model based on boundary enhancement and hybrid transformer. *Computers in Biology and Medicine*. 2022;151:106330.
146. Prabakaran J, Selvaraj P. An Approach to Recognise Lung Diseases Using Segmentation and Classification. *Measurement Science Review*. 2023;23(6):254–9.
147. Kadry S, Herrera-Viedma E, Crespo RG, Krishnamoorthy S, Rajinikanth V. Automatic detection of lung nodule in CT scan slices using CNN segmentation schemes: A study. *Procedia Computer Science*. 2023;218:2786–94.
148. Chen W, Wang Y, Tian D, Yao Y. CT Lung Nodule Segmentation: A Comparative Study of Data Preprocessing and Deep Learning Models. *IEEE Access*. 2023;11:34925–31.
149. Riaz Z, Khan B, Abdullah S, Khan S, Islam MS. Lung Tumor Image Segmentation from Computer Tomography Images Using MobileNetV2 and Transfer Learning. *Bioengineering*. 2023;10(8):981.
150. Usman M, Shin Y-G. DEHA-Net: A Dual-Encoder-Based Hard Attention Network with an Adaptive ROI Mechanism for Lung Nodule Segmentation. *Sensors*. 2023;23(4):1989.
151. Cai G-W, Liu Y-B, Feng Q-J, Liang R-H, Zeng Q-S, Deng Y, et al. Semi-Supervised Segmentation of Interstitial Lung Disease Patterns from CT Images via Self-Training with Selective Re-Training. *Bioengineering*. 2023;10(7):830.
152. Gopinath A, Gowthaman P, Venkatachalam M, Saroja M. Computer aided model for lung cancer classification using cat optimized convolutional neural networks. *Measurement: Sensors*. 2023;30:100932.
153. Peng X, Peng J, Ruan P, Lv W. Research and Implementation of Lung Tumor Segmentation Method Based on Improved U-Net Network. 2022 International Conference on Image Processing and Computer Vision (IPCV); 2023;91–5.

154. Akash V, Monish SR, Amith KB, Jayanhi MG, Kannadaguli P. Lung Nodule Segmentation and Classification Using Conv-Unet Based Deep Learning. 2023 International Conference on Network, Multimedia and Information Technology (NMITCON); 2023. <https://doi.org/10.1109/NMITCON58196.2023.10276037>
155. Masood A, Naseem U, Razzak I. Multi-Scale Swin Transformer Enabled Automatic Detection and Segmentation of Lung Metastases Using CT Images. 2023 IEEE 20th International Symposium on Biomedical Imaging (ISBI). 2023;1–5.
156. Bruntha PM, Dhanasekar S, Ahmed LJ, Govindaraj V, Pandian SIA, Abraham SS. Lung_RUNET: a Segmentation Framework for Lung Nodules. 2023 9th International Conference on Advanced Computing and Communication Systems (ICACCS). 2023;1:658–61.
157. Fernandes L, Oliveira HP, editors. Multitask learning approach for lung nodule segmentation and classification in CT images. 2023 IEEE International Conference on Bioinformatics and Biomedicine (BIBM); 2023 5–8 Dec. 2023.
158. Bhattacharjee A, Murugan R, Goel. Automated Lung Nodule Segmentation Using Federated Learning Through CT Images. 2023 8th International Conference on Robotics and Automation Engineering (ICRAE).2023;136–41.
159. Wang C, Xu R, Xu S, Meng W, Xiao J, Zhang X. Accurate Lung Nodule Segmentation With Detailed Representation Transfer and Soft Mask Supervision. IEEE Transactions on Neural Networks and Learning Systems. 2023;1–13.
160. Beghouri I, Hamoudi R, Benssalah M, Sbargoud F. Comparative Analysis Between a Deep Learning Model and a Region-Based Active Contour Model for Pulmonary Nodules Segmentation on CT. 2023 2nd International Conference on Electronics, Energy and Measurement (IC2EM). 2023;1–6.
161. Fan X, Lu Y, Hou J, Lin F, Huang Q, Yan C. DMC-UNet-Based Segmentation of Lung Nodules. IEEE Access. 2023;11:110809-26.
162. Hou J, Yan C, Li R, Huang Q, Fan X, Lin F. Lung Nodule Segmentation Algorithm With SMR-UNet. IEEE Access. 2023;11:34319-31.
163. Ji Z, Zhao Z, Zeng X, Wang J, Zhao L, Zhang X, et al. ResDSda_U-Net: A Novel U-Net-Based Residual Network for Segmentation of Pulmonary Nodules in Lung CT Images. IEEE Access. 2023;11:87775-89.
164. Wang T, Li G, Tang C, Kamiya T. Res BCDU-SCB Net: A Lung CT Image Segmentation based on SC-BLSTM. Proceedings of the 2023 8th International Conference on Biomedical Signal and Image Processing. 2023;42–6.
165. Cai Y, Liu Z, Zhang Y, Yang Z. MDFN: A Multi-level Dynamic Fusion Network with self-calibrated edge enhancement for lung nodule segmentation. Biomedical Signal Processing and Control. 2024;87:105507.
166. R EK, Rajakumar G, Bruntha PM. Performance Analysis of Modified CNN Model In Lung Nodule Segmentation. 2024 International Conference on Inventive Computation Technologies (ICICT). 2024;1343–48.
167. Venkatraman K, Reddy SNP. Augmenting Clinical Decisions with Deep Learning Lung Cancer Image Abnormality Segmentation. 2024 14th International Conference on Cloud Computing, Data Science & Engineering (Confluence). 2024; 674–78
168. Thirupathi P, Ram UN, Kumar VK, Malathi M. Pulmosegnet: CT Nodule Segmentation with Mask R-CNN. 2024 5th International Conference on Innovative Trends in Information Technology (ICITIIT); 2024;1–5.
169. Chen X, Lu Y, Zou W, Hu N, Wang J, Zhang B. Segmentation of Metastasized Pulmonary Nodules of Patients with Thyroid Cancer. Proceedings of the 2024 3rd Asia Conference on Algorithms, Computing and Machine Learning. 2024;367–71.
170. Li A-HA, Aruperes CD, Lai Y-J, Chien T-Y, Chiu Y-L, Chan C-L. Lung Nodule Analysis in CT Images: Deep Learning for Segmentation and Measurement. Proceedings of the 2024 8th International Conference on Medical and Health Informatics. 2024;13–7.
171. Wu Y, Luo X, Xu Z, Guo X, Ju L, Ge Z, et al., Diversified and Personalized Multi-rater Medical Image Segmentation. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition; 2024.11470–79.
172. Yadav DP, Sharma B, Webber JL, Mehbodniya A, Chauhan S. EDTNet: A spatial aware attention-based transformer for the pulmonary nodule segmentation. PloS one. 2024;19(11):e0311080.
173. Albalawi E, Neal Joshua ES, Joys NM, Bhatia Khan S, Shaiba H, Ahmad S, et al., Hybrid healthcare unit recommendation system using computational techniques with lung cancer segmentation. Frontiers in medicine. 2024;11:1429291.
174. Cai J, Zhu H, Liu S, Qi Y, Chen R. Lung image segmentation via generative adversarial networks. Frontiers in Physiology. 2024;15.
175. Das SP, Mitra S. A Comparative Analysis of Deep Learning Architectures for Segmentation in Lung. 2024 IEEE Region 10 Symposium (TENSYP). 2024;1–6.
176. Dugășescu A, Chiru CM, Nan M, Trăscău M, Florea AM. Explainable Cancer Segmentation Through Classification. 2024 IEEE 20th International Conference on Intelligent Computer Communication and Processing (ICCP). 2024;1–8.
177. Modak S, Abdel-Raheem E, Rueda L. GPD-Module: A Lightweight Lung Nodule Detection and Segmentation Framework on Computed Tomography Images Using Uniform Superpixel Generation. IEEE Access. 2024;12:154933-48.
178. Sarah Prithvika PC, Jani Anbarasi L. LA-ResUNet: An Efficient Linear Attention Mechanism in ResUNet for the Semantic Segmentation of Pulmonary Nodules. IEEE Access. 2024;12:182894-907.
179. Ivusic D, Zeger I, Grgic S. Segmentation of Lung Cancer in CT Images using Modified U-Net Architecture. 2024 International Symposium ELMAR. 2024;341–46.
180. Ismail HM, McKee GT. U-Net Segmentation of Lung Cancer CT Scans for 3D Rendering. 2024 5th International Conference on Artificial Intelligence, Robotics and Control (AIRC). 2024; <https://doi.org/10.1109/AIRC61399.2024.10671940>.
181. Pal S, Mitra S, Shankar BU. Weighted Deformable Network for Efficient Segmentation of Lung Tumors in CT. IEEE Transactions on Systems, Man, and Cybernetics: Systems. 2025;55(2):898-909.
182. Janzen I, Ho C, Melosky B, Ye Q, Li J, Wang G, et al. Machine Learning and Computed Tomography Radiomics to Predict Disease Progression to Upfront Pembrolizumab Monotherapy in Advanced Non-Small-Cell Lung Cancer: A Pilot Study. Cancers; 2024;17(1):58.
183. Sharma N, Aggarwal LM. Automated medical image segmentation techniques. J Med Phys. 2010;35(1):3-14.
184. Pang S, Du A, Yu Z, Orgun MA. 2D medical image segmentation via learning multi-scale contextual dependencies. Methods. 2022;202:40-53.
185. Wang X, Peng Y, Lu L, Lu Z, Bagheri M, Summers RM. ChestX-ray: Hospital-Scale Chest X-ray Database and Benchmarks on Weakly Supervised Classification and Localization of Common Thorax Diseases. In: Lu L, Wang X, Carneiro G, Yang L, editors. Deep Learning and Convolutional Neural Networks for Medical Imaging and Clinical Informatics. Cham: Springer International Publishing; 2019. p. 369-92.
186. Pu Q, Xi Z, Yin S, Zhao Z, Zhao L. Advantages of transformer and its application for medical image segmentation: a survey. BioMedical Engineering OnLine. 2024;23(1):14.

187. Müller D, Soto-Rey I, Kramer F. Towards a guideline for evaluation metrics in medical image segmentation. *BMC Research Notes*. 2022;15(1):210.
188. Salmanpour MR, Alizadeh M, Mousavi G, Sadeghi S, Amiri S, Oveisi M, et al., Machine Learning Evaluation Metric Discrepancies across Programming Languages and Their Components: Need for Standardization. *arXiv preprint*. 2024;241112032.
189. Liu Z, Mhlanga JC, Xia H, Siegel BA, Jha AK. Need for Objective Task-Based Evaluation of Image Segmentation Algorithms for Quantitative PET: A Study with ACRIN 6668/RTOG 0235 Multicenter Clinical Trial Data. *J Nucl Med*. 2024;65(3):485–92.
190. Voulodimos A, Doulamis N, Doulamis A, Protopapadakis E. Deep learning for computer vision: A brief review. *Computational intelligence and neuroscience*. 2018;7068349.
191. Ronneberger O, Fischer P, Brox T, editors. U-Net: Convolutional Networks for Biomedical Image Segmentation. *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*; 2015 2015//; Cham: Springer International Publishing.
192. Wang G, Li W, Zuluaga MA, Pratt R, Patel PA, Aertsen M, et al. Interactive Medical Image Segmentation Using Deep Learning With Image-Specific Fine Tuning. *IEEE Transactions on Medical Imaging*. 2018;37(7):1562–73.
193. Lee G, Bacon S, Bush I, Fortunato L, Gavaghan D, Lestang T, et al. Barely sufficient practices in scientific computing. *Patterns*. 2021;2(2):100206.
194. Brito JJ, Li J, Moore JH, Greene CS, Nogoy NA, Garmire LX, et al. Recommendations to enhance rigor and reproducibility in biomedical research. *GigaScience*. 2020;9(6):giaa056.
195. Goodman SN, Fanelli D, Ioannidis JPA. What does research reproducibility mean? *Science Translational Medicine*. 2016;8(341):341ps12–ps12.
196. Haibe-Kains B, Adam GA, Hosny A, Khodakarami F, Waldron L, Wang B, et al. Transparency and reproducibility in artificial intelligence. *Nature*. 2020; 586(7829):E14–E16.
197. Kapoor S, Narayanan A. Leakage and the reproducibility crisis in machine-learning-based science. *Patterns*. 2023;4(9):100804.
198. Zech JR, Badgeley MA, Liu M, Costa AB, Titano JJ, Oermann EK. Variable generalization performance of a deep learning model to detect pneumonia in chest radiographs: A cross-sectional study. *PLOS Medicine*. 2018;15(11):e1002683.
199. Ball P. Is AI leading to a reproducibility crisis in science? *Nature*. 2023;624(7990):22–5.
200. Colliot O, Thibeau-Sutre E, Burgos N. Reproducibility in Machine Learning for Medical Imaging. *Machine Learning for Brain Disorders*. New York, NY: Springer US; 2023. p. 631–53.
201. Renard F, Guedria S, Palma ND, Vuillerme N. Variability and reproducibility in deep learning for medical image segmentation. *Scientific Reports*. 2020;10(1):13724.
202. Baker M. 1,500 scientists lift the lid on reproducibility. *Nature*. 2016;533(7604):452–4.
203. Raff E. A step toward quantifying independently reproducible machine learning research. *Proceedings of the 33rd International Conference on Neural Information Processing Systems*. 2019;492.
204. Maassen E, van Assen MALM, Nuijten MB, Olsson-Collentine A, Wicherts JM. Reproducibility of individual effect sizes in meta-analyses in psychology. *PloS one*. 2020;15(5):e0233107.
205. Artner R, Verliefde, T., Steegen, S., Gomes, S., Traets, F., Tuerlinckx, F., & Vanpaemel, W. . The reproducibility of statistical results in psychological research: An investigation using unpublished raw data. . *Psychological Methods*. 2021;26(5):527–46.
206. Hardwicke TE, Bohn M, MacDonald K, Hembacher E, Nuijten MB, Peloquin BN, et al. Analytic reproducibility in articles receiving open data badges at the journal *Psychological Science*: an observational study. *Royal Society Open Science*. 2021;8(1):201494.
207. Gundersen OE, Shamsaliei S, Isdahl RJ. Do machine learning platforms provide out-of-the-box reproducibility? *Future Generation Computer Systems*. 2022;126:34–47.
208. Yousefirizi F, Klyuzhin I, Girum K, Uribe C, Buvat I, Rahmim A. Federated testing of AI techniques: Towards sharing of implementations, not just code. *Journal of Nuclear Medicine*. 2023;64(supplement 1):1482.
209. Harris JK, Johnson KJ, Carothers BJ. Use of reproducible research practices in public health: A survey of public health analysts. *PloS one*. 2018;13(9):e0202447.
210. N. Heffer .Addressing the Reproducibility Crisis, One Step at a Time. 2020; <https://www.technologynetworks.com/informatics/articles/addressing-the-reproducibility-crisis-one-ste>.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.