Deanship of Graduate Studies
Al-Quds University

Cross-Modality Transfer Learning for Reliable Lung Cancer Nodule Classification in Low-Dose CT

Sara Rasheed Saleh Asfour

M.Sc. Thesis

Jerusalem – Palestine

1447/2026

Cross-Modality Transfer Learning for Reliable Lung Cancer Nodule Classification in Low-Dose CT

Prepared By:

Sara Rasheed Saleh Asfour

BSc Medical Imaging -Al-Quds University – Palestine

Supervisor: Prof. Radwan Qasrawi

Co-Supervisor: Dr. Hussein ALMasri

This Thesis submitted in partial fulfillment of requirements for the degree of Master of Medical Imaging Technology. Faculty of Graduate studies-Al-Quds University

1447/2026

Al-Quds University
Deanship of Graduate Studies
Medical Imaging Technology

Thesis Approval

Cross-Modality Transfer Learning for Reliable Lung Cancer Nodule Classification in Low-Dose CT

Prepared by: Sara Rasheed Saleh Asfour

Registration No: 22311525

Supervisor: Prof. Radwan Qasrawi

Co-Supervisor: Dr. Hussein ALMasri

The master's thesis was submitted and accepted, Date: 08/01/2026
The names and signatures of the examining committee members are as follows:

1- Head of Committee: Prof. Radwan Qasrawi	Signature:
2- Co-Supervisor: Dr. Hussein ALMasri	Signature:
3- Internal Examiner: Dr. Mohammad Hjouj	Signature:
4- External Examiner: Prof. Derar Eleyan	Signature:

Jerusalem – Palestine

1447/2026

Dedication

I dedicate this work to my family, who stood by me throughout this journey. Your love and support gave me strength when I needed it most.

To my parents, thank you for your endless encouragement and for teaching me the value of hard work and perseverance. Your belief in me made all the difference.

To my teachers, whose knowledge, guidance, and patience shaped not only this research but also my growth as a scholar.

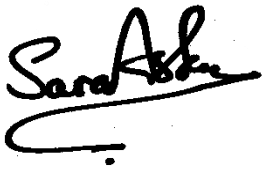
I also dedicate this work to the radiologists, radiologic technologists, and colleagues who generously shared their expertise, time, and support. Your clinical insight and collaboration were essential to the completion of this study.

To my friends, who offered laughter during difficult times and understanding during long absences, your presence made this journey brighter.

This achievement belongs to all of you as much as it belongs to me.

Declaration

I certify that this thesis submitted for the degree of Master is the result of my own research, except where otherwise acknowledged, and that this study (or any part of the same) has not been submitted for a higher degree to any other university or Institution.

Signed: 

Sara Rasheed Saleh Asfour

Date: 08/01/2026

Acknowledgements

I would like to express my sincere gratitude to my supervisor, Dr. Radwan Qasrawi, for his invaluable guidance, continuous support, and constructive feedback throughout the development of this research. His expertise, critical insight, and encouragement were fundamental in shaping the direction, methodology, and scientific rigor of this work.

I am also grateful to Dr. Hussein AlMasri for his valuable suggestions, thought-provoking questions, and constructive critiques, which significantly enriched the quality of this research and broadened my understanding of the subject matter, particularly in refining the methodological framework and interpretation of results.

I would like to express my special appreciation to the two consultant radiologists, Dr. Tawfiq Abu Kishik and Dr. Yazan Dibas, who generously contributed their time and clinical expertise to this study. Their active collaboration, insightful clinical perspectives, and constructive critiques were essential to the study design, data annotation, multi-reader consensus labeling, and clinical validation. Their contributions substantially enhanced the scientific rigor and clinical relevance of this research and deepened my understanding of lung cancer imaging and risk stratification.

I would like to acknowledge the Department of Medical Imaging Technology at Al-Quds University for providing a supportive academic environment, advanced facilities, and computational resources essential to this thesis. I also thank my colleagues at the Al-Quds Business Center for Innovation, Technology & Entrepreneurship, particularly Sulieman Thwib, for their collaboration, insightful discussions, and valuable feedback that strengthened this work.

My sincere appreciation is extended to the radiology departments, clinical teams, and physicians at Al-Makassed Hospital, particularly Arfat Injas, Mofeed Shatat, Mohammad Mshahreh, and Riham Idress, as well as the radiology technicians, for their technical support and cooperation, which were essential to ensuring the clinical relevance and practical grounding of this study. I also extend my appreciation to the teams at Augusta Victoria Hospital, including Yasser Qasem, Sameeh Anati, Yousef and Arwa Salameh, Hussein Bader, Moath Halaheh, Ranin Qumsieh, and Rasha Lahham, for facilitating access to clinical CT datasets and supporting this research.

Special gratitude is extended to Dr. Marwan Qubaja for his significant role in histopathological validation, and to Amira Iriqat and the laboratory technicians for their technical support. Their cooperation was essential to the successful completion of this work.

I also thank colleagues and collaborators who facilitated this research through their cooperation, encouragement, and support, contributing to a productive and supportive research environment.

Finally, I am profoundly grateful to my family and friends for their unconditional love, encouragement, and unwavering support throughout this academic journey. I am especially

thankful to my parents, Rasheed and Sondos Asfour, for instilling in me the values of perseverance, resilience, and the pursuit of knowledge, and to my brothers and sisters for their constant encouragement. To my friends, thank you for being a source of support, motivation, and balance during this journey.

To all those who contributed directly or indirectly to the successful completion of this research, I extend my heartfelt gratitude. Your support, guidance, and generosity were invaluable.

AI Tools Declaration

During the preparation of this work, the author utilized ChatGPT-5 to enhance readability, refine language, and perform grammar checking. All content was subsequently reviewed and edited by the author, who takes full responsibility for the accuracy, integrity, and originality of the publication.

Abstract

Lung cancer is the leading cause of cancer-related mortality worldwide and remains a major health burden in Palestine. According to the Palestine Annual Health Report 2024, lung and bronchus cancers accounted for 316 newly diagnosed cases (10.5 per 100,000 population) and 266 deaths, with 85% occurring among males. These epidemiological patterns highlight the urgent need for locally validated AI solutions that support early lung cancer detection and robust clinical decision-making, particularly in resource-limited healthcare settings where radiological expertise and structured screening programs remain limited.

This study proposes a dose-aware hybrid deep-learning–machine-learning framework for multi-class Lung-RADS classification using heterogeneous chest CT datasets collected from multiple Palestinian institutions. The framework integrates optimized contrast enhancement, deep feature extraction using VGG16, and five classical machine-learning classifiers (LR, SVM, RF, GB, and DT). Contrast enhancement was systematically evaluated using two subsets of 120 malignant cases, one LDCT and one SDCT which is demonstrating that the hybrid CLAHE-USM provided the most balanced improvements (EME = 20.648, PSNR = 19.711, SSIM = 0.912). Dose-optimized parameters, including a clip limit of 3 for LDCT and 4 for SDCT, confirmed the importance of dose-specific preprocessing in stabilizing image quality.

The classification results obtained across the LDCT, combined, external validation, and clinical validation datasets demonstrate that the proposed VGG16-machine learning framework provides robust and consistent performance for Lung-RADS-based nodule risk stratification.

On the LDCT dataset, all classifiers achieved clinically meaningful performance, although variability was observed across Lung-RADS categories. GB and decision tree classifiers demonstrated powerful performance in LR2 and higher-risk categories, achieving accuracies above 0.85. A clear performance improvement was observed with increasing risk of malignancy. For LR4A and LR4B, most classifiers achieved higher accuracy than in lower-risk categories, reflecting the greater structural distinctiveness of high-risk nodules.

The most important performance gains were observed on the combined LDCT-SDCT dataset, particularly for LR3 and LR4A-LR4B. In this setting, SVM and RF achieved near-ceiling performance, with accuracies exceeding 0.97 and AUC values approaching unity. External validation on an independent test set from Al-Makassed Hospital further confirmed the robustness of the proposed framework. For LR2, all major classifiers achieved accuracies above 0.84, whereas for LR4B, all major classifiers achieved accuracies above 0.86.

The most stringent assessment of diagnostic performance was provided by Clinical validation on biopsy-confirmed LR4B cases from Augusta Victoria Hospital. SVM achieved the highest accuracy and AUC, followed closely by LR.

Overall, this framework demonstrates strong potential as a practical decision-support tool for improving Lung-RADS-based risk stratification and supporting early lung cancer detection in resource-constrained healthcare settings.

Keywords: Lung, Pulmonary Nodules, Cancer, Deep Learning, Machine Learning, Computed Tomography, Low-Dose, Standard-Dose.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

Lung cancer represents a significant global public health burden (WHO, 2023) and remains one of the most prevalent and fatal malignancies in Palestine, where it contributes substantially to cancer incidence and mortality and is frequently identified at advanced stages. (The Annual Health Report for 2024, n.d.). This chapter begins by setting the clinical and public health context of the study, outlining the global and national epidemiology of lung cancer, summarizing key risk factors, and reviewing current strategies for screening and early detection, with particular emphasis on low-dose CT. It then discusses the limitations of conventional CT interpretation, such as inter-reader variability, high false-positive rates, and the practical difficulties of consistently applying structured reporting systems like Lung-RADS in everyday practice. Against this backdrop, the chapter introduces the emerging role of artificial intelligence and deep learning in pulmonary nodule detection and risk stratification, leading to a problem statement that is specifically anchored in the Palestinian healthcare setting. Finally, it presents the research justification, objectives, questions, hypotheses, expected contributions, and overall thesis structure, providing a clear roadmap for the development and evaluation of the proposed dose-aware AI framework for Lung-RADS–based nodule classification.

1.1 Overview of Lung Cancer

Lung cancer (LC) represents one of the most serious global health challenges, responsible for substantial illness and death each year. It develops when abnormal lung cells proliferate uncontrollably, forming malignant growths within pulmonary tissue(WHO, 2023). Histologically, lung cancer is classified into two main categories based on the microscopic appearance of tumor cells: Non–Small Cell Lung Cancer (NSCLC), which accounts for approximately 80–85% of cases, and Small Cell Lung Cancer (SCLC), representing about 15–20%(Inamura, 2017; Zheng, 2016). LC remains a major public health concern worldwide due

to its high incidence and mortality rates. According to the International Agency for Research on Cancer (IARC), lung cancer was the leading cause of cancer-related deaths globally in 2022, with nearly 2.5 million new cases and over 1.8 million deaths, representing approximately 12.4% of all newly diagnosed cancers and 18.7% of total cancer deaths(Bray et al., 2024; Ferlay J, 2024).

Lung cancer ranks first among males and second among females in both incidence and mortality(Bray et al., 2024). The estimated number of prevalent cases of lung cancer over 5 years for both sexes in 2022 is 6%(Ferlay J, 2024). Whereas, in most countries 5-year survival rate with little variation according to human development accounts below 20%(Bray et al., 2024).

Lung cancer is one of the most common cancer cases in Palestine. According to the Palestine annual health report dataset records of 2024, lung cancer and bronchus occupy the third place with a total of 316 new cases of lung and bronchus, representing an incidence rate of 10.5 cases per 100,000 population. (The Annual Health Report for 2024, n.d.). as shown in Figure 1.1

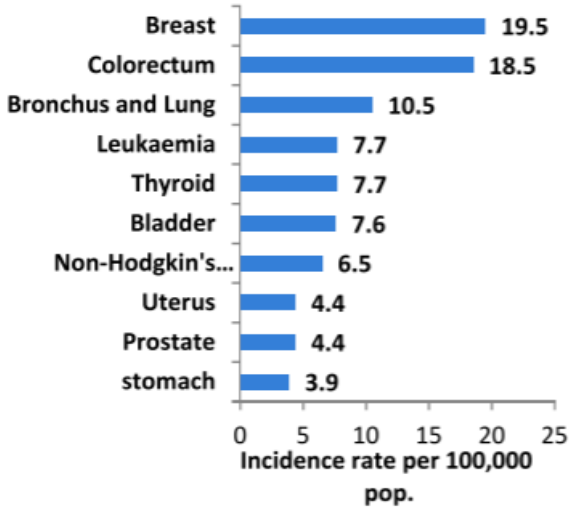


Figure 1.1 Incidence Rate per 100,000 Population for the Top Ten Reported Cancers in the West Bank, Palestine 2024. Adapted from (The Annual Health Report for 2024, n.d.)

In addition, lung cancer ranks second place in males after colorectal cancer with an incidence of 17.5, but in ninth place for the female population with an incidence of 3.3(The Annual Health Report for 2024, n.d.), as described in Figure 1.2 (a&b).

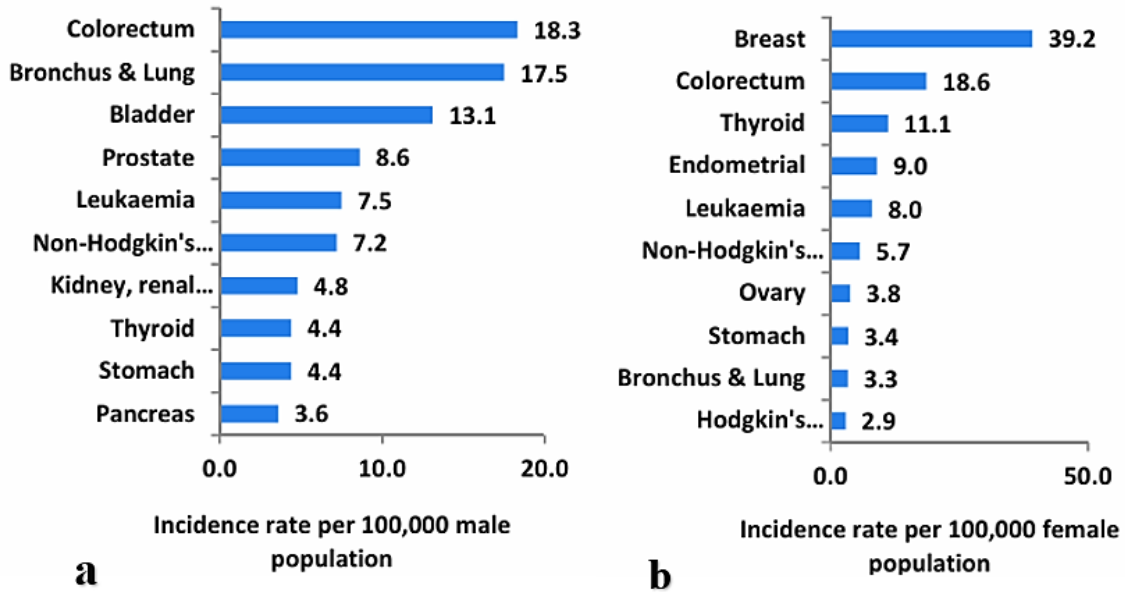


Figure 1.2 Top ten cancer incidence rates among males (a) and females (b) per 100,000. West Bank, Palestine, 2024. Adapted from (The Annual Health Report for 2024, n.d.)

Moreover, the ten most common causes of cancer-related deaths were generally more frequent among males than females, except for breast cancer. For instance, lung cancer represents the primary cause of cancer-related deaths in West Bank- Palestine, responsible for 266 deaths, where 85% of these cases occurred among males(The Annual Health Report for 2024, n.d.), as shown in Figure 1.3. The estimated 5-years survival rate of lung cancer and bronchus from 2019 to 2024 was 24% (The Annual Health Report for 2024, n.d.)

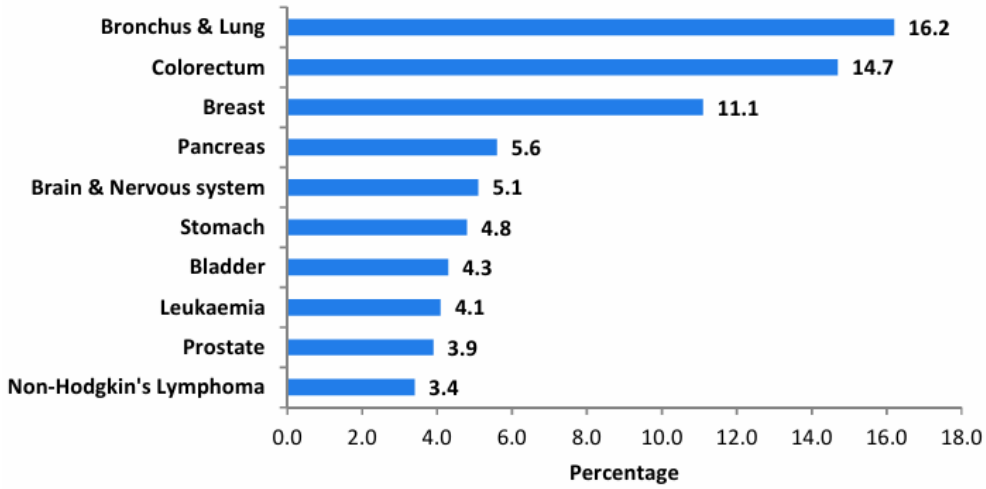


Figure 1.3 Top ten causes of cancer deaths in the West Bank, Palestine (2024). Adapted from (The Annual Health Report for 2024, n.d.)

Several diagnostic methods for lung cancer can be used, including physical examinations, Medical Imaging modalities (such as Chest X-rays and computed tomography scans, Magnetic Resonance Imaging, etc), bronchoscopy to visualize the airways, and histopathologic analysis with tissue biopsy to determine the specific subtype (NSCLC versus SCLC). Moreover, molecular and genetic tests are used to identify genetic mutations and biomarkers that guide the best treatment options and plans(Xiang et al., 2013). These global and national statistics emphasize the significant public health issue of lung cancer and the urgent need to identify the major factors contributing to its expansion.

1.2 Risk Factors, Screenings and Diagnostic Challenges in Lung Cancer

Tobacco smoking is widely recognized as the dominant etiological contributor to lung carcinogenesis, which remains the most well-established cause of lung cancer. In addition to smoking, many other related factors increase susceptibility to lung cancer, including genetic factors, occupational exposures (such as asbestos, silica, and radon), and environmental air pollution. A comprehensive understanding of these risk factors is essential for identifying high-risk populations and implementing effective screening and prevention strategies (Corrales et al., 2020; Malhotra et al., 2016). Notably, tobacco exposure has been directly associated with approximately 85% of lung cancer cases, underscoring the critical importance of targeted screening among high-risk individuals to enable early detection and improve survival outcomes(WHO, 2023).

Treatment options for lung cancer depend on multiple factors, including histological subtype, patient comorbidities, and disease stage at diagnosis. Despite advances in diagnostic techniques and therapeutic approaches, most patients continue to be diagnosed at advanced stages, when curative treatment is often no longer feasible(WHO, 2023). Consequently, the implementation of effective screening programs remains a cornerstone strategy for reducing lung cancer–related mortality(“Reduced Lung-Cancer Mortality with Low-Dose Computed Tomographic Screening,” 2011).

Several screening modalities have been investigated, including sputum cytology, chest radiography, and, most notably, low-dose computed tomography (LDCT), which is superior in detecting early-stage disease(Xiang et al., 2013). Early lung cancer screening tests for high-risk individuals using sputum cytology and chest radiography did not demonstrate a statistically significant reduction in disease-specific mortality(Manser et al., 2013). However, a major turning point occurred with the National Lung Screening Trial (NLST), which reported a 20 % reduction in lung-cancer mortality among high-risk individuals screened with low-dose computed tomography (LDCT) compared with chest X-ray. This reduction was observed among patients aged 55-74 years who had a 30-pack-year history of smoking or were former smokers who had quit within the previous 15 years(“Reduced Lung-Cancer Mortality with Low-Dose Computed Tomographic Screening,” 2011). Based on this evidence, the implementation

of LDCT screening is endorsed by the American College of Radiology (ACR), emphasizing the weight of appropriate patient selection, high-quality low-dose CT acquisition, expert interpretation by qualified radiologists, and structured reporting systems to ensure optimal outcomes(Kazerooni et al., 2016). Detection of pulmonary nodules remains challenging and susceptible to variations among readers, despite management guidelines.

Despite established management guidelines, pulmonary nodule detection remains challenging and is susceptible to inter-reader variability. Recent studies have demonstrated that artificial intelligence (AI)-based computer-aided detection (CAD) systems can significantly enhance radiologists' performance. CAD assistance has been shown to reduce oversight of pulmonary nodules, particularly among less experienced readers, resulting in improved detection of both medium-sized (6–10 mm) and small nodules (3–6 mm). With CAD support, overall sensitivity increased from 68.0% to 85.1%, while average reading time decreased by 11.3%, highlighting both diagnostic and workflow benefits (Kozuka et al., 2020). Another study evaluated radiologists' performance in lung cancer detection, comparing readers from cancer-specialized and non-specialized centers. In this investigation, thirty radiologists independently interpreted 60 chest CT examinations. The overall mean sensitivity among all readers was 0.749. When stratified by institutional background, radiologists from specialized cancer centers demonstrated a slightly higher sensitivity (0.80) compared with those from non-specialized centers (0.719), indicating that experience and training environment may influence diagnostic accuracy(Al Mohammad et al., 2019).

Radiological image interpretation is inherently dependent on human visual assessment and cognitive decision-making processes, which are inherently subject to variability and potential error, affecting the overall accuracy of lung cancer diagnosis. The NLST reported false-positive rates of 21% for LDCT and 9% for chest radiography, with 7% and 4% of false-positive cases, respectively, leading to invasive diagnostic procedures. These false positives may increase patient anxiety, require additional follow-up imaging, and occasionally result in unnecessary invasive testing (Croswell et al., n.d.). Although LDCT is associated with higher false-positive rates due to its superior sensitivity for detecting small pulmonary nodules, this trade-off enables earlier cancer detection and substantial mortality reduction, reinforcing the need for structured reporting systems and advanced decision-support tools to minimize unnecessary interventions.

Conversely, false-negative findings represent another limitation of screening programs, with studies reporting that up to 15% of lung cancers may be missed, and false-negative rates ranging from 8% to 15% (Horeweg et al., 2014; Veronesi et al., 2014). To accelerate the diagnostic process and reduce radiologists' workload, deep learning (DL) methods have emerged as powerful tools in medical image analysis. These approaches not only aid in detecting and characterizing pulmonary nodules but also contribute to analyzing histopathological slides of various cancers. In particular, convolutional neural networks (CNNs) have shown remarkable performance in classifying histopathology images of lung, brain, skin, breast, and colorectal cancers.

For instance, a previous study achieved three-class classification of lung cancer histology, normal tissue, adenocarcinoma, and squamous cell carcinoma, using architectures such as ResNet-50, VGG-19, Inception-ResNet-V2, and DenseNet. These models employed triplet loss functions to maximize inter-class separation and minimize intra-class variation, thereby enhancing feature representation and diagnostic accuracy(Hejbari Zargar et al., n.d.). These findings underscore the transformative potential of deep-learning techniques in lung-cancer imaging and highlight their capacity to address current research gaps.

1.3 Problem Statement

Lung cancer remains the leading cause of cancer-related deaths worldwide, with an incidence of 12.4% (Ferlay J, 2024). According to the Palestinian Health Information Center (PHIC), Annual Health Report 2024, Ministry of Health, bronchus and lung cancer ranks third among the most common cancers in the West Bank, with an incidence rate of 10.5 per 100,000 population. It is also one of the leading causes of cancer mortality, with 266 reported deaths in 2023-85% occurring among males(Palestinian Health Information Center, 2025).

Evidence from major international randomized trials has demonstrated that low dose computed tomography (LDCT) screening enables the detection of up to 70% of lung cancers at early, more treatable stages, where survival rates exceed 65% (de Koning et al., 2020). In organized screening programs, the effectiveness of LDCT is strongly supported by standardized reporting and risk stratification systems, most notably Lung Imaging Reporting and Data System (Lung-RADS)(American College of Radiology, 2022; Christensen et al., 2024a), which was created to reduce false-positive findings, standardize follow-up recommendations, and enhance clinical decision-making. The lack of a structured screening framework highlights the need for advanced, reproducible tools that can assist with LDCT interpretation and risk classification, especially in resource-limited healthcare settings.

Accurate interpretation of chest CT scans is challenging due to the subtle appearance of lung nodules, variations in scanner protocols, radiologist fatigue, and differences in experience. Recent studies have shown that artificial intelligence (AI) systems can improve diagnostic accuracy, increase sensitivity for small or non-solid nodules, and reduce inter-observer variability(Peters et al., 2024)(D. Liu et al., 2024)(Y. Zhang et al., 2021). However, most AI models have been developed on international datasets, with limited application to Palestinian clinical imaging.

This gap highlights the urgent need for AI-based diagnostic solutions tailored to local imaging conditions and population characteristics. Therefore, this study aims to develop and validate a hybrid deep learning–machine learning model for accurate lung nodule detection and classification using CT images from Palestinian healthcare facilities, with the goal of improving early diagnosis and supporting clinical decision-making.

1.4 Research Justifications

Lung cancer remains the leading cause of cancer-related mortality worldwide, and survival is strongly dependent on early detection. Although low-dose CT (LDCT) screening has demonstrated significant reductions in mortality, its effectiveness is limited by high false-positive rates, the presence of subtle early-stage nodules, and variability in radiologists' interpretations. Detecting small pulmonary nodules is particularly challenging because visibility and accurate characterization depend on nodule size, morphology, associated parenchymal abnormalities, underlying disease, and technical CT acquisition parameters. These challenges highlight the need for advanced computational tools that enhance diagnostic accuracy, improve consistency, and support radiologists in distinguishing between benign and malignant nodules.(Peters et al., 2024; "Reduced Lung-Cancer Mortality with Low-Dose Computed Tomographic Screening," 2011)

Artificial intelligence (AI) has emerged as a promising solution to these limitations. Recent studies show that AI-assisted interpretation significantly increases sensitivity for detecting small pulmonary nodules (3–10 mm), reduces reading time, and decreases the likelihood of overlooked abnormalities(Kozuka et al., 2020). Transfer learning and pretrained convolutional neural networks further improve diagnostic performance by enabling robust feature extraction from limited annotated datasets. These approaches also address key sources of, variability, including imaging noise, protocol differences, and heterogeneous nodule morphology, while contributing to reductions in false-positive rates, one of the major limitations of LDCT screening programs.

Despite the development of standardized reporting frameworks such as the Fleischner guidelines(MacMahon et al., 2017a) and the ACR Lung-RADS system(American College of Radiology, 2022; Christensen et al., 2024a), their real-world application remains complex. Radiologists must evaluate hundreds of CT slices per examination within a highly heterogeneous lung parenchyma and apply increasingly detailed criteria for nodule characterization, contributing to interpretive variability, diagnostic uncertainty, and inconsistent management recommendations. AI-driven systems therefore hold considerable potential to standardize assessment, improve reproducibility, and integrate imaging features directly with established reporting guidelines.

This need is especially critical in Palestine, where lung cancer constitutes a significant health burden, yet no national LDCT screening program currently exists. Furthermore, no AI-based systems for pulmonary nodule classification have been developed or validated using Palestinian clinical CT datasets, limiting the applicability of models trained on external populations. This gap underscores the importance of developing a reliable, population-specific AI framework

capable of supporting radiologists and improving early detection within the local healthcare system.

Given these clinical, technical, and regional gaps, this research is justified in its aim to develop and validate an integrated AI framework that enhances image quality, extracts discriminative features, and performs accurate Lung-RADS–based nodule classification. Such a system has the potential to improve diagnostic consistency, reduce false positives, and support earlier detection of lung cancer in Palestinian healthcare settings.

1.5 Research Objectives

The overarching aim of this study is to design, develop, and evaluate hybrid deep learning–machine learning models for accurate lung nodule classification using CT images. The intended outcome is to enhance diagnostic accuracy, reduce radiologist interpretation variability, and support early lung cancer detection within the Palestinian healthcare system.

Specific Research Objectives

1. To develop and curate a cross-modality chest CT dataset from Palestinian healthcare facilities, ensuring the identification of robust, dose-invariant radiological features and a diverse representation of normal lung parenchyma, benign nodules, and biopsy-confirmed malignant lesions.
2. To investigate and implement optimal preprocessing and enhancement techniques, and to develop a Hybrid VGG16 + ML Lung-RADS classification pipeline capable of achieving high sensitivity and specificity while maintaining computational efficiency suitable for resource-constrained clinical environments.
3. To develop and evaluate a hybrid classification framework in which deep features extracted from a pre-trained VGG16 backbone are classified using traditional machine-learning algorithms, with performance assessed through quantitative metrics and clinical validation on biopsy-confirmed cases.

1.6 Research Questions

This research aims to address the following key questions related to developing and evaluating a hybrid lung-nodule classification framework within the Palestinian healthcare context:

1. How effectively can a hybrid VGG16-based deep learning and machine learning (DL+ML) Lung-RADS classification pipeline identify, classify, and monitor suspicious

lung nodules with diagnostic performance comparable to expert radiologists while maintaining real-time or near–real-time processing efficiency?

2. Which preprocessing, enhancement, and optimization techniques most significantly improve the performance of the automated classification system when applied to a multimodal chest CT dataset combining publicly available low-dose CT images and standard-dose clinical CT scans obtained from Palestinian healthcare facilities?
3. To what extent can the proposed automated chest CT interpretation framework provide reliable and consistent decision support for radiologists working in Palestine, particularly across diverse imaging conditions, scanner types, and acquisition protocols?

1.7 Research Hypotheses

Based on the identified research questions and objectives, this study proposes the following hypotheses.

1. Advanced preprocessing, particularly the combined CLAHE and unsharp masking approach, will significantly improve lung nodule classification accuracy compared with standard preprocessing methods.

2. Hybrid models integrating VGG16 feature extraction with classical machine learning classifiers will outperform standalone deep learning or traditional machine learning approaches in lung nodule classification.

3. The proposed hybrid AI framework will demonstrate strong generalization across multimodal chest CT datasets, including both public LDCT scans and standard-dose CT images from Palestinian healthcare facilities.

1.8 Study Contributions

This study presents an integrated AI framework for lung nodule classification that combines optimized image enhancement, VGG16-based deep feature extraction, and classical machine learning classifiers aligned with Lung-RADS v2022. It represents the first effort to develop and validate a Lung-RADS–based AI system using biopsy-confirmed CT data from Palestinian hospitals, addressing a critical gap in regional clinical AI research. By applying CLAHE and unsharp masking, the framework improves the visibility of small, low-contrast pulmonary nodules, enabling more reliable feature extraction and classification. The study further conducts a systematic evaluation of VGG16 deep feature embeddings in combination with logistic regression, random forest, gradient boosting, and decision tree classifiers to identify the most

effective configuration for Lung-RADS classification. Collectively, the proposed approach demonstrates strong potential as a clinical decision-support tool, with the capacity to reduce diagnostic variability and assist radiologists in early lung cancer detection, particularly in healthcare settings without established low-dose CT screening programs.

1.9 Thesis Structure

This thesis comprises six chapters that systematically address the research questions and present the study's findings.

Chapter One establishes the research foundation by introducing lung cancer detection challenges in Palestine. It outlines the problem statement, research questions, objectives, and hypotheses, thereby defining the scope and motivation of the study.

Chapter Two provides essential technical and clinical background, including the histological types and imaging characteristics of lung cancer, radiological features of pulmonary nodules, clinical screening and diagnostic procedures, reporting systems, and the role of medical imaging in lung cancer detection. This chapter serves as a knowledge bridge between the clinical problem context and the proposed computational solutions.

Chapter Three presents a comprehensive review and critical analysis of relevant literature, covering image preprocessing in lung CT imaging, deep learning and computer-aided processing methods for lung cancer analysis, transfer learning strategies, and related AI-based frameworks. This chapter identifies key research gaps and positions the present study within the broader scientific context.

Chapter Four describes the proposed methodology in detail, including data curation, preprocessing pipelines, feature extraction, machine-learning classification, and evaluation strategies. Ethical considerations and implementation details are also addressed to ensure methodological rigor and reproducibility.

Chapter Five presents the experimental results, structured around contrast enhancement performance, classification outcomes, different validation tests, and longitudinal case analyses. Quantitative metrics, comparative evaluations, and statistical results are reported to objectively assess the proposed framework.

Chapter Six discusses the findings in relation to the original research objectives, examines their clinical and practical implications, analyzes study limitations, and outlines directions for future research. The chapter concludes by summarizing the main contributions of this work toward advancing lung cancer classification in resource-constrained healthcare environments.

Chapter 2

Clinical and Theoretical Background

This chapter outlines the key clinical and technical foundations underlying lung cancer assessment on chest CT. It begins with the epidemiology and CT characteristics of pulmonary nodules, focusing on radiologic features that inform malignancy risk. The role of low-dose CT (LDCT) screening and the structure of Lung-RADS v2022 are then introduced, alongside distinctions between LDCT and standard-dose CT (SDCT) that affect image quality and model performance. The chapter also summarizes essential image-processing and enhancement techniques, as well as core deep learning principles relevant to feature extraction in medical imaging.

Emphasis is placed on approaches that improve robustness across heterogeneous CT acquisition protocols, reduce false positives in low-dose CT screening, and enable practical deployment in resource-limited settings, including those in Palestine.

2.1 Histological Types and Imaging Characteristics of Lung Cancer

A clear understanding of lung-cancer histology provides the essential foundation for interpreting imaging findings and developing advanced diagnostic tools. Lung cancer remains a major global health challenge despite significant advances in medical imaging and diagnostic technology. Early and accurate detection remains essential for enhancing survival rates and informing effective treatment strategies. In general, lung cancer is mainly classified into two primary histological categories based on the microscopic characteristics of tumor cells: Non-Small Cell Lung Cancer (NSCLC) and Small Cell Lung Cancer (SCLC). NSCLC accounts for approximately 80–85 % of all cases and includes three fundamental subtypes, adenocarcinoma (55%) with two main types of adenocarcinomas are in situ and invasive lung

adenocarcinoma(Pang et al., 2020), squamous-cell carcinoma (35%), and large-cell carcinoma (10%), each differing in growth pattern, location, and prognosis. In contrast, SCLC is identified by growing rapidly, invading early into other organs and tissues, with a powerful correlation to tobacco smoking, which represents about 15–20 % of cases (Ciofiac et al., 2024).

Lung cancer in the early stage is classified pathologically into 4 types: Firstly, Invasive adenocarcinoma is mostly a wall-like growth, acinar, papillary, and solid; secondly, Adenocarcinoma in situ is a localized small adenocarcinoma with a diameter of the lesion of 3 cm or less; thirdly, Squamous cell carcinoma has common lobulation features; lastly, SCLC is a lung mass growing beside the long axis of the bronchus(Pang et al., 2020). These CT images of the four pathological types of lung cancer are shown in Figure 2.1.

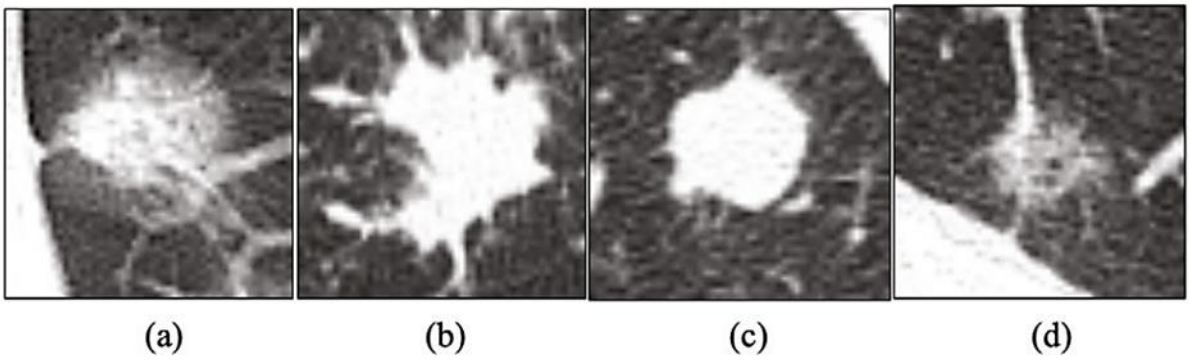


Figure 2.1 CT Imaging Examples of Major Lung Cancer Histological Subtypes.

(a) invasive adenocarcinoma, (b) squamous cell carcinoma; (c) small cell lung cancer; (d) adenocarcinoma in situ. Adapted from (Pang et al., 2020)

Furthermore, a previous study using CT scan images found that SQCC and SCLC subtypes tend to be centrally located, with internal cavitations being more common in SQCC cases. In contrast, ADC tends to be observable peripherally and in a solid texture (Ciofiac et al., 2024). These histological differentiations are clinically significant because they affect imaging appearance, staging, treatment planning, and survival outcomes. The correlation between histological subtype and its CT appearance provides the basis for understanding how tumor biology manifests radiologically.

2.2 Radiological Features of Pulmonary Nodules

Pulmonary nodules may constitute either non-cancerous (benign) or cancerous (malignant) growths. Benign nodules are non-invasive lesions that do not spread into or damage surrounding tissues. They usually grow slowly and are sometimes detected incidentally. In most cases,

benign nodules can be surgically removed, with minimal possibility of recurrence. However, large benign nodules may occasionally compress neighboring organs, such as blood vessels or nerves, leading to clinical symptoms. On the other hand, malignant nodules are composed of cancerous cells that can invade neighboring tissues and spread to distant organs via the lymphatic system or bloodstream. This aggressive process of district invasion and distant spread is referred to as metastasis(Chand Bansal et al., n.d.). Awareness of these fundamental biological differences provides the basis for radiologic assessment and AI-based classification of pulmonary nodules as benign or malignant.

Lung nodules are commonly characterized on chest computed tomography (CT) as small, rounded, or irregular opacities that are well or poorly defined, with diameters up to 3 cm. Lung nodules can be evaluated effectively when CT scans are done, with contiguous thin sections reconstructed (≤1.5 mm in thickness), and include off-axis reconstructions, enabling precise depiction of small lesions(J. G. Nam & Goo, 2022). Furthermore, the optimum evaluation and measurement CT imaging techniques for nodules are achieved using lung window setting and a sharp filter (edge-enhancing reconstruction). Whereas nodules could be rendered partially invisible when viewed using the mediastinal window setting. It is recommended to use similar CT parameters and reconstruction techniques during follow-up examinations, including dose reduction techniques, slice thickness, and the reconstruction filter, to reduce interscan differences(MacMahon et al., 2017b; J. G. Nam & Goo, 2022).

Lung nodules have different classifications according to their appearance on CT scan, specifically their attenuation characteristic (density) and the presence or absence of ground-glass opacity (GGO)(MacMahon et al., 2017b; J. G. Nam & Goo, 2022). Nodules with homogeneous soft-tissue attenuation are classified as solid, whereas those containing any ground-glass component are considered subsolid. As well as a hazy area of increased attenuation in the lung with preserved bronchial and vascular markings, which refers to ground-glass opacity. In addition, Subtypes of Sub-solid nodules are part-solid nodules and pure ground-glass nodules (GGNs). A part-solid nodule consists of a combination of ground-glass and solid attenuation, while a pure GGN only contains a ground-glass component (J. G. Nam & Goo, 2022), which is presented in Figure 2.2.

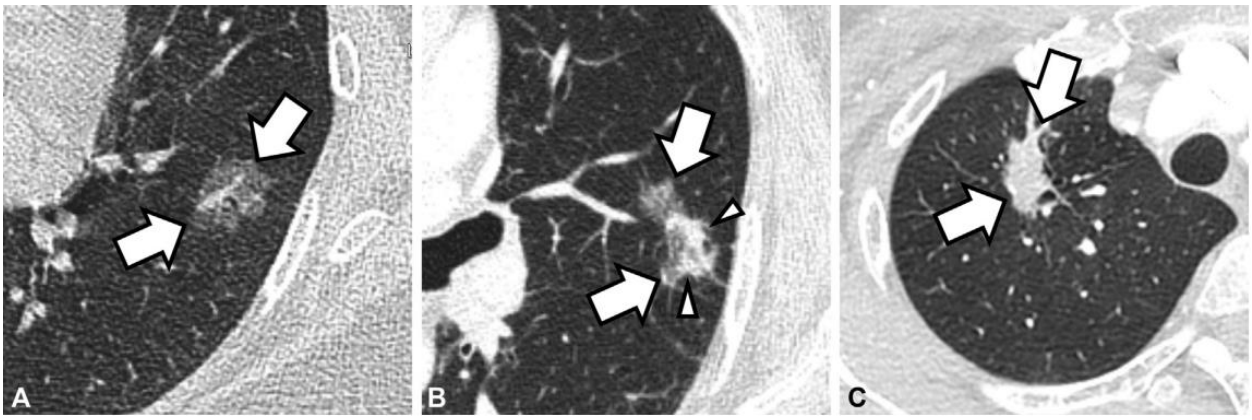


Figure 2.2 Examples of Pulmonary Nodules with varying Morphological appearances.

(A) Pure ground-glass nodule exhibiting hazy increased attenuation without obscuring bronchial or vascular margins. (B) Part-solid nodule containing both ground-glass and solid elements. (C) Solid nodule characterized by uniform soft-tissue attenuation. Adapted from (J. G. Nam & Goo, 2022)

Multiple factors and imaging characteristics have been linked to an increased likelihood of malignancy. Notably, marginal spiculation has long been recognized as one of the strongest radiologic indicators of cancer, with recent studies confirming its strong association with malignant nodules(MacMahon et al., 2017b). Nodule location also contributes to malignancy risk, particularly when lesions are situated in the upper lobes, which show a predilection for primary lung cancer, most often in the right lung(MacMahon et al., 2017b). However, the relationship between nodule multiplicity and malignancy is not linear. Data from both the NELSON trial and the PanCan model demonstrated that patients with one to four nodules exhibit a progressively higher risk of primary lung cancer, whereas those with five or more nodules show a comparatively lower malignancy risk(MacMahon et al., 2017b). This inverse association suggests that multiple nodules are more often related to benign etiologies, such as prior granulomatous infection. Compared with solitary nodules, the presence of multiple nodules is therefore associated with a relative reduction in cancer risk(MacMahon et al., 2017b).

Moreover, growth rate varies widely depending on nodule morphology and histologic subtype. For solid malignant nodules, an increase in diameter of approximately 26% corresponds to a doubling in volume, with most exhibiting volume-doubling times (VDTs) within the 100–400-day range(MacMahon et al., 2017b). In contrast, subsolid cancerous nodules, which often represent primary adenocarcinomas, tend to grow much more slowly, with mean doubling times ranging from 3 to 5 years(MacMahon et al., 2017b). The comprehensive evaluation of pulmonary nodules, based on morphology, attenuation, location, and growth behavior, provides the essential framework for determining the risk of malignancy and guiding patient management. These radiologic parameters are also the fundamental imaging features utilized in

traditional computer aided detection (CAD) systems and deep-learning algorithms, which aim to enhance radiologists' accuracy and diagnostic efficiency.

2.3 Clinical Lung Cancer Screening and LDCT Reporting System

In 1980, the American Cancer Society (ACS) withdrew its recommendation for routine lung cancer screening with chest X-rays in smokers after multiple randomized trials conducted in the 1970s showed no clear mortality benefit(Wolf et al., 2024).

Thirty-three years later, evidence from the National Lung Screening Trial (NLST) demonstrated that three annual rounds of LDCT screening conducted a 20% reduction in lung cancer mortality compared with chest radiography("Reduced Lung-Cancer Mortality with Low-Dose Computed Tomographic Screening," 2011b; Wolf et al., 2024). Based on these findings, the ACS issued a new recommendation supporting annual LDCT screening for adults aged 55–74 years with a smoking history of at least 30 pack-years, who were either current smokers or had quit within the past 15 years, and who did not have life-limiting comorbidities(Wender et al., 2013; Wolf et al., 2024).

The updated 2013 ACS guideline also incorporated emerging evidence on LDCT effectiveness, lung cancer risk among long-term former smokers, the potential benefits and harms of screening beyond age 80, screening considerations for adults with an anticipated life expectancy greater than five years, and revised assessments of radiation-related risks using modern imaging doses.

The most recent ACS recommendations advise annual low-dose CT (LDCT) lung cancer screening for adults aged 50–80 years who currently smoke or formerly smoked and are considered high risk due to a ≥20 pack-year smoking history. Notably, the ACS no longer recommends using the number of years since quitting (YSQ) as a criterion to initiate or discontinue screening in eligible former smokers who meet the age and pack-year threshold.(Wender et al., 2013)

Efforts to establish lung cancer screening strategies date back to the late 1950s and early 1960s, when combinations of sputum cytology and chest radiography (CXR) were tested in high-risk individuals. These approaches, however, did not demonstrate a reduction in lung cancer mortality compared with CXR alone. In 2000, Berlin further noted that the existing trials did not provide convincing evidence that annual CXR failed to reduce mortality, underscoring ongoing uncertainty regarding effective screening modalities.(Berlin, 2000)

Clear evidence emerged decades later with the National Lung Screening Trial (NLST), whose first results were published in 2011. The NLST demonstrated a 20% relative reduction in lung cancer mortality among high-risk individuals screened annually with LDCT compared with those screened using CXR. This landmark finding established LDCT as the superior modality

and directly informed subsequent updates to ACS screening guidelines("Reduced Lung-Cancer Mortality with Low-Dose Computed Tomographic Screening," 2011b).

Annual low-dose CT (LDCT) is recommended for lung cancer screening because early detection substantially improves the likelihood of effective treatment. However, LDCT screening has several important limitations. Not all lung cancers are detected by LDCT, and not all cancers identified at an early stage will prevent mortality. Additionally, LDCT screening is associated with a considerable risk of false-positive findings, often necessitating repeat imaging and, in some cases, invasive diagnostic procedures such as biopsies to distinguish malignant lesions from benign or incidental abnormalities(Wender et al., 2013).

LDCT screening is characterized by a relatively high rate of benign, noncalcified nodule detection. Across published studies, thresholds for defining a positive finding and nodule detection rates per screening round vary widely. Most trials reported only baseline or baseline plus first-round detection rates; the NLST was unique in reporting results across three consecutive screening rounds. In the NLST LDCT arm, 27.3% of participants had a positive screen in the first round, and 39.1% had at least one abnormal LDCT result across all three rounds("Reduced Lung-Cancer Mortality with Low-Dose Computed Tomographic Screening," 2011b; Wender et al., 2013).

According to the American College of Radiology Imaging Network (ACRIN) NLST protocol, individuals with indeterminate or positive findings were recommended to undergo follow-up diagnostic, evaluation, including repeat LDCT, limited thin-section CT, or additional testing, at intervals of 3, 6, 12, or 24 months, depending on nodule size and malignancy suspicion. Many participants with small nodules and low concern for malignancy required only a second LDCT for confirmation; if no interval growth was observed, no further imaging was recommended until the next annual screening round. If interval growth or suspicious morphological changes were noted, earlier repeat imaging (3–6 months) or more definitive diagnostic procedures were pursued. Importantly, most individuals with abnormal LDCT findings required only follow-up imaging to determine whether a detected nodule was enlarging(Wender et al., 2013).

Although LDCT uses a relatively small amount of ionizing radiation, repeated annual screening does entail a minimal lifetime risk of radiation-induced cancer. Nonetheless, current evidence indicates that the benefits of LDCT screening clearly outweigh the very low radiation risk, particularly in high-risk individuals.(Wender et al., 2013)

The ACR–STR Practice Parameter for Lung Cancer Screening CT outlines the technical and reporting standards for low-dose CT (LDCT) in high-risk, asymptomatic adults aged 50–80 years with ≥20 pack-years of smoking. LDCT is the only imaging test proven to reduce lung cancer mortality, and therefore must be performed with optimized low-dose parameters because screening may continue annually for decades. The primary goal of LDCT is early detection of suspicious pulmonary abnormalities, with additional evaluation of incidental thoracic findings following the ACR–SABI–SPR Thoracic CT Practice Parameter(ACR–STR, n.d.).

A standard lung cancer screening CT should be performed using multidetector helical acquisition during a single breath-hold, extending from the lung apices to the costophrenic angles. Axial images must be reconstructed at ≤2.5 mm, with 1.0 mm or thinner slices preferred to optimize nodule detection. Maximum Intensity Projection (MIP) images can further increase sensitivity for small pulmonary nodules, and multiplanar reconstructions (MPRs) are particularly useful for characterizing juxtapleural nodules. All scans should be obtained at full suspended inspiration, covering the entire lung volume, with a patient-specific field of view. Intravenous contrast is not used for screening examinations. Interpretation and reporting of LDCT screening studies must follow the ACR Lung-RADS v2022 classification scheme. Appropriate window and level settings should be applied to evaluate all relevant anatomy within the scan range, including lung parenchyma, mediastinum, pleura, chest wall, bones, lower neck, and upper abdomen.(ACR–STR, n.d.)

Pulmonary nodules and focal lesions should be reported systematically, specifying anatomic location (lobe and segment), series and image number, size, attenuation characteristics (e.g., soft tissue, calcification pattern, fat content), nodule type (solid, ground-glass, or part-solid), and margin features (e.g., smooth, lobulated, spiculated). Comparison with prior imaging is essential for accurate assessment. Direct comparison with previous CT images is strongly recommended rather than relying solely on earlier reports. When evaluating temporal changes, radiologists should review both the oldest available and the most recent prior examinations to identify subtle progression. Volumetric nodule analysis may also be incorporated to improve accuracy in assessing growth.(ACR–STR, n.d.)

The American College of Radiology (ACR) introduced the Lung CT Screening Reporting and Data System (Lung-RADS) version 1.0 in 2014 to standardize CT reporting for lung cancer screening, provide consistent management recommendations, and support systematic outcomes monitoring(Christensen et al., 2024a). The most recent update, Lung-RADS v2022, was released in November 2022, coinciding with National Lung Cancer Awareness Month. This version was developed by the ACR Lung-RADS Committee, which includes 15 experts in diagnostic radiology, thoracic surgery, and pulmonary medicine representing a wide range of clinical practice settings(Christensen et al., 2024a).

Lung-RADS Category: Each screening exam is assigned a Lung-RADS category from 0 to 4, based on the nodule with the highest level of suspicion.
Lung-RADS Management: Follow-up intervals are determined from the date of the exam interpretation. For example, a Lung-RADS 2 finding requires a 12-month screening LDCT from the date of the current exam. Management of category 4A lesions follows a stepped approach, depending on whether stability or reduction in size is demonstrated on follow-up imaging(American College of Radiology, 2022).

In the Practicing Audit Definitions, categories 1 and 2 are considered a negative screen, while categories 3 and 4 are classified as a positive screen. Importantly, a negative result does not rule out the presence of lung cancer(American College of Radiology, 2022).

Accurate nodule measurement is essential for correct classification. Nodule size is determined by calculating the mean diameter, using the long and short axes measured to one decimal place (mm), and reporting the average to one decimal place. These measurements should be performed on thin axial slice axial CT images to reflect the true nodule dimensions. However, multiplanar reconstructions may be used when it's necessary to delineate the true maximal nodule dimension. When available, volumetric measurements should be reported to the nearest whole number in mm^3(American College of Radiology, 2022).

Size thresholds apply both to newly detected nodules and existing nodules that enlarge into a higher size category. If a nodule crosses a new size threshold for another Lung-RADS category, regardless of whether it meets the formal definition of growth, it must be reclassified accordingly. A nodule is considered to have grown if its mean diameter increases by >1.5 mm (or >2 mm^3 in volume) within a 12-month interval(American College of Radiology, 2022).

Lung-RADS distinguishes two types of slow-growing nodules(American College of Radiology, 2022):

- Slow-growing non-solid (ground-glass) nodules:

A ground-glass nodule that shows gradual enlargement across multiple screening exams but does not exceed the >1.5 mm growth threshold in any 12-month period may remain classified as Lung-RADS 2 until it meets criteria for another category, such as developing a solid component, in which case it should be managed according to the part-solid nodule criteria(American College of Radiology, 2022).

- Slow-growing solid or part-solid nodules:

A solid or part-solid nodule showing growth over several exams, where the increase exceeds 1.5 mm in any 12-month interval, is considered suspicious and may be categorized as Lung-RADS 4B. These lesions may not demonstrate increased metabolic activity on PET/CT; therefore, biopsy or surgical evaluation is often the most appropriate management strategy.

Finally, when prior imaging is unavailable, a temporary Lung-RADS 0 may be assigned until comparison studies become available, after which a definitive Lung-RADS category can be applied(American College of Radiology, 2022).

- Juxtapleural Nodules

Screen-detected peri-fissural nodules that meet specific size and morphological, criteria, namely solid density, a triangular, ovoid, or lentiform shape, and a maximum diameter <10 mm, are generally regarded as benign. These lesions most commonly represent intrapulmonary lymph nodes, and multiple studies have reported a 0% malignancy rate for nodules fulfilling these characteristics(Christensen et al., 2024a).

Recent evidence shows that up to 32% of solid nodules measuring 6–10 mm in mean diameter is juxta-pleural (including peri-fissural, costal pleural, peri-mediastinal, and peri-diaphragmatic locations). Importantly, none of these nodules were malignant when evaluated using the same size and morphological criteria established for peri-fissural nodules(Christensen et al., 2024a).

Expanding the peri-fissural nodule classification framework to encompass all juxta-pleural nodules, regardless of their pleural location, has been shown to reduce false-positive rates and improve Lung-RADS specificity without compromising diagnostic sensitivity. These Juxtapleural Pulmonary Nodules described clearly in Figure 2.3 (A, B, C) (Christensen et al., 2024a).

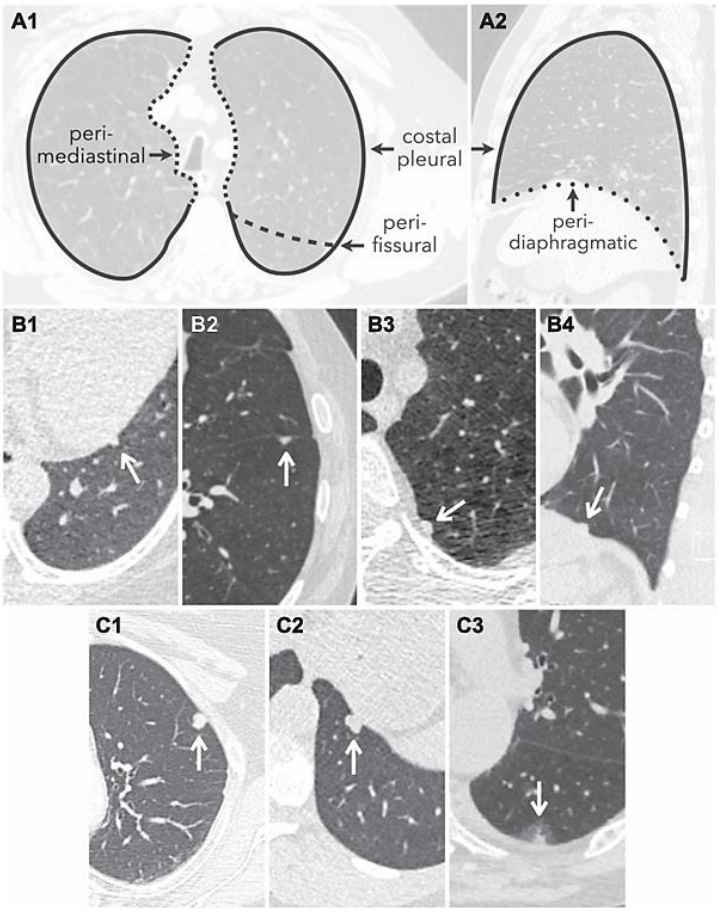


Figure 2.3 Juxtapleural Pulmonary Nodules**.** Adapted from (Christensen et al., 2024a)

(A) Schematic illustration of the anatomical distribution of juxtapleural nodules, including perimediastinal, perifissural, costal pleural, and peridiaphragmatic locations, shown in axial (A1) and sagittal (A2) planes. (B) Juxtapleural nodules with a mean diameter < 10 mm that are solid, have smooth margins, and exhibit oval, lentiform, or triangular morphology (arrows) are considered benign and classified as Lung CT Screening Reporting and Data System (Lung-RADS) category 2. Representative benign examples include perimediastinal (B1), perifissural (B2), costal pleural (B3), and peridiaphragmatic (B4) nodules. (C) Juxtapleural nodules that do not meet Lung-RADS 2 criteria (arrows) are classified according to size and composition, including a 9-mm solid round costal pleural nodule (Lung-RADS 4A; C1), an 11-mm solid lobulated peridiaphragmatic nodule (Lung-RADS 4A; C2), and a 15-mm part-solid costal pleural nodule with a 6-mm solid component (Lung-RADS 4A; C3). Adapted from (Christensen et al., 2024a)

Lung-RADS v2022 provides detailed recommendations for managing pulmonary and airway findings that may resemble malignant nodules but frequently represent benign infectious, inflammatory, or obstructive processes. Accurate classification is essential to prevent unnecessary diagnostic procedures while preserving the high specificity and sensitivity of lung cancer screening.

1. Suspected Infectious or Inflammatory Pulmonary Findings

Some CT abnormalities detected during lung cancer screening may reflect an indeterminate infectious or inflammatory process rather than malignancy. In these situations, short-term follow-up is recommended to allow potential resolution and to exclude an underlying suspicious lesion.

1.1. Lung-RADS 0: Indeterminate Findings Requiring 1–3 Month Follow-up

A temporary Lung-RADS 0 category with repeat LDCT in 1–3 months may be assigned when the following abnormalities are present:

- Segmental or lobar consolidation
- Multiple newly appearing nodules (>6)
- Large solid nodules (≥8 mm) developing over a short interval
- New nodules in high-risk clinical contexts (e.g., immunocompromised patients)
- Any finding that obscures lung parenchyma and may conceal a malignant nodule

The follow-up examination must be reclassified based on the most suspicious finding.

1.2. Findings More Consistent with Malignancy

New solid or part-solid nodules whose morphology favors malignancy over infection, and that meet Lung-RADS 4B size criteria, should be directly categorized as Lung-RADS 4B, prompting diagnostic imaging or clinical evaluation.

1.3. Infectious Findings That Do Not Require Short-term Follow-up

Some characteristic infectious patterns do not require 1–3-month reevaluation, including:

- Tree-in-bud nodularity
- New ground-glass nodules < 3 cm

These findings should be assessed using standard Lung-RADS size and composition criteria, most commonly corresponding to Lung-RADS 2 with 12-month LDCT. These Findings described in Figure 2.4 (A, B, C)

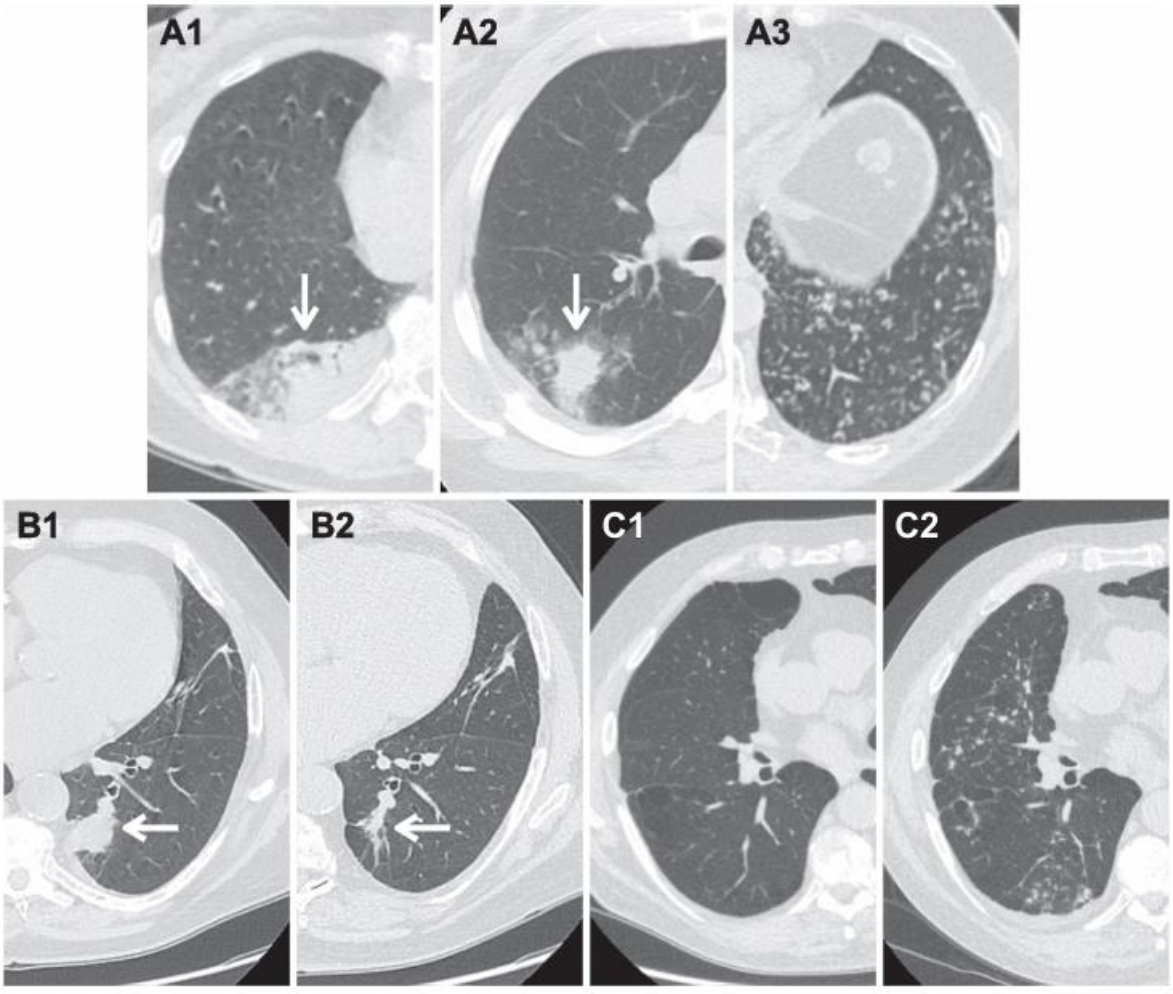


Figure 2.4 Infectious or Inflammatory Findings. Adapted from (Christensen et al., 2024a)

(A) Lung cancer screening (LCS) findings suggestive of an infectious or inflammatory process include new segmental or lobar consolidation (A1, arrow), the rapid appearance of large new

nodules (≥ 8 mm) over a short interval (A2, arrow), or the presence of multiple new nodules (A3). Such findings are typically classified as Lung CT Screening Reporting and Data System (Lung-RADS) category 0 when clinical suspicion for infection or inflammation is high, with a recommendation for short-interval follow-up low-dose CT (LDCT) at 1–3 months to confirm resolution. Subsequent classification is based on the most concerning residual finding at follow-up.

(B) A 55-year-old patient with a baseline LCS CT previously classified as Lung-RADS 3 (not shown) underwent a 6-month follow-up LDCT, which demonstrated a new lobulated soft-tissue nodule measuring 28 mm in the left lower lobe (B1, arrow). Given the lesion's size and its absence on prior imaging, the examination was classified as Lung-RADS 0, favoring an infectious etiology. Follow-up LDCT at 3 months showed a marked reduction in nodule size without complete resolution (B2, arrow), with no additional new findings and stability of other nodules. The study was therefore reclassified as Lung-RADS 2, with a recommendation for routine annual LCS CT. Had the nodule remained stable, increased in size, or shown minimal regression, an infectious or inflammatory cause would be less likely, and the examination would be upgraded to Lung-RADS 4B with a recommendation for diagnostic evaluation.

(C) A 61-year-old patient undergoing annual LCS CT was found to have multiple new bilateral tree-in-bud nodules with basilar predominance (C2) compared with the prior annual examination (C1). Although an infectious process was considered most likely, short-interval follow-up LDCT was deemed unnecessary. Instead, the examination was classified based on the size and composition of the nodules (solid, < 4 mm), corresponding to Lung-RADS category 2.

2. Airway Nodules

Airway abnormalities can reflect secretions, mucous plugging, inflammation, or less, commonly, airway neoplasms. Lung-RADS v2022 classifies these findings based on location, morphology, number, and persistence(Christensen et al., 2024a).

2.1. Location

- Segmental or more proximal airway nodules (>3 mm):

Classified as Lung-RADS 4A, with a 3-month LDCT to evaluate for resolution.

- Subsegmental or distal airway opacities:

Typically represent benign mucous plugging or infection and may be classified as Lung-RADS 2 or Lung-RADS 0, depending on the degree of diagnostic uncertainty.

2.2. Morphology

- Features that favor airway secretions include:

- Tubular or complex configuration
- Internal air
- Absence of soft-tissue density
- Attenuation < 21.7 HU

These findings are generally classified as Lung-RADS 2 with 12-month LDCT follow-up.

Airway nodules without these benign features may be assigned Lung-RADS 4A with repeat imaging at 3 months.

2.3. Number of Lesions

- Single airway nodules are more suspicious for malignancy.
- Multiple airway opacities (tree-in-bud, multifocal mucous impaction) typically reflect benign infectious or inflammatory etiologies and may not necessitate a 4A designation.

2.4. Persistence on Follow-up

Persistence is the most decisive feature:

- Segmental or proximal airway nodules that persist at 3-month follow-up must be upgraded to Lung-RADS 4B.
- Further evaluation includes PET/CT (if solid component ≥8 mm) or clinical assessment with bronchoscopy.

Unlike pulmonary nodules, persistent airway nodules must not be downgraded despite stability.

2.5. National Comprehensive Cancer Network (NCCN) vs. Lung-RADS Timing Considerations

NCCN guidelines recommend a 1-month follow-up CT for airway nodules, whereas Lung-RADS v2022 recommends 3 months. Both systems advise bronchoscopy for persistent lesions. Due to limited comparative data, neither approach has demonstrated superiority. That is demonstrated in Figure 2.5 (A, B, C) (Christensen et al., 2024a)

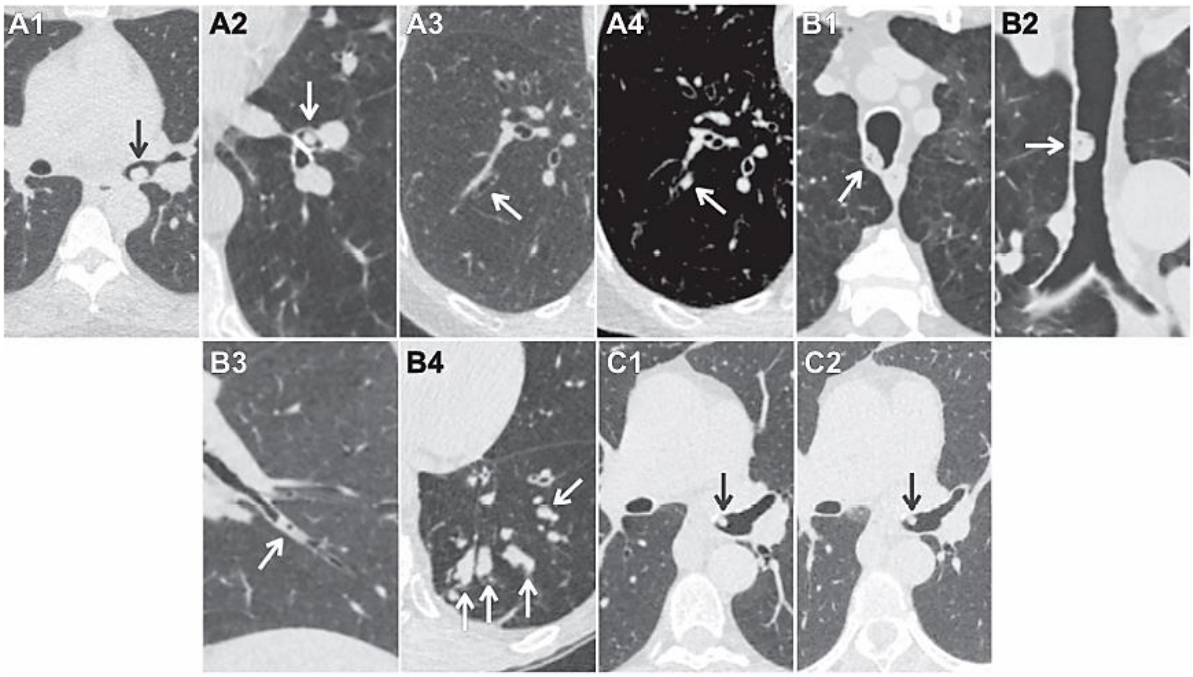


Figure 2.5 Morphological Variants of Airway Nodules. Adapted from (Christensen et al., 2024a)

(A) In Lung CT Screening Reporting and Data System (Lung-RADS) version 2022, airway nodules are assessed based on their location, morphology, number, and persistence. Newly detected airway nodules located at the segmental or more proximal bronchi that lack benign features are classified as Lung-RADS category 4A, with a recommendation for short-interval follow-up low-dose CT (LDCT) at 3 months to assess for resolution (A1: left mainstem bronchus nodule [arrow]; A2: left lower lobe segmental bronchus nodule [arrow]). In contrast, subsegmental airway nodules are classified as Lung-RADS category 2 (A3: baseline lung cancer screening [LCS] CT demonstrating a patent subsegmental airway [arrow]; A4: annual LCS CT showing a new subsegmental airway nodule [arrow]).

(B) The presence of air within an airway nodule, particularly in the absence of a soft-tissue component and with a mean attenuation of less than 21 Hounsfield units, favors secretions rather than true nodules (B1–B3, arrows). Multiple tubular airway opacities are suggestive of mucus plugging (B4, arrows). These findings are typically classified as Lung-RADS category 2.

(C) A 70-year-old patient presenting for annual LCS CT demonstrates a soft-tissue nodule in the left mainstem bronchus (C1, arrow), classified as Lung-RADS 4A. Follow-up LDCT at 3 months shows persistence of the lesion (C2, arrow). Airway nodules that persist on short-interval follow-up remain suspicious and are upgraded to Lung-RADS category 4B, prompting diagnostic evaluation, most commonly with referral for clinical assessment and bronchoscopy. Adapted from (Christensen et al., 2024a)

In summary, Lung-RADS standardizes the interpretation of LDCT screening examinations by assigning each study a category from 0 to 4, based on the nodule with the highest level of suspicion. Categories 1 and 2 represent negative or benign findings and typically require routine 12-month LDCT follow-up. Category 3 includes probably benign nodules and warrants a 6-month LDCT. Categories 4A and 4B reflect increasing levels of suspicion for malignancy, requiring 3-month LDCT, PET/CT, or further diagnostic evaluation, depending on nodule size and composition. Category 0 indicates an incomplete evaluation, necessitating comparison with prior imaging or a short-interval follow-up (1–3 months). These categories streamline management decisions, reduce unnecessary interventions, and enhance the accuracy and consistency of lung cancer screening interpretations(American College of Radiology, 2022).

2.4 Clinical Thoracic Diagnostic CT Protocols

Thoracic computed tomography (CT) plays a central role in the evaluation of pulmonary nodules, interstitial lung disease, and thoracic malignancies. Clinical diagnostic protocols vary according to the specific clinical indication, required spatial resolution, and radiation dose considerations. The following subsections describe the technical characteristics and clinical applications of conventional thoracic CT and high-resolution CT (HRCT) protocols.

2.4.1 Thoracic Computed Tomography (CT)

The ACR–SABI–SPR–STR Practice Parameter for Thoracic CT provides comprehensive standards for the performance of high-quality thoracic CT imaging. CT is the primary modality for evaluating thoracic pathology and requires familiarity with thoracic anatomy, pathophysiology, CT acquisition techniques, and radiation safety. The goal of thoracic CT is to accurately demonstrate normal and abnormal intrathoracic structures to support clinical diagnosis and management.(ACR-SABI-SPR-STR, n.d.)

Thoracic CT may serve as a primary diagnostic examination or complement chest radiography. Its clinical indications are broad and include evaluation of abnormalities detected on previous imaging, lung cancer screening, staging and follow-up of thoracic malignancies, assessment of metastatic disease, characterization of cardiothoracic manifestations of systemic disorders, evaluation of congenital and acquired cardiovascular abnormalities, investigation of suspected pulmonary embolism or pulmonary hypertension, assessment of parenchymal and airway disease, trauma evaluation, CT-guided interventions, and assessment of chest wall, pleural, and mediastinal pathology. Thoracic CT is also essential for surgical and radiotherapy planning, monitoring response to oncologic therapies, evaluating postoperative complications, and assessing complex infections, particularly in immunocompromised patients. Appropriate documentation of signs, symptoms, and relevant clinical history is necessary for appropriate protocol selection.(ACR-SABI-SPR-STR, n.d.)

A routine thoracic CT examination extends from the lung apices to the posterior costophrenic sulci. Standard reconstructions use 3–5 mm axial slices with soft-tissue and high-spatial-frequency kernels, while thin-section imaging (≤1.2 mm) is preferred when volumetric reconstructions or detailed parenchymal assessment are required. Coronal and sagittal reformations at similar slice thickness enhance anatomic evaluation. Maximum intensity projection (MIP) slabs (5–10 mm) support pulmonary nodule detection, whereas minimum intensity projection (MinIP) images (<10 mm) assist in assessing emphysema and air trapping. Most scans are obtained during suspended full inspiration. For low-dose CT examinations, such as lung cancer screening or nodule surveillance, the scan range should be restricted to the lungs. Field-of-view reduction supports optimal spatial resolution, particularly for high-resolution imaging. Use of intravenous contrast depends on the clinical indication. Reporting should follow the ACR Practice Parameter for Communication of Diagnostic Imaging Findings.(ACR, n.d.; ACR-SABI-SPR-STR, n.d.)

2.4.2 High-Resolution Computed Tomography (HRCT)

The ACR–SPR–STR Practice Parameter for HRCT of the Lung defines standards for the performance and clinical use of HRCT in evaluating diffuse parenchymal lung disease. HRCT is primarily used for known or suspected interstitial lung disease (ILD). Its main objective is to detect, characterize, and quantify parenchymal and small-airway abnormalities using standardized nomenclature.

HRCT relies on thin-section imaging (≤1.5 mm, preferably ≤1.0 mm) reconstructed with a high-spatial-frequency kernel to optimize visualization of fine pulmonary detail. With modern multidetector CT (MDCT), HRCT employs near-isotropic volumetric data acquired in a single breath-hold, improving spatial resolution and diagnostic accuracy. Indications for HRCT include evaluation of diffuse lung disease (obstructive, restrictive, or interstitial), assessment of small airway disease, estimation of fibrosis extent, guidance for biopsy site selection, and evaluation of occupational lung diseases such as pneumoconiosis. Reporting should adhere to the ACR Practice Parameter for Communication of Diagnostic Imaging Findings.(ACR, n.d.; ACR-SPR-STR, n.d.)

2.5 Medical Imaging in Lung Cancer

The evaluation of lung cancer relies on several imaging modalities, including chest radiography, computed tomography (CT), magnetic resonance imaging (MRI), and hybrid techniques such as positron emission tomography combined with CT (PET/CT). Among these, chest radiography remains the most performed diagnostic imaging test worldwide for the assessment of numerous pulmonary diseases. Its advantages include simplicity, low cost, minimal radiation exposure,

and wide availability. Standard posteroanterior (PA) and lateral radiographic views obtained in the standing position provide optimal visualization and can reveal subtle parenchymal abnormalities. However, small lesions may still be obscured by overlying anatomical structures. Dual-energy radiography, which acquires images using both high- and low-energy X-rays, has demonstrated improved detection of pulmonary nodules and may enhance screening performance.(de Wever et al., 2011)

Historically, chest radiography was the first modality evaluated for lung cancer screening, and multiple large, randomized trials in the United States and Europe investigated its effectiveness. The consistently disappointing results from these studies, showing no significant mortality reduction, led to the abandonment of chest radiography as a screening tool. Although limited for early detection, chest radiography still plays a role in staging by demonstrating lesion size, post-obstructive atelectasis, pleural effusion, or signs of extrapulmonary spread. Nevertheless, it lacks the ability to accurately assess chest wall invasion or subtle mediastinal disease(de Wever et al., 2011).

Computed tomography (CT) represents the second most important modality in thoracic imaging and has become central to lung cancer diagnosis and staging. Advances in multidetector CT (MDCT) technology have significantly enhanced diagnostic capabilities. Improved temporal resolution allows rapid acquisition of thin-section images during a single breath-hold, facilitating optimal contrast enhancement and minimizing motion artifacts. The increased spatial resolution of MDCT enables continuous thin-slice acquisition, improving the quality of multiplanar reconstruction (MPR) images. Coronal and sagittal MPRs are particularly valuable for assessing the relationship of tumors to adjacent anatomical structures that may not be well evaluated on axial images alone, thereby increasing radiologists' confidence when evaluating fissural involvement, diaphragmatic invasion, or mediastinal extension. See Figure 2.6 (a, b, c) (de Wever et al., 2011)

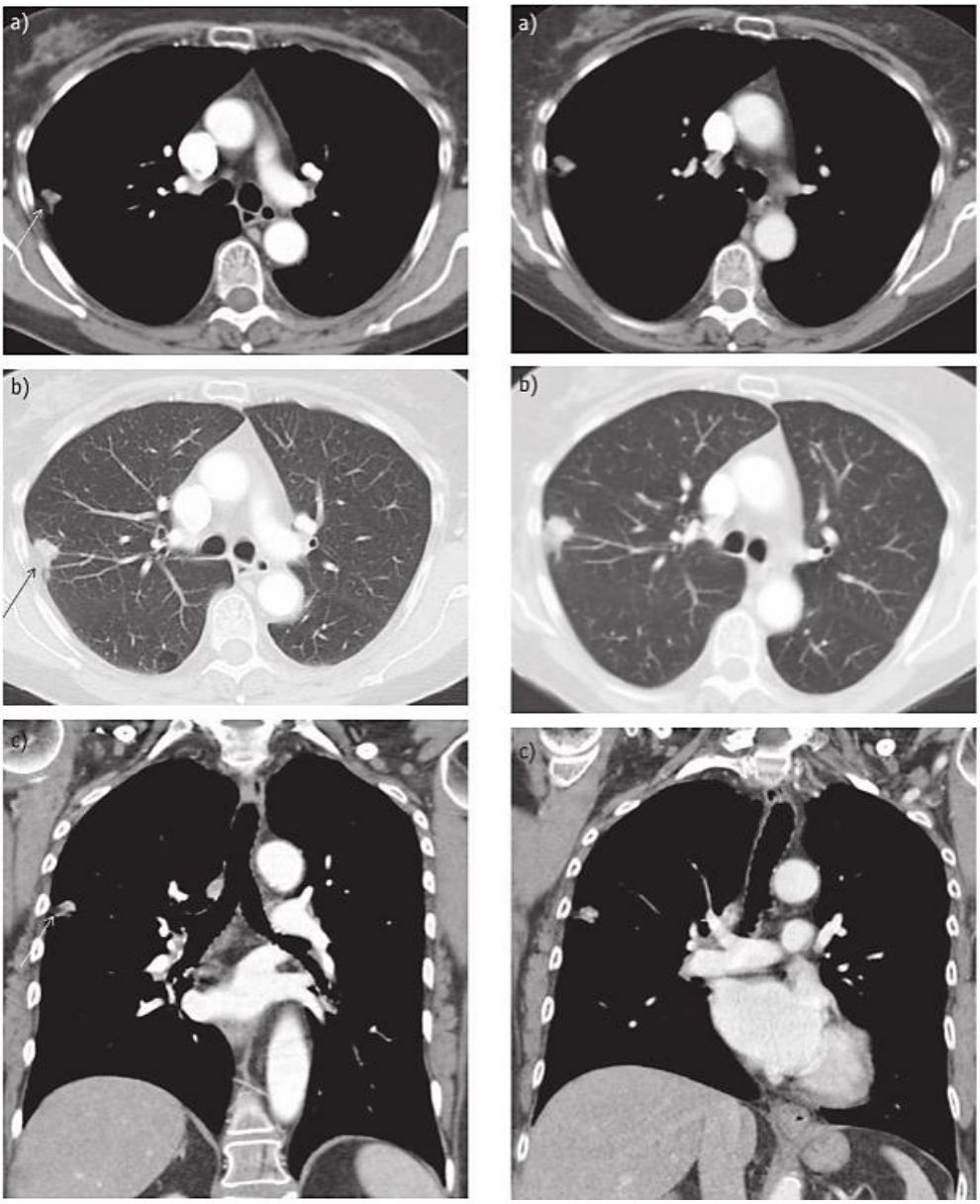


Figure 2.6 Incidental Detection of a Nodular Lesion in the Right Upper Lobe. Adapted from (de Wever et al., 2011)

(a) Axial CT image in the mediastinal window, (b) axial CT image in the lung window, and (c) coronal multiplanar reconstruction demonstrating a right upper lobe nodule with a pleural tail (arrow). Adapted from (de Wever et al., 2011)

Additional MDCT reconstruction techniques, such as maximum intensity projection (MIP) and three-dimensional (3D) rendering, further enhance diagnostic performance. MIP images improve the detection of small pulmonary nodules and have been shown to reduce the number of overlooked early cancer, see Figure 2.7. They also minimize radiologist workload by reducing the number of images requiring review compared with conventional thin-slice axial

datasets. Three-dimensional airway reconstructions facilitate the identification of mild or focal airway stenoses and can produce endoluminal views that resemble bronchoscopic images.

CT remains the imaging modality with the highest sensitivity for pulmonary nodule detection. Moreover, technical developments in MDCT have allowed reductions in radiation dose without sacrificing diagnostic utility. Low-dose CT (LDCT) has emerged as the preferred modality for lung cancer screening in high-risk individuals. Clinical trials and real-world screening programs have shown that LDCT detects a high proportion of early-stage non-small cell lung cancers at baseline, with decreasing detection rates on subsequent annual examinations. Although LDCT may yield false-positive findings, structured diagnostic algorithms have minimized unnecessary invasive procedures, improving the overall safety and effectiveness of screening.

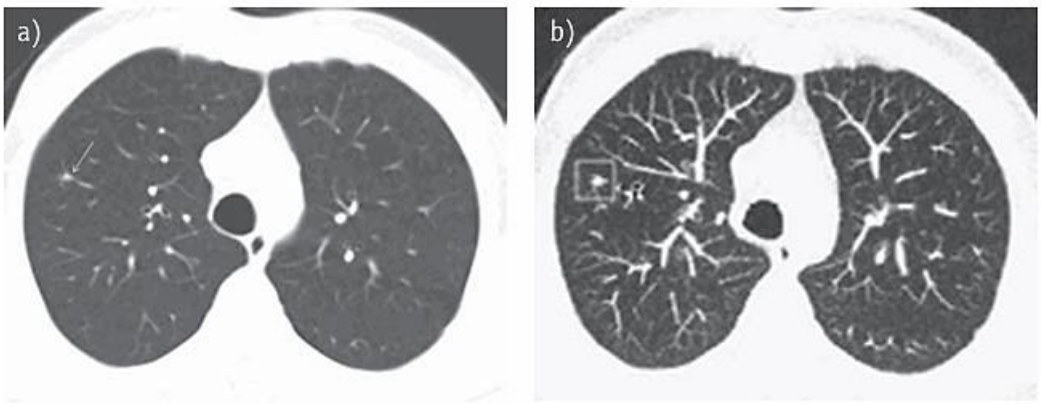


Figure 2.7(a) Axial 5-mm CT image shows no definite lung nodule, with only a subtle irregularity in the right upper lobe (arrow). (b) Corresponding MIP image clearly reveals a pulmonary nodule (square), illustrating the improved sensitivity of MIP for small nodules. Adapted from (de Wever et al., 2011)

Magnetic resonance imaging (MRI) has historically played a limited role in lung evaluation due to several inherent challenges. These limitations include its high susceptibility to motion artifacts caused by cardiac pulsation and respiratory motion, the intrinsically low proton density of lung parenchyma, and the marked signal loss at air–soft tissue interfaces. Consequently, for many years MRI was considered primarily a problem-solving modality, used selectively in cases where CT was insufficient. Typical indications included:

- assessing tumor invasion of the chest wall or mediastinal structures, particularly in superior sulcus (Pancoast) tumors.
- differentiating between solid and vascular hilar masses.
- evaluating diaphragmatic dysfunction; and
- staging and monitoring treatment response in mediastinal lymphoma

Recent advancements, however, have expanded the clinical utility of MRI in thoracic oncology. Whole-body MRI (WBMR) has demonstrated increased sensitivity for the detection of distant

metastatic disease and may play an important role in comprehensive staging. In addition, diffusion-weighted imaging (DWI) has emerged as a powerful functional technique that provides information on tumor cellularity and cell membrane integrity. Because water mobility is more restricted in tissues with high cellular density, such as malignant tumors, DWI enables improved tumor detection, characterization, and treatment response assessment. see Figure 2.8. As these applications continue to advance, MRI is increasingly used as a complementary tool, particularly in situations where CT findings are inconclusive

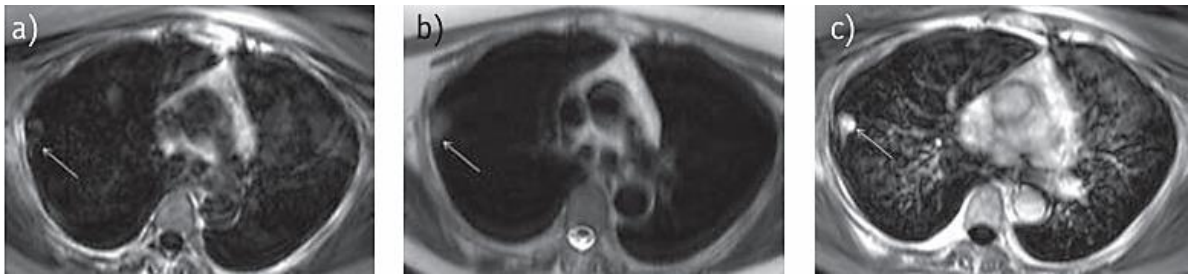


Figure 2.8 Axial magnetic resonance (MR) images of the right upper lobe. (a) T1-weighted and (b) T2-weighted images demonstrate a nodular lesion (arrow). (c) Post contrast T1-weighted image shows contrast enhancement of the lesion (arrow). Adapted from (de Wever et al., 2011)

Positron emission tomography (PET) is a highly sensitive and specific functional imaging modality that enables in vivo visualization of metabolic activity within human tissues. In thoracic oncology, the most commonly used radiotracer is fluorodeoxyglucose (FDG), a glucose analogue. After cellular uptake through glucose transporters, FDG is phosphorylated but not further metabolized, resulting in intracellular trapping. Because malignant cells exhibit increased glucose metabolism, they accumulate significantly higher levels of FDG compared to normal tissue. The standardized uptake value (SUV) is the most widely used semi-quantitative metric for evaluating FDG accumulation(de Wever et al., 2011).

Integrated PET/CT systems combine the metabolic information provided by PET with the anatomical detail of CT within a single imaging session. This integrated approach offers several diagnostic advantages, including improved detection of lesions not clearly visible on CT or PET alone; enhanced anatomical localization of abnormal FDG uptake; better delineation of lesions from surrounding structures; and more accurate characterization of lesions as benign or malignant.

PET/CT is regarded as the most accurate noninvasive imaging technique for assessing the T-stage of lung cancer. A key advantage is its ability to distinguish viable tumor tissue from distal obstructive atelectasis, a distinction that is often difficult on CT alone Figure 2.9. By integrating metabolic and anatomical data, PET/CT significantly improves diagnostic confidence, staging accuracy, and treatment planning in lung cancer.(de Wever et al., 2011)

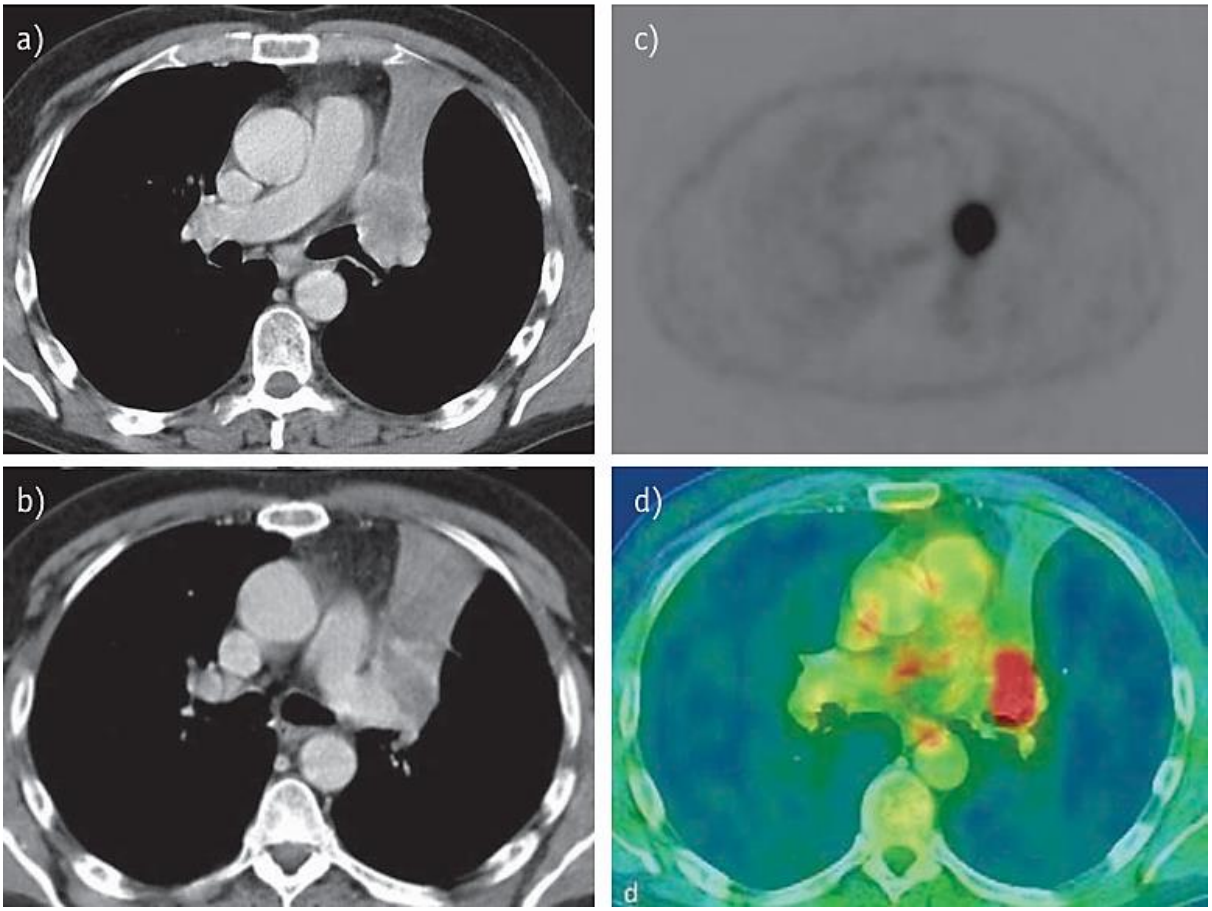


Figure 2.9 Integrated PET/CT of a central left lung tumor. (a–b) Axial CT images show a central tumor with associated retro-obstructive atelectasis. (c) PET demonstrates FDG uptake within the tumor but not within the atelectatic lung, enabling accurate tumor localization on fused PET/CT images. Adapted from (de Wever et al., 2011)

Chapter 3

Literature Review

This chapter describes the main evolutions in AI-based analysis of pulmonary nodules on chest CT, providing the conditions needed to position the present study within the existing research landscape. It reviews key approaches in nodule detection, deep learning–based classification, radiomics, multimodal fusion, enhancement techniques, and Lung-RADS–aligned risk stratification.

The literature is structured to highlight prevailing trends, recurring dataset limitations, and challenges such as dose variability, inconsistent preprocessing pipelines, and limited external validation. Particular attention is given to studies that incorporate both LDCT and SDCT data, apply enhancement techniques, or address multi-class nodule classification, as these themes closely align with the objectives of this thesis. Together, the gaps identified across prior research underscore the need for a dose-aware, enhancement-optimized Lung-RADS classification framework.

3.1 Image Preprocessing in Lung CT Computer-Aided Diagnosis

Image preprocessing is a fundamental component of lung CT based computer-aided diagnosis (CADx) systems, playing a crucial role in determining the performance, robustness, and generalizability of deep learning models. Raw CT data acquired in clinical practice are inherently heterogeneous due to variations in scanner manufacturers, acquisition protocols, reconstruction kernels, slice thickness, and radiation dose. Consequently, the objective of preprocessing is to transform these raw data into a standardized and structured representation

that enhances image quality, reduces irrelevant variability, and ensures compatibility with downstream learning algorithms.

A broadly consistent preprocessing pipeline has been adopted across recent studies, typically including conversion of raw pixel values to Hounsfield Units (HU), spatial resampling to isotropic voxel resolution, lung parenchyma segmentation, intensity windowing and normalization, and dataset-level intensity centering(Marinakis et al., 2024). These steps are configured not only to improve visual image quality but also to stabilize feature distributions and optimize the learning behavior of convolutional neural networks.

The conversion of raw image intensities into the standardized Hounsfield Unit (HU) scale is the initial core step in lung CT preprocessing. The HU scale is derived by linearly transforming the measured linear attenuation coefficients, such that distilled water at standard temperature and pressure is assigned a value of 0 HU, whereas air is defined as −1000 HU. This standardized representation enables consistent tissue characterization across scanners and acquisition settings, independent of photon statistics or detector configuration. By mapping tissues to well-defined HU ranges, lung parenchyma, soft tissues, and high-density structures such as bone can be reliably distinguished. HU-based thresholding is therefore commonly employed to suppress unwanted components, including surrounding air and non-pulmonary regions, thereby improving contrast and focusing the analysis on relevant lung-specific anatomies(Marinakis et al., 2024).

After HU conversion, spatial resampling is applied to address geometric inconsistencies inherent in volumetric CT data. Clinical CT scans often display anisotropic voxel spacing, with in-plane resolution differing from slice thickness. Such non-uniformity can distort anatomical proportions and adversely affect convolutional operations, particularly in three-dimensional deep learning models. To mitigate these effects, studies typically resample CT volumes to an isotropic voxel resolution, ensuring equal spacing along all spatial axes. In investigations using benchmark datasets such as LIDC-IDRI(Armato, McLennan, Bidaut, McNitt-Gray, et al., 2011) and LUNA16(Setio et al., 2017a), an isotropic resolution of 1 mm × 1 mm × 1 mm is commonly adopted. Accurate spatial feature extraction is facilitated by this standardization, which improves reproducibility across studies and enhances the robustness of learned representations(Marinakis et al., 2024).

Another critical preprocessing stage is represented by Lung segmentation, which aims at isolating the lung parenchyma from surrounding anatomical structures, including the chest wall, ribs, mediastinum, and adjacent soft tissues. Removing these irrelevant regions reduces background noise, limits feature contamination and decreases computational complexity. Significantly, segmentation algorithms must be designed with particular care near the pleural boundaries. Juxta-pleural nodules may lie at or slightly beyond the lung edge, and excessively aggressive segmentation risks excluding clinically relevant findings. Accordingly, most studies preserve a narrow peripheral margin around the lungs to ensure that subtle nodular structures are retained during mask generation.

Intensity windowing and normalization are applied to further enhance image quality and learning stability, following spatial and anatomical standardization. Lung CT images span a wide dynamic range of HU values, typically from approximately −1024 HU to values exceeding 2000 HU. To assert pulmonary structures and suppress high-density tissues such as bone, intensities are commonly clipped to a lung-specific window of −1000 to 400 HU. This windowing strategy, widely adopted in challenges such as LUNA16(Setio et al., 2017a), enhances local contrast within the lung parenchyma while reducing the influence of irrelevant high-intensity regions. Following clipping, voxel values are linearly normalized to a fixed range, most often [0, 1], to ensure numerical stability and prevent any single intensity range from dominating the learning process. This normalization procedure is mathematically expressed in Equation. (Marinakis et al., 2024)

$$Normalized\ Value = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (1)$$

where X denotes the current value, $X_{min}$is the minimum value, and X_{max} is the maximum value.

As a final improvement step, zero-centering is applied considerably to improve optimization efficiency during network training. Zero-centering shifts the normalized intensity distribution so that its mean becomes zero, facilitating faster convergence and more stable gradient updates. This is completed by subtracting a global mean pixel value that is computed across the entire dataset from each voxel. For example, the LUNA16 dataset(Setio et al., 2017a) reports an average normalized pixel intensity of approximately 0.25, which is subtracted from all images after normalization. CT scanners are calibrated to produce consistent HU measurements; per-image mean subtraction is discouraged, as it may introduce bias. Instead, the mean must be calculated over the entire dataset to ensure a stable and representative normalization. This preprocessing step is expressed in Equation 2 (Marinakis et al., 2024).

$$Zero\_centered\_image = image\_pixel\_values - Pixel\_Mean \ (2)$$

where image_pixel_values denote the image array values, and PIXEL_MEAN is the average value of the dataset.

Collectively, a coherent and widely adopted pipeline that forms these preprocessing steps, which enhances lung CT image quality, reduces inter-scan variability, and improves the robustness and generalizability of deep learning–based CADx systems. By standardizing spatial resolution, intensity distributions, and anatomical focus, this preprocessing framework provides a definitive foundation for accurate pulmonary nodule detection and classification.

3.2 Patch and ROI Extraction

Another essential preprocessing step is patch extraction. For training deep learning models designed for pulmonary nodule detection, segmentation, or classification, nodule samples must be extracted from the available CT datasets. Using radiologist-provided pixel-level annotations, pulmonary nodules were first localized and directly defined as regions of interest (ROIs). Each ROI was obtained by cropping a volumetric region centered on the nodule annotation, with an additional contextual margin included to preserve surrounding anatomical information. These ROIs were subsequently used to generate representative training samples. Depending on the network architecture, the samples were generated as two-dimensional slices or three-dimensional patches. Three-dimensional patches, often referred to as multi-view patches, preserve volumetric information across the axial, coronal, and sagittal planes, thereby providing richer contextual information.(Marinakis et al., 2024)

In the LIDC-IDRI dataset(Armato, McLennan, Bidaut, McNitt-Gray, et al., 2011), radiologists assign each nodule a malignancy score ranging from 1 to 5. Common practice in the literature is to merge classes 1 and 2 (Highly Unlikely and Moderately Unlikely) into a single benign category, and to combine classes 4 and 5 (Moderately Suspicious and Highly Suspicious) into a malignant category. Class 3, representing an indeterminate category, is typically excluded due to its diagnostic uncertainty.(Marinakis et al., 2024)

3.3 Data Augmentation in Lung CT

Data augmentation is a critical preprocessing step that can substantially improve the performance and generalization of deep learning models. Its primary purpose is to generate synthetic variations of the training data to reduce overfitting and enhance model robustness. Augmentation is especially important when dealing with imbalanced datasets, where one or more classes are underrepresented.

To address class imbalance, traditional augmentation techniques such as rotations, translations, scaling, shearing, flipping, cropping, and duplicating nodule samples are commonly applied to increase the diversity of training examples. Another approach involves adjusting the sampling rates of positive and negative nodules to ensure a more balanced distribution across classes. In addition to these classical methods, more advanced data generation techniques, such as Generative Adversarial Networks (GANs), have been explored to create realistic synthetic nodules for improved model training.(Marinakis et al., 2024)

In addition to the anatomical and geometric preprocessing steps described above, intensity-based contrast enhancement is frequently used to improve the visibility of lung structures and pulmonary nodules before model training.

3.4 Contrast-Enhancement Techniques (HE, AHE, CLAHE)

Contrast enhancement is fundamental to CT interpretation. Clinically, radiologists adjust window width (WW) and window level (WL) to optimize visualization of soft tissues, lung parenchyma, and bone, each requiring distinct window settings, as seen in Figure 3.1 Although digital viewing systems allow rapid switching between window presets, repeatedly adjusting WW/WL can be inefficient. Computational enhancement methods that simultaneously improve visibility across multiple tissue ranges can therefore streamline interpretation.

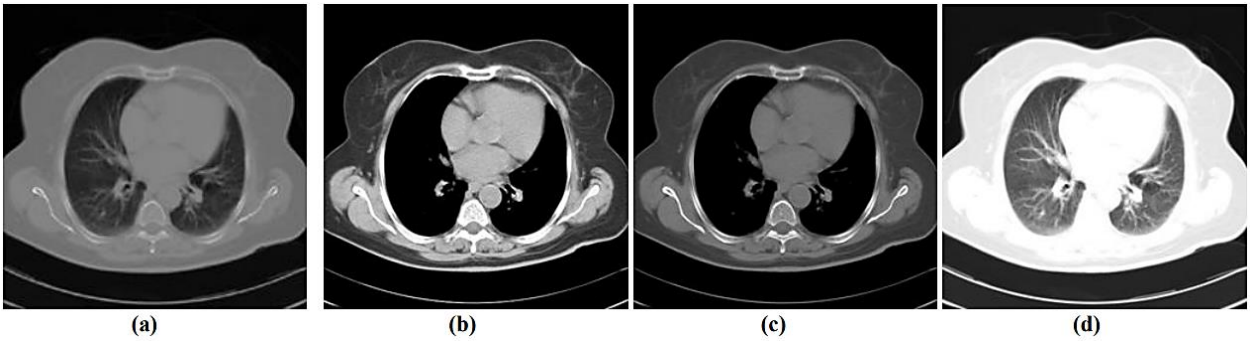


Figure 3.1 Chest CT image shown with multiple window settings: (a) original image and optimized brightness/contrast for (b) soft tissue, (c) bone, and (d) lung parenchyma. Adapted from (Jin et al., 2001a)

Within histogram-based enhancement methods, two important techniques are Histogram Equalization (HE) and Contrast Limited Adaptive Histogram Equalization (CLAHE)(Jin et al., 2001a). Both have shown strong potential for improving the quality and diagnostic usefulness of medical images. In many CT images, important features occupy only a small portion of the available grayscale range. Contrast enhancement techniques expand this limited grayscale region so that the structures of interest are displayed with greater clarity while maintaining the overall image quality. The goal is to determine a suitable transformation between the original gray levels and the output intensities so that the contrast between adjacent structures is maximized.

Histogram-based techniques are among the most widely used computational approaches for contrast enhancement in medical imaging. They aim to expand the limited grayscale range occupied by diagnostically relevant features, thereby improving the separation between adjacent structures. Histogram Equalization (HE) redistributes intensities to approximate a uniform

histogram, effectively enhancing global contrast in images with narrow tonal ranges. However, HE often amplifies noise and produces artificial-looking images, making it less suitable for diagnostic CT (Jin et al., 2001a).

Adaptive Histogram Equalization (AHE) improves upon HE by computing local histograms within pixel neighborhoods, enabling spatially adaptive enhancement. Yet, AHE tends to over-amplify noise in homogeneous regions and may introduce halo or ring artifacts around strong edges.

Contrast-Limited Adaptive Histogram Equalization (CLAHE) was designed to reduce these limitations. CLAHE applies histogram equalization to small image tiles but constrains the amplification of local histograms using a clipping limit, thereby suppressing noise while preserving fine structural detail. Bilinear interpolation between neighboring tiles ensures smooth transitions without boundary artifacts.

A multi-scale framework using AHE was described by(Jin et al., 2001a), who applied AHE at different dyadic decomposition levels with window sizes adapted to each sub-band's frequency content. In qualitative comparisons among unsharp masking, HE, AHE, and CLAHE, expert readers consistently preferred CLAHE-processed images for providing balanced enhancement without excessive noise. Consequently, CLAHE has become a widely accepted reference method for local contrast enhancement in CT and other modalities. Beyond CT, CLAHE has been successfully applied in ultrasound, MRI, and X-ray imaging to enhance lesions and subtle structures while maintaining noise stability. Its advantages include controlled local enhancement, robustness to intensity heterogeneity, computational efficiency, and preservation of diagnostic detail. Nonetheless, performance remains sensitive to parameter selection, and CLAHE may not address global intensity disparities or may produce unnatural contrast transitions in smoothly varying regions.

Recent research has explored noise-aware variants, hybrid global local enhancement pipelines, and deep learning-based methods for automated parameter optimization, further improving the adaptability and performance of CLAHE in clinical imaging workflows.

The study by (Jin et al., 2001a) compared several enhancement methods, including unsharp masking, HE, AHE, and CLAHE, shown by Figure 3.2. Expert radiologists consistently preferred the CLAHE-processed images, noting that they provided clearer structural detail without over-enhancement or noise amplification. Based on their findings, CLAHE was selected as the reference enhancement method for subsequent analyses.

.

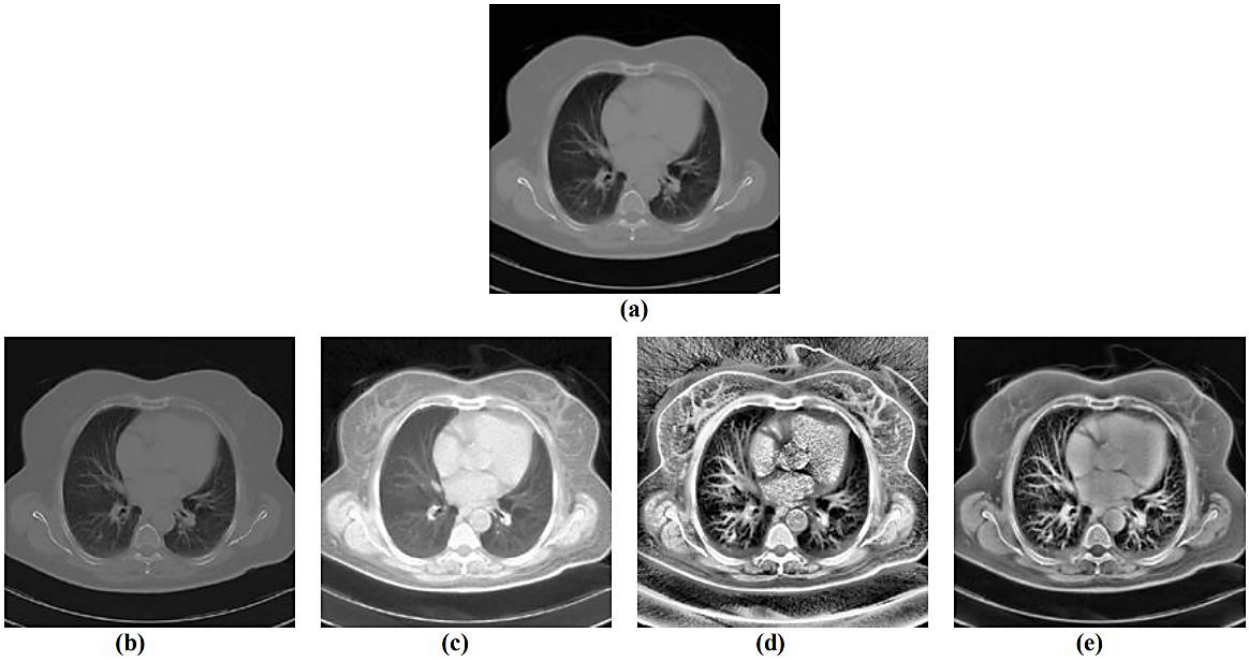


Figure 3.2 Contrast enhancement of a chest CT image using traditional techniques: (a) original image; (b) unsharp masking; (c) histogram equalization (HE); (d) adaptive histogram equalization (AHE); and (e) contrast-limited adaptive histogram equalization (CLAHE). Adapted (Jin et al., 2001a)

3.5 Unsharp Masking and Fractional-Derivative Filters

CT images often contain low-quality pixels, distortions, and various anomalous details that can obscure lesion visibility and contribute to false detections. Image enhancement is therefore necessary to mitigate these issues. One of the most widely used enhancement methods is unsharp masking (UM), which generates a sharpened output image by adding a scaled high-pass–filtered version of the input image back to the original. Although effective for contrast and edge enhancement, conventional UM suffers from notable limitations: it tends to overshoot high-contrast regions, introduces unwanted artifacts, and amplifies existing noise in the CT scan.(Satish et al., n.d.; Shukla et al., 2022a)

To address these issues, previous study (Shukla et al., 2022a) have proposed integrating fractional-derivative-based operators with UM. In particular, combining a Grunwald–Letnikov (GL) fractional derivative mask with a Laplacian framework has been shown to improve texture, edge definition, and overall contrast. The GL mask, an extension of the classical Euler method, acts as a refined high-pass operator, reducing the excessive noise amplification characteristic of traditional UM.

In the modified approach, the CT image is first converted to grayscale, after which the GL fractional derivative mask is applied and scaled using a factor between 0 and 1. The resulting filtered image is then added back to the original grayscale image to produce the enhanced output. Comparative results as provide in Figure 3.3 demonstrate that the modified UM technique alleviates over-brightness and noise amplification, outperforming conventional UM in terms of noise suppression, contrast improvement, and edge sharpening. An increase in the EME percentage corroborates the visual enhancements achieved by the proposed method(Shukla et al., 2022a).

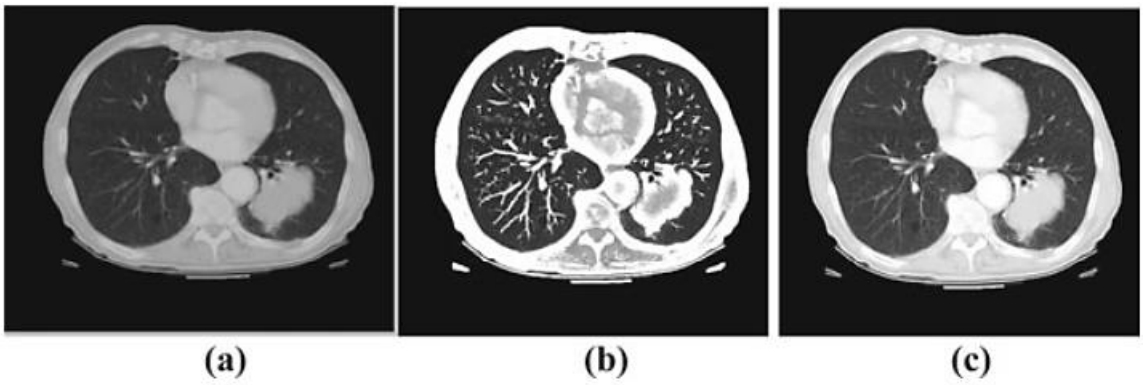


Figure 3.3: Unsharp masking comparison: (a) original image, (b) conventional UM, and (c) GL-based improved UM. Adapted from(Shukla et al., 2022a)

3.6 Comprehensive Literature Review of the study

Deep learning has rapidly advanced the detection and classification of pulmonary nodules on chest CT images. Early CAD systems based on handcrafted features have been replaced by 3D CNNs, transfer learning, and multi-scale architectures, which greatly improve nodule identification and benign–malignant discrimination. Recent trends emphasize attention mechanisms, lightweight models for real-time use, and CT enhancement techniques such as super-resolution to boost CAD accuracy. Multimodal approaches integrating radiomics, clinical data, and longitudinal CT information further enhance malignancy prediction. Collectively, these advances point toward more accurate, efficient, and clinically integrated AI tools for lung cancer screening and diagnosis.

This section synthesizes major research contributions in lung nodule detection and classification over the past decade. Studies are reviewed in sequential and methodological order, highlighting datasets used, preprocessing pipelines, deep learning architectures, and evaluation metrics. Each study's strengths and limitations are analyzed to identify gaps relevant to the development of the proposed framework.

(J. Gong et al., 2018) proposed an automated computer-aided detection (CADe) scheme for pulmonary nodule detection that combines 3D tensor filtering with local image feature analysis. The system was evaluated on two benchmark datasets: LUNA16(Setio et al., 2017b) and ANODE09(B. S. M. P. M. et al. van Ginneken, 2010). LUNA16 comprises 888 CT scans with 1,186 nodules extracted from the LIDC-IDRI database (slice thickness >2.5 mm, 120–140 kV, 40–627 mA), with each nodule annotated by four radiologists and the reference standard defined as nodules ≥3 mm agreed upon by at least three radiologists. ANODE09 provided an independent validation set of five CT scans containing 39 nodules from the NELSON trial, reconstructed at 0.7 mm thickness and annotated by expert radiologists.

The proposed CADe workflow included lung volume segmentation, isotropic CT resampling, andidate detection using a 3D tensor filtering algorithm with local feature analysis, boundary refinement via 3D level-set segmentation, and feature selection using a Correlation-based Feature Selection (CFS) evaluator. Final classification was performed using a random forest classifier, with comparisons to J48 decision tree, logistic regression, and SVM models. Using 10-fold cross-validation, the method achieved a sensitivity of 79.3% at 4 false positives per scan (FP/scan) on LUNA16 and 84.62% at 2.8 FP/scan on ANODE09. The combination of volumetric tensor filtering and machine-learning-based false-positive reduction demonstrated clear improvements in candidate detection and overall robustness.

However, several limitations were identified, including the small size of the independent ANODE09 validation set, lack of detailed nodule subtype labels, and assumptions that nodules

are ball- or dot-like, which may reduce accuracy for irregular or attached lesions. Additionally, the study lacked clinical validation, leaving the real-world applicability of the system untested.

(L. Gong et al., 2019) proposed a fully automated computer-aided detection (CADe) system for pulmonary nodule detection using 3D deep convolutional neural networks augmented with squeeze-and-excitation residual (SE-ResNet) modules. The framework integrates a 3D U-Net–based region proposal network (RPN) for candidate generation with a 3D SE-ResNet classifier for false-positive reduction, enabling accurate identification of nodules across a broad size range, including very small lesions (<5 mm). The system was assessed on the LUNA16 dataset(Setio et al., 2017b), comprising 888 CT scans and 1,186 nodules, with reference labels established by the consensus of at least three of four expert radiologists. Performance was evaluated using tenfold cross-validation to ensure robustness.

The method demonstrated high detection capability, achieving sensitivities of 93.6% and 95.7% at 1 and 4 false positives per scan, respectively, and a CPM score of 0.904. Its strong performance on multi-size nodules, particularly extremely small nodules, was emphasized as a major advantage. Nonetheless, the study acknowledged several limitations, including the substantial computational burden of 3D DCNNs, challenges in segmenting nodules with irregular shapes or adjacent structures, and underrepresentation of certain nodule types within the LUNA16 dataset(Setio et al., 2017b). The authors highlighted the need to extend the system toward diagnostic classification in future work to enhance its clinical utility.

(Kozuka et al., 2020) evaluated whether a deep learning–based CAD system could enhance pulmonary nodule detection by less experienced radiologists and reduce reading time. The reference standard for nodules ≥3 mm was established by a panel of three expert radiologists (26, 6, and 12 years of experience). Two senior radiologists independently marked all nodules, and disagreements were resolved by a third expert. This expert consensus served as the gold standard for evaluating reader performance.

Two less experienced radiologists (5 and 1 year of experience) then interpreted 120 chest CT scans in two sessions: first without CAD, and after a 14-day interval, with CAD assistance. True-positive, false-positive, and false-negative findings were determined based on agreement with the expert-derived gold standard. The CAD system, built on a two-stage Faster R-CNN architecture comprising a region proposal network (RPN), ROI pooling, and classification/regression heads, analyzed 1-mm reconstructed images and detected solid, part-solid, calcified, and ground-glass nodules.

CAD support substantially improved detection performance. Overall sensitivity increased from 20.9% to 38.0%, with significant gains for small nodules (3–6 mm: 13.7% to 32.4%) and

medium nodules (6–10 mm: 33.3% to 47.6%). Although PPV decreased slightly (70.5% to 61.8%), the F1-score rose from 32.2% to 47.0%, and reading time decreased by 11.3%. CAD alone achieved a sensitivity of 70.3%, PPV of 57.9%, and an F1-score of 63.5%, highlighting its potential value as an assistive tool for novice readers.

The authors noted several limitations: absence of histologic confirmation, reliance on 1-mm thin-slice images rather than standard 5-mm clinical protocols, and possible unfamiliarity of junior readers with thin-slice interpretation. Lower sensitivity compared with prior literature likely reflected reader inexperience and the inherent ambiguity of defining lung nodules, especially ground-glass and pleural-adjacent lesions. The study emphasized the need for broader validation and the integration of automated nodule detection to support more reliable deployment in clinical settings.

(S. Zhang et al., 2019a) presented a transfer learning-enhanced convolutional neural network based on the classic LeNet-5 architecture for pulmonary nodule classification using thoracic CT images from the LIDC-IDRI dataset(Armato, McLennan, Bidaut, McNitt-Gray, et al., 2011). Their approach employed a two-stage binary classification scheme: initially differentiating malignant nodules from non-nodule structures ≥3 mm, followed by classifying malignant nodules into serious- and mild-malignant categories according to radiologist-assigned malignancy scores. Image annotations were provided by four experienced thoracic radiologists, and the dataset was divided into training (80%), validation (10%), and testing (10%) subsets. To mitigate the constraints of limited sample size, the authors applied rotation- and shift-based augmentation and incorporated transfer learning to enhance feature representation. Ten-fold cross-validation was used to evaluate model robustness and generalizability.

The LeNet-5 model demonstrated strong performance, achieving 97.041% accuracy in distinguishing malignant nodules from non-nodules and 96.685% accuracy in differentiating malignancy severity, outperforming AlexNet on the same dataset. Despite these favorable results, the authors acknowledged key limitations, including reliance on transfer learning, restricted availability of annotated training data, and potential loss of spatial detail resulting from resizing CT images from 512 × 512 to 64 × 64, which may limit broader clinical applicability.

(Y. Wang et al., 2020) introduced a multi-path, multi-scale convolutional neural network (ConvNet) for classifying pulmonary nodules as benign or malignant, addressing the wide variability in nodule size, shape, and appearance in CT scans. The architecture employs multi-scale convolutional layers to capture both fine-grained local details and broader global structures, while parallel feature-extraction paths enhance representation diversity. The model was trained and evaluated on the LUNGx Challenge dataset(Armato et al., 2015a) using fivefold cross-validation. Preprocessing included contrast normalization and extensive augmentation random rotations (0–359°), horizontal-vertical flips, and 20% bounding-box translations to

generate spatial variability and reduce overfitting. Implemented in Caffe and optimized with Adagrad, the proposed ConvNet achieved a sensitivity of 0.887, specificity of 0.924, and an AUC of 0.948, marking a 14% improvement over state-of-the-art unsupervised approaches and surpassing several existing ConvNets. When used as a CADx tool, it further enhanced radiologists' diagnostic performance, yielding a 23% increase in AUC.

Despite these strengths, the study faced limitations, including a small dataset, class imbalance, and the need for manually cropped nodule regions, which limits scalability and clinical applicability. The authors recommended integrating automated nodule detection in future work to support fully end-to-end medical imaging workflows.

(Raza et al., 2023) introduced Lung-EffNet, a transfer-learning framework built on EfficientNetB1 for classifying lung CT images into three categories: benign, malignant, and normal. The study used the IQ-OTH/NCCD dataset(alyasriy, 2023) consisting of 1,097 CT slices (120 benign, 561 malignant, 416 normal) and applied an 80:20 train–test split. Preprocessing included lung-region cropping to remove irrelevant background, resizing to 240 × 240 × 3, standardization, normalization, and label encoding. To address severe class imbalance, the authors implemented targeted augmentation in the training set: benign images were augmented 13-fold, malignant 2-fold, and normal 3-fold using rotations, flips, brightness changes, zoom, and spatial shifts. EfficientNet variants (B0–B4) were fine-tuned for 50 epochs, with 10% of training images randomly selected each epoch for validation to monitor overfitting; EfficientNetB1 demonstrated the strongest performance.

Lung-EffNet achieved 99.10% accuracy, 98.63% precision, and ROC-AUC values between 0.97 and 0.99, outperforming several state-of-the-art CNNs. The model showed perfect precision for malignant cases and high specificity for normal images, highlighting the effectiveness of targeted augmentation and fine-tuning in enhancing discriminative performance.

The authors noted limitations related to the small dataset size, the risk of overfitting inherent to deep CNNs, and restrictions on publicly available annotated medical data. They recommended expanding datasets, exploring synthetic augmentation methods such as GANs, and validating the framework on larger, multi-institutional cohorts to improve generalizability.

Saha, Ganie, Pramanik, and colleagues (Saha et al., 2024) introduced VER-Net, an ensemble transfer-learning framework for multiclass lung cancer detection using CT images. The model integrates VGG19, EfficientNetB0, and ResNet101 to enhance feature representation across four lung cancer classes. The dataset, sourced from Kaggle, initially contained 1,000 CT images of adenocarcinoma, squamous cell carcinoma, and large-cell carcinoma, which were expanded to 1,653 images through random oversampling. Image preprocessing involved standardization, normalization, and resizing to 460 × 460 × 3, followed by data augmentation using shear (0.2),

zoom (0.2), rotation (24°), and both horizontal and vertical flipping. The final dataset was divided into training (64.48%), testing (26.98%), and validation (8.52%) subsets. VER-Net demonstrated strong performance, surpassing eight competing transfer-learning models and achieving 91% accuracy, 92% precision, 91% recall, and an F1-score of 91.3%. These results highlight the effectiveness of ensemble fusion in stabilizing predictions and improving multiclass discrimination relative to individual CNNs.

Despite its strengths, the study acknowledged several challenges, including potential distortions introduced by aggressive augmentation, the inclusion of low-resolution images in parts of the dataset, and increased computational demands inherent to ensemble architectures. The authors suggested that future work explore more efficient variants of the model and extend the framework to chest X-ray analysis and lung cancer severity assessment.

(Kumaran S et al., 2024a) proposed an integrated deep learning framework for lung cancer classification that combines features extracted from three pretrained CNNs, VGG16, ResNet50, and InceptionV3. The study used the publicly available IQ-OTH/NCCD dataset(alyasriy, 2023), comprising 120 benign, 561 malignant, and 416 normal CT images collected from an oncology center in Iraq. Preprocessing included resizing to 256 × 256 pixels, grayscale conversion, normalization, and augmentation (rotation, translation, and Gaussian blurring). Class imbalance was addressed using SMOTE. Features generated by the three CNN backbones were concatenated and passed through dense layers to perform final classification. The model achieved high performance, with an overall accuracy of 98.18% and strong precision, recall, and F1-scores across all categories. It demonstrated near-perfect detection of malignant cases and robust discrimination of benign and normal images, indicating the benefit of multi-architecture feature fusion. Grad-CAM visualizations were used to highlight salient regions influencing network predictions, improving interpretability from a clinical perspective.

However, the authors noted several limitations, including limited dataset diversity, potential demographic and acquisition-related biases, and challenges in generalizing performance to broader clinical environments. They emphasized the need for validation on heterogeneous datasets, improved model explainability, and strategies to enhance computational efficiency for real-world deployment.

(Majumder et al., 2024) proposed MENet, an ensemble deep learning framework that integrates the prediction outputs of three pretrained CNN architectures: Xception, InceptionResNetV2, and MobileNetV2, that using a fuzzy ranking fusion strategy based on the Mitscherlich function. This approach was designed to exploit complementary feature representations from each network and enhance overall classification robustness.

The model was trained and tested on two CT datasets: the IQ-OTH/NCCD dataset(alyasriy, 2023), comprising 120 benign, 561 malignant, and 461 normal images, and the LIDC-IDRI

dataset (Armato, McLennan, Bidaut, McNitt-Gray, et al., 2011)for external generalization. Transfer learning was applied to each base network prior to ensemble fusion. MENet achieved 99.54% accuracy on IQ-OTH/NCCD and 95.75% accuracy on LIDC-IDRI, outperforming individual classifiers and several prior state-of-the-art models. The results underscore the effectiveness of ensemble strategies in stabilizing predictions across heterogeneous imaging datasets.

Despite its strong performance, the model exhibited reduced sensitivity for micro-nodules (<3 mm) and generated false-positive and false-negative predictions that could impact clinical applicability. The authors also noted challenges in comparing results across studies due to dataset inconsistencies and highlighted the need for lightweight architectures and improved interpretability to facilitate real-world deployment.

(Kumar et al., 2024) developed a deep learning–based decision support system for lung cancer prediction, focusing on improving diagnostic accuracy across four CT-based categories: normal lung (Class 0), large cell carcinoma (Class 1), adenocarcinoma (Class 2), and squamous cell carcinoma (Class 3). Using 1,000 DICOM images from the LIDC-IDRI dataset(Armato, McLennan, Bidaut, McNitt-Gray, et al., 2011), the authors applied transfer learning to three pretrained CNN architectures: ResNet-50, EfficientNet-B3, and ResNet-101, each fine-tuned to classify the four categories. Data augmentation was employed to mitigate overfitting, and an ensemble fusion strategy was used to stabilize predictions and enhance robustness. Models were trained using the Adam optimizer (learning rate = 0.001) for 35 epochs with a batch size of 32, incorporating batch normalization, dropout, fully connected layers, and softmax activation.

The ensemble framework achieved superior performance relative to all individual models, reaching 99.44% accuracy and 100% precision for squamous cell carcinoma, demonstrating the diagnostic value of combining multiple pretrained architectures. However, the authors noted several challenges, including the limited dataset size, potential overfitting, and reliance on a narrow range of model architectures. They emphasized the need for larger and more diverse imaging datasets, improved annotation quality, and further architectural refinement to enhance clinical applicability and generalizability.

(Klangbunrueang et al., 2025) investigated the performance of a pretrained VGG16 model for CT-based lung cancer detection in a three-class classification setting. Using 1,097 CT images from a public dataset, the authors applied rotation, scaling, and flipping for augmentation and adapted VGG16 through ImageNet-based transfer learning followed by fine-tuning on the medical dataset. Model performance was assessed using a 10% holdout test set to ensure that evaluation data remained unseen, with the authors noting that cross-validation was not implemented due to computational limitations. To enhance interpretability, Grad-CAM heatmaps were generated and overlaid on CT images, with radiologists annotating relevant areas such as nodules and lesions. This approach provided visual confirmation that the network's

attention aligned with clinically meaningful regions, thereby improving confidence in model decisions.

Among all evaluated architectures: including ResNet50, MobileNetV2, and InceptionV3, VGG16 achieved the highest accuracy (98.18%), establishing it as a strong baseline model for lung cancer classification using 2D CT images. However, the authors acknowledged limitations related to the small dataset size, manual data splitting, and reliance on 2D inputs, which may restrict generalizability. They recommended future work involving larger datasets, 3D or multi-view imaging, and the exploration of advanced architectures such as Vision Transformers or 3D CNNs, as well as the use of interpretable AI techniques like SHAP or LIME.

(UrRehman et al., 2024) proposed a customized CNN architecture incorporating a dual attention mechanism to enhance the detection and classification of pulmonary nodules, particularly addressing challenges related to limited labeled samples and background interference in CT images. The model was trained on the LUNA16 dataset and used SimpleITK for preprocessing, with data augmentation applied to broaden sample variability. Training employed stochastic gradient descent with adaptive learning-rate scheduling and binary cross-entropy loss to optimize convergence. The proposed framework demonstrated strong diagnostic capability, achieving 94.69% sensitivity and an AUC of 98.00%, outperforming several recent approaches and showing reliable discrimination between benign and malignant nodules. The study also acknowledged practical constraints, including the high memory and computational demands of 3D architectures, the need to balance accuracy with available computing resources, and challenges in selecting optimal decision thresholds for clinical deployment.

(Canayaz et al., 2024) proposed an integrated deep learning framework for pulmonary nodule analysis, combining classification and segmentation to support early lung cancer diagnosis. The dataset consisted of CT images from a clinical cohort: 199 images from 29 malignant cases, 202 images from 68 benign cases, and 343 normal lung images from 67 patients, used across both tasks. For segmentation, the authors employed U-Net, LinkNet, and FPNet architectures with InceptionV3, DenseNet121, and SeResNet101 backbones, while nodule classification was performed using their hybrid C-EffxNet model. All models were trained with the Adam optimizer (learning rate 0.0005), categorical cross-entropy loss, batch size 8, and 100 epochs. The proposed system demonstrated strong performance. C-EffxNet achieved 97.98% accuracy, 98.02% precision, and an F1-score of 0.9798, with a Cohen's kappa of 0.969. For segmentation, the best results reached a Jaccard index of 0.8026 and a Dice coefficient of 0.8877, indicating robust delineation of nodule boundaries across architectures. The authors noted that clinical applicability could be further improved through larger datasets and a unified interface combining both segmentation and classification in a single workflow.

(Li et al., 2023) proposed a multimodal, graph-based deep learning framework for multiclass classification of lung adenocarcinoma nodules, representing one of the first attempts to integrate

convolutional radiomics, semantic clinical descriptors, and graph neural networks (GNNs) within a unified predictive model. The study included 338 patients, divided into 80% training and 20% testing, with fivefold cross-validation. Regions of interest extracted from CT scans were processed through both CNNs and radiomics pipelines, and the resulting feature sets were fused using a graph-embedding representation module. A graph database incorporating nine semantic and clinical attributes such as nodule type, spiculation, lobulation, vacuole sign, air bronchogram, vessel morphology, tumor–lung interface, and pleural indentation, served as the structural foundation for GNN-based node classification.

To address incomplete or uncertain graph connections, an edge-generation network using GCN-based convolution and sigmoid functions was introduced to improve relational consistency and model stability. Node classification was performed using Graph Convolutional Networks (GCN) and Graph Attention Networks (GAT), with standard data augmentation applied to ROI patches. Classical machine-learning models (SVM, Random Forest, AdaBoost) were included as baselines. The multimodal GNN framework outperformed traditional classifiers and achieved performance comparable to state-of-the-art deep learning approaches, reporting an accuracy of 66.3%, AUC of 75.9%, F1-score of 64.0%, and MCC of 48.4%. Incorporating the edge-generation module consistently improved predictive performance, while the combined use of semantic and clinical features enhanced model interpretability and better aligned predictions with radiological reasoning.

Limitations included the instability of graph construction due to subjective semantic annotations, challenges in augmenting multimodal data, and reliance on accurate manual labeling of nodule signs, which restricts applicability in fully automated workflows. The authors highlighted the need for more robust graph-building strategies and improved multitask detection of radiological signs to support broader generalizability.

(Qiu et al., 2022) proposed a deep learning approach leveraging sequential CT imaging to distinguish benign from malignant pulmonary ground-glass nodules (GGNs). The retrospective study included 762 patients, with 508 used for model development (1,524 CT sections) and 256 reserved for testing. A DenseNet121-derived network was employed to extract temporal features from two CT scans obtained at different time points, capturing longitudinal changes in nodule appearance. Data augmentation techniques including random shifts, rotations, flips, and zooming were applied to reduce overfitting, and the model was pretrained on ImageNet and fine-tuned using the GGN dataset. Extracted deep features were combined with clinical predictors and incorporated into a malignancy-prediction framework using ridge regression. Model performance was benchmarked against two trained radiologists through an observer study. The integrated model achieved an AUC of 0.841, outperforming a junior radiologist and performing comparably to a senior radiologist. Diagnostic accuracy improved for larger nodules (>10 mm) and with longer follow-up intervals, emphasizing the added value of temporal imaging features.

The authors noted limitations, including potential selection bias due to the single-center retrospective design and the limited interpretability of the deep learning model. They highlighted the need for external validation to strengthen generalizability and for future methods capable of enhancing transparency in model decision-making.

(Y. Liu et al., 2024) investigated whether incorporating the fibrotic microenvironment surrounding lung nodules could improve malignancy classification on chest CT. Their framework used two datasets: the publicly available LIDC-IDRI dataset (n = 1018) for nodule segmentation and model pre-training, and an in-house cohort of 1088 CT scans collected from 4500 patients for classification. In the in-house dataset, a senior radiologist annotated each nodule with a single-point marker using the VGG Image Annotator (VIA) tool.

Two 3D deep learning models were employed: a segmentation network to localize nodules and a classification network designed to differentiate malignant from benign nodules while also detecting fibrotic tissue in the surrounding parenchyma. To assess the contribution of the microenvironment, the authors generated three types of inputs: (1) nodule-only volumes, (2) nodules with their immediate microenvironment, and (3) nodules supplemented with semantic fibrosis metadata derived from a 3D attention-gated network (3D AG-Net). Classification was performed on 64 × 64 × 64 voxel patches using biopsy findings and radiologist reports as ground truth, and performance was evaluated through ten-fold cross-validation. The results demonstrated that including microenvironment information improved malignancy prediction, and the addition of fibrosis metadata yielded the highest accuracy and AUC. These findings highlight the potential diagnostic value of fibrosis-related contextual features, given their higher prevalence around malignant nodules.

Several limitations were noted. Shrinking nodules were automatically labeled as benign, which may underestimate classification accuracy because some decreasing lesions can still represent malignancy. In addition, segmentation errors such as partial removal of nodule boundaries or leakage of surrounding tissue could propagate into the classification stage and influence performance estimates. The authors suggested that more advanced segmentation approaches (e.g., AMSUnet or Dual-Branch U-Net) may reduce such errors. To their knowledge, this study is the first to evaluate the contribution of the fibrotic microenvironment in deep learning–based lung nodule classification.

(Lin et al., 2024) developed a multi-task learning framework that integrates deep learning, handcrafted radiomics, and clinical variables to support comprehensive lung nodule characterization, including benign–malignant classification, pathological subtype prediction, and Lung-RADS scoring. The study utilized three datasets: LUNA16 (n = 1004) for benign/malignant classification, the Lung Nodule Received Operation (LNOP) cohort (n = 1027) for pathological subtype prediction, and the Lung Nodule in Health Examination (LNHE) dataset (n = 1525) for Lung-RADS categorization. A stacked ensemble based on an AutoGluon-

Tabular classifier combined features extracted from a modified 3D CNN architecture (NASLung), radiomics features generated using PyRadiomics, and structured clinical information such as patient demographics and nodule characteristics. Image preprocessing included voxel resampling, Hounsfield unit normalization, and cropping to standardize input volumes. Three-dimensional tumor ROIs were manually segmented slice-by-slice by experienced thoracic specialists using a Python-based GUI, applying a lung window to ensure precise delineation of nodule boundaries while excluding mediastinal and chest wall structures. Model performance was evaluated using ten-fold cross-validation across the three tasks. The integrated model achieved 92.8% accuracy for benign–malignant classification, an F1-score of 75.5% for pathological subtype prediction, and an F1-score of 80.4% for Lung-RADS classification. Results showed higher accuracy for clearly defined categories such as Lung-RADS 2 and 4B/4X, whereas intermediate categories (Lung-RADS 3 and 4A) were more challenging, reflecting their inherent ambiguity. The model demonstrated strong discriminative ability for severe categories, achieving 93.3% accuracy for the 4B/4X group.

The authors identified several limitations. Manual segmentation was required to extract radiomics features, limiting scalability and introducing operator dependency. Additionally, the model could not incorporate temporal assessment of nodule growth an essential component of Lung-RADS categorization because longitudinal CT scans were unavailable. They emphasized that future research should integrate automatic or semi-automatic segmentation tools and temporal imaging analysis to more accurately model growth patterns and improve Lung-RADS risk stratification

(K. Wang et al., 2025) proposed an integrated predictive framework that combines radiomics and 2.5D deep transfer learning (DTL) to distinguish benign from malignant Lung-RADS 3 and 4A nodules. Their multicenter retrospective cohort included 298 patients from three institutions. Dataset allocation consisted of a training set (n=172) and test set (n=75) from Center 1, with an independent validation set from Centers 2 and 3 (n=51). Clinical variables such as age, gender, smoking status, and nodule characteristics were also included to enhance diagnostic modeling.

The authors developed three models: a radiomics-only model (Rad), a DTL model, and a fused deep transfer radiomics (DTR) model. Radiomics features comprising shape descriptors, first-order intensity metrics, and texture features were extracted from manually segmented ROIs using PyRadiomics. Reproducibility was assessed through intraclass correlation coefficients (ICC), retaining features with ICC ≥ 0.75. Highly correlated features (Pearson $r > 0.9$) were removed to reduce redundancy. Least Absolute Shrinkage and Selection Operator (LASSO) regression with 10-fold cross-validation was then applied for final feature selection and to minimize overfitting. For DTL, the slice with the largest ROI cross-sectional area was selected, together with ±1 and ±2 adjacent slices, producing a 2.5D five-slice representation that preserved spatial context. Deep features were extracted from the penultimate layer of the optimized network, selected using cross-entropy loss and Grad-CAM visualization. Principal

component analysis (PCA) was used to reduce dimensionality and enhance feature compactness. Classification across all models was performed using Support Vector Machines. For the integrated DTR approach, deep and radiomics features were fused using a pre-fusion strategy. Hyperparameters for both LASSO and SVM were optimized using cross-validation.

The DTR model demonstrated the best performance, achieving AUCs of 0.975 (training), 0.851 (testing), and 0.727 (external validation), and showing strong specificity (0.895) and a high positive predictive value (0.956). These findings suggest that combining radiomics with deep semantic features can enhance risk stratification for indeterminate Lung-RADS categories and potentially reduce unnecessary follow-up interventions. However, the study is limited by its modest sample size, restricted focus on Lung-RADS 3 and 4A nodules, and the inherent selection bias of a retrospective design. Broader validation across diverse populations and nodule categories is needed to confirm the model's clinical utility.

(Jeong et al., 2024) analyzed a dataset of 970 chest CT scans collected at the Kangbuk Samsung Medical Center (KBSMC) between 2017 and 2019, of which 832 were reviewed by radiologists 500 with pulmonary nodules and 332 without. All nodule annotations were performed independently by three radiologists. The dataset, retrieved from the PACS system, also included scanner and demographic metadata. Their study investigated the effectiveness of a deep learning–based slice thickness reduction (STR) technique, designed to reconstruct 1-mm thin-slice images from 5-mm scans to enhance CAD performance. Using a retrospective dataset of 687 chest CT scans, the authors applied VUNO Med LungCT-AI, a CNN-based super-resolution model trained on paired thick- and thin-slice CT images. The training used MSE loss, the ADAM optimizer (learning rate = 0.0001), and augmentation techniques such as rotation and flipping. Since only 5-mm images were available, the 5-mm STR model was exclusively used. The STR technique demonstrated meaningful gains, improving CAD AUC from 0.867 to 0.902 and increasing nodule-level sensitivity from 0.826 to 0.916, particularly for small nodules. It also improved differentiation of ground-glass versus part-solid nodules, enhancing diagnostic precision. However, as the authors acknowledge, the single-center design limits generalizability, and the absence of true 1-mm reference scans prevents fully independent validation. As a result, qualitative assessments were based on synthetic 1-mm reconstructions rather than true ground truth. Future studies should incorporate multi-center datasets and real thin-slice references to more rigorously evaluate STR performance.

(Mahmoud et al., 2025) presented a recent evaluation of lightweight deep learning architectures for lung cancer classification in CT images, aiming to establish efficient models suitable for deployment in resource-limited clinical environments. The study compared three pretrained lightweight networks, MobileOne-S0, FastViT-S12, and MambaOut-Femto using two heterogeneous datasets that differed in scanner vendors, reconstruction kernels, and institutional protocols. The private dataset included 274 cases (936 images) acquired on Siemens Definition

AS+ scanners, while the public Jian dataset (Jian et al., 2024)contained 95 cases (308 images) reconstructed using GE, Siemens, and UIH systems with high-frequency kernels (B60f, B70f).

Model performance was optimized using a held-out 70/15/15 split to select the best hyperparameter configuration, which was then applied across stratified 5-fold cross-validation to assess robustness. Training incorporated standard augmentation and early-stopping strategies, and evaluation metrics included AUC, accuracy, recall, specificity, and class-wise performance. Across both datasets, MambaOut-Femto achieved the strongest balance of accuracy, speed, and memory efficiency, outperforming the other lightweight models and demonstrating clear suitability for low-resource clinical settings. The authors noted limitations such as small dataset size, slice-level partitioning effects, and limited generalizability, underscoring the need for larger multi-institutional validation and more quantitative interpretability assessments.

A recent study (J. Liu et al., 2024) developed and validated an interpretable radiomics-based model to distinguish likely benign from likely malignant pulmonary nodules using synthetic LDCT images generated from standard-dose CT (SDCT) scans. The study demonstrated that synthetic LDCT images are both feasible and effective for building a relatively accurate radiomics model. This approach helps address the limited availability of annotated LDCT datasets and may contribute to improved lung cancer detection while reducing false-positive rates. Using degraded SDCT images also provides a practical solution to data scarcity, enabling training on both SDCT and synthetic LDCT images and supporting the development of more generalizable models. The findings further highlight the important contribution of shape and size features to improving classification performance. To generate synthetic LDCT images, SDCT scans were degraded using a sinogram-domain low-dose simulation technique described by (Zeng et al., 2015). The dataset was divided into training, validation, and testing sets. Radiomic features were extracted using PyRadiomics, and three feature sets were evaluated: (1) shape and size (SS) features only, (2) all features excluding SS features, and (3) all features. Feature selection was applied to each subset, and multiple machine-learning models were trained and validated. SHAP (Shapley Additive exPlanations) analysis was used to interpret the final model predictions(Lundberg et al., n.d.). The results showed that synthetic LDCT images can be used effectively for radiomics-based lung nodule classification. A logistic regression model using only three shape and size features achieved high performance, reinforcing the relevance of these features within the dataset. The use of degraded SDCT images provided a practical solution to the limited availability of annotated LDCT scans.

The study has several limitations. The dataset was not specifically designed for nodule classification or screening, and the work remains preliminary, requiring further validation. Due to the absence of paired LDCT–SDCT data, the degradation method could not be fully validated. External validation is also needed to confirm the findings. Although the analysis was based on the large, multi-center LIDC-IDRI dataset (Armato, McLennan, Bidaut, McNitt-Gray, et al., 2011) and therefore relatively robust, additional testing on independent datasets would

strengthen the conclusions. Future work should include validating the model on external datasets and training deep learning models on degraded SDCT images to compare their performance with other state-of-the-art approaches.

3.7 Analysis of Literature Review

This section is highlighted every study contribution and its limitation and how our study fills theses gaps.

(J. Gong et al., 2018) proposed a CADe system for pulmonary nodule detection combining 3D tensor filtering with machine-learning–based false-positive reduction, evaluated on LUNA16(Setio et al., 2017a) and ANODE09(B. van Ginneken et al., 2010). However, the study focused solely on detection, relied on heterogeneous datasets without dose-aware analysis, used limited external validation, and lacked clinical or standardized risk stratification.

(L. Gong et al., 2019) developed a high-performing 3D CNN–based CADe framework using U-Net and SE-ResNet architectures for pulmonary nodule detection, achieving strong sensitivity on LUNA16. Nevertheless, evaluation was restricted to internal cross-validation, without external or clinical validation, dose-aware modeling, or clinically standardized risk classification.

(S. Zhang et al., 2019a) proposed a transfer learning–based LeNet-5 CNN for pulmonary nodule classification using LIDC-IDRI(Armato, McLennan, Bidaut, McNitt-Gray, et al., 2011). However, evaluation relied on internal cross-validation only, used coarse image down-sampling that may degrade spatial detail, lacked dose-aware analysis and external or clinical validation, and did not align classification with standardized clinical risk systems.

(Kozuka et al., 2020) evaluated a deep learning–based CAD system as an assistive tool for less experienced radiologists and showed improved detection sensitivity and reduced reading time. However, the study lacked histopathological validation, relied on thin-slice (1-mm) images rather than standard clinical protocols, did not assess dose variability, and focused on reader assistance rather than clinically standardized risk stratification or external validation.

(Y. Wang et al., 2020) proposed a multi-path, multi-scale CNN for benign–malignant pulmonary nodule classification, evaluated on the LUNGx Challenge dataset(Armato et al., 2015a). However, the study relied on a small and imbalanced dataset, required manually cropped nodules, used internal cross-validation only, and lacked dose-aware analysis, automated end-to-end workflows, and external or clinical validation.

(Raza et al., 2023) proposed Lung-EffNet, an EfficientNet-based transfer-learning framework for lung CT image classification using the IQ-OTH/NCCD dataset(alyasriy, 2023). However, the study relied on a small single-source dataset with slice-level labels, used internal train–test splitting only, lacked dose-aware analysis, clinical (biopsy) validation, and alignment with standardized clinical risk stratification frameworks.

(Saha et al., 2024b) proposed VER-Net an ensemble transfer-learning framework for multiclass lung cancer classification using a Kaggle CT dataset, (Saha et al., 2024b) proposed VER-Net.

However, the study relied on a single non-clinical dataset with aggressive augmentation, used internal validation only, lacked dose-aware analysis and biopsy-confirmed outcomes, and did not align predictions with standardized clinical risk frameworks.

(Kumaran S et al., 2024b) proposed a multi-CNN feature fusion framework for lung cancer classification using the IQ-OTH/NCCD dataset. However, the study relied on a single-source dataset with internal validation only, lacked dose-aware analysis and biopsy-confirmed outcomes, and did not align predictions with standardized clinical risk stratification, limiting generalizability to real-world clinical settings.

An ensemble CNN framework for lung CT classification (Majumder et al., 2024) proposed MENet, which is evaluated on IQ-OTH/NCCD and LIDC-IDRI. However, the study relied on slice-level datasets with internal validation, lacked dose-aware analysis and biopsy-confirmed outcomes, showed reduced sensitivity for micro-nodules, and did not align predictions with standardized clinical risk stratification.

(Kumar et al., 2024) developed an ensemble deep learning decision-support system for multiclass lung cancer prediction using LIDC-IDRI. However, the study relied on a limited single-source dataset with internal validation only, lacked dose-aware analysis and biopsy-confirmed outcomes, and did not align predictions with standardized clinical risk stratification, restricting real-world generalizability.

(Klangbunrueang et al., 2025) evaluated a VGG16-based transfer-learning model for three-class lung cancer classification using 2D CT images. However, the study relied on a small single-source dataset with a simple holdout split, lacked cross-validation, dose-aware analysis, 3D contextual modeling, and biopsy confirmed or standardized clinical risk stratification, limiting generalizability.

(UrRehman et al., 2024) proposed a dual-attention CNN for pulmonary nodule classification, evaluated on LUNA16. However, the study relied on a single benchmark dataset with internal evaluation only, lacked dose-aware analysis, external or clinical validation, and standardized risk stratification, while high computational demands limit practical clinical deployment.

(Canayaz et al., 2024) proposed a combined segmentation classification deep learning framework for pulmonary nodule analysis using a small clinical CT cohort. However, the study relied on limited single center data, lacked dose-aware analysis, external or biopsy confirmed validation, and did not align classification outputs with standardized clinical risk stratification, constraining generalizability and clinical deployment.

(Li et al., 2023) proposed a multimodal graph based deep learning framework combining CNN features, radiomics, and semantic clinical descriptors for multiclass lung adenocarcinoma classification. However, the approach relied on manually annotated semantic features, exhibited

modest classification performance, lacked dose-aware analysis, external or biopsy confirmed validation, and faces scalability challenges for fully automated clinical deployment.

(Qiu et al., 2022) proposed a longitudinal deep learning model combining sequential CT imaging and clinical predictors for classifying pulmonary ground-glass nodules. However, the study was limited by a single center retrospective design, internal validation only, lack of dose-aware analysis and biopsy-confirmed outcomes, and absence of standardized clinical risk stratification, restricting generalizability.

(Y. Liu et al., 2024) investigated incorporating fibrotic microenvironment features into deep learning–based lung nodule malignancy classification using LIDC-IDRI and an in-house cohort. However, evaluation relied on internal cross-validation only, involved heuristic label assumptions, lacked dose-aware analysis and standardized risk stratification, and remains sensitive to segmentation errors affecting clinical robustness.

(Lin et al., 2024) proposed a multi-task framework integrating deep learning, radiomics, and clinical variables for lung nodule classification and Lung-RADS scoring using LUNA16 and in-house datasets. However, the approach relied on manual segmentation, internal cross-validation only, lacked dose-aware and longitudinal growth modeling, and showed reduced performance for intermediate Lung-RADS categories, limiting scalability and clinical robustness.

(K. Wang et al., 2025) proposed a radiomics-2.5D deep transfer learning framework for risk stratification of indeterminate Lung-RADS 3 and 4A nodules using a multicenter retrospective cohort. However, the study relied on manual segmentation, showed declining performance on external validation, lacked dose-aware and longitudinal growth modeling, and was limited to a narrow subset of Lung-RADS categories, restricting broader clinical generalizability.

(Jeong et al., 2024) evaluated a deep learning–based slice thickness reduction (STR) technique to enhance CAD performance on chest CT. However, the study relied on a single-center retrospective cohort, lacked true thin-slice (1-mm) ground truth for independent validation, did not assess dose-aware effects, and used synthetic reconstructions, limiting generalizability and clinical reliability.

(Mahmoud et al., 2025) evaluated lightweight deep learning architectures for lung cancer classification on heterogeneous CT datasets to support deployment in resource-limited settings. However, the study relied on small slice-level datasets, used internal cross-validation only, lacked dose-aware and volumetric modeling, biopsy-confirmed clinical validation, and alignment with standardized clinical risk stratification frameworks, limiting clinical generalizability.

(J. Liu et al., 2024) developed an interpretable radiomics model for pulmonary nodule classification using synthetic LDCT images generated from SDCT scans based on LIDC-IDRI. However, the approach relied on simulated rather than true LDCT data, lacked paired LDCT-

SDCT validation, external clinical testing, dose-aware modeling on real acquisitions, and alignment with standardized clinical risk stratification.

Across the reviewed literature, most computer-aided approaches for lung nodule and lung cancer analysis focus on either nodule detection or binary/multiclass classification, often relying on single public datasets, internal cross-validation, and slice-level or manually cropped inputs. Heterogeneous CT acquisitions are commonly treated as pooled cohorts, without explicit consideration of radiation dose variability, and preprocessing pipelines are typically fixed rather than optimized for differences between low-dose screening CT (LDCT) and standard-dose diagnostic CT (SDCT). Clinical applicability is further constrained by the frequent absence of biopsy-confirmed ground truth, independent external validation, and alignment with standardized clinical reporting frameworks. In addition, many methods depend on manual segmentation, simulated LDCT data, or non-clinical datasets, limiting scalability and real-world generalizability. Even studies incorporating radiomics, ensemble learning, or Lung-RADS scoring often lack dose aware modeling, temporal growth assessment, or robust validation across multi-source clinical data.

To address these gaps, the present study proposes a dose-aware, enhancement-optimized machine-learning framework for lung nodule risk assessment that is explicitly aligned with Lung-RADS v2022. The framework integrates clinically acquired LDCT and SDCT data, applies systematic preprocessing and contrast enhancement optimization tailored to dose conditions, and leverages deep feature extraction combined with classical machine-learning classifiers for robust multi-class risk stratification. Unlike prior work, model performance is evaluated using a multi-level validation strategy, including clinical validation with biopsy-confirmed cases, external validation on an independent hold-out test set, and additional testing on a combined multi-source lung CT dataset, thereby strengthening both generalizability and clinical reliability. Collectively, this study advances existing CADe and CADx systems beyond detection-centric or dataset specific solutions toward a clinically actionable, dose-robust, and Lung-RADS which aligned decision-support framework suitable for real-world deployment.

Chapter 4

Methodology

This chapter outlines the methodology employed in this research, which integrates data curation, preprocessing, machine learning, and evaluation to develop a robust framework for advanced lung cancer detection. The methodology is structured into several key component, each designed to address specific challenges in medical imaging and ensure the reliability and accuracy of the proposed system.

Figure 4.1 provides an overview of the overall methodology, illustrating the workflow from enhancement and preprocessing to model training, validation, and performance evaluation.

4.1 Data Curation

Data curation is central to this study and spans the entire data lifecycle, from ethical approval and secure acquisition to preprocessing and analysis readiness. A structured workflow is used to ensure data quality, integrity, and usability for developing robust and generalizable machine learning models. After obtaining ethical approvals and protecting patient confidentiality, heterogeneous chest CT scans are collected from multiple clinical settings and systematically organized, annotated, and augmented to address class imbalance, protocol variability, and labeling reliability. This process ensures a high-quality dataset that reflects real-world clinical conditions. The following subsections outline the ethical framework, data acquisition, and key data preparation and annotation procedures.

4.2 Ethical statement

This research was conducted under full ethical oversight, with primary approval granted by the Deanship of Scientific Research at Al-Quds University, Abu Dis. The study protocol underwent rigorous evaluation and received formal approval from the university's Institutional Review Board (IRB) Ref No: 433/REC/2024. Prior to initiating data collection, comprehensive ethical clearance was obtained from the administration and ethics committee of each hospital involved.

A robust framework for data protection and patient confidentiality was strictly implemented throughout the study. All patient-identifiable information including names, contact details, medical record numbers, and treatment dates was systematically removed during the data extraction phase. The de-identification process followed standardized protocols to ensure full compliance with institutional regulations and internationally recognized ethical guidelines governing research involving human subjects.

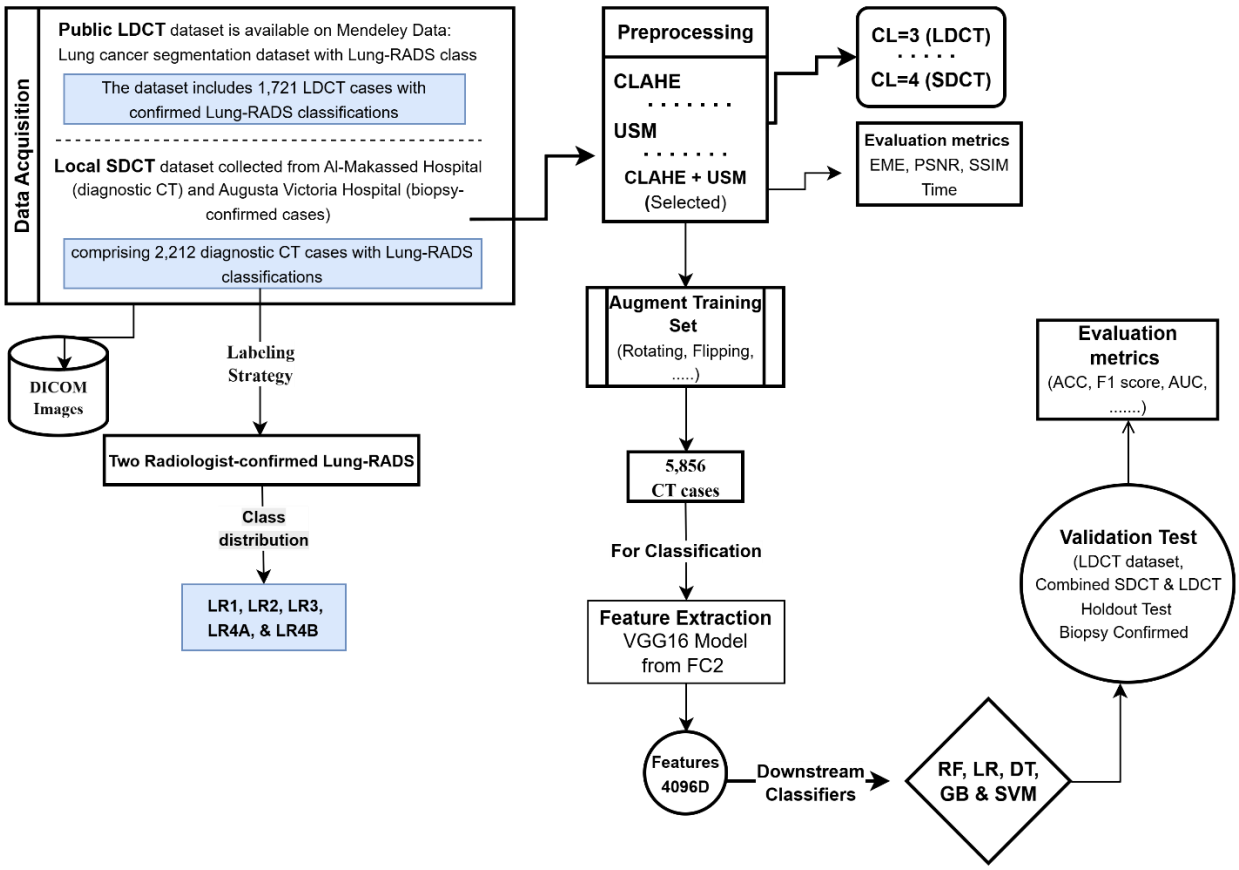


Figure 4.1 Overview of the proposed pipeline for automated lung nodule classification in chest CT

4.3 Data Collection and Labeling

Data were retrospectively collected from two institutions: diagnostic chest CT scans from Al-Makassed Hospital and biopsy-confirmed chest CT cases from Augusta Victoria Hospital. The collection period extended from 2024 to 2025.

The primary dataset consists of approximately 46 helical thoracic CT scans collected from the Picture Archiving and Communication Systems (PACS) of the two participating hospitals. All imaging data were standardized and archived in Digital Imaging and Communications in Medicine (DICOM) format to ensure adherence to medical imaging standards and to support future interoperability. Of these scans, nine included confirmatory biopsy results, which were incorporated into the final clinical validation of the proposed framework.

Standardized imaging protocols were applied to maintain diagnostic relevance. All CT scans were reviewed and adjusted using two clinically established window settings: the mediastinal window (width: 350–400 Hounsfield Units [HU]; level: 40–60 HU) for evaluation of mediastinal structures, and the lung window (width: 1500–1600 HU; level: −400 to −700 HU) for detailed assessment of the lung parenchyma. Image reconstructions were generated with slice thicknesses ranging from 0.9 mm to 3 mm, providing high-resolution cross-sectional visualization suitable for nodule detection and morphological analysis.

The study employed defined inclusion and exclusion criteria to ensure data quality and diagnostic reliability. Eligible cases were required to have a histopathologically confirmed diagnosis and a radiologist-verified chest CT report. Cases were excluded if the patient was younger than 18 years, if the lung-window thin-section slice thickness exceeded 3 mm, or if the CT examination was not fully evaluable. Non-evaluable scans included those with severe motion artifacts, such as patient breathing or movement, or incomplete lung coverage in which part or all of the lung fields could not be adequately visualized.

This study utilized a multimodal imaging dataset that combines two complementary CT acquisition protocols: one for routine diagnostic-dose chest CT scans and another for low-dose CT scans. The primary dataset was collected from Al-Makassed Hospital, comprising a total of 2,212 cases from patients who underwent routine diagnostic-dose chest CT scans during clinical care. These cases were retrospectively classified into Lung-RADS categories based on radiological findings and clinical follow-up. Of these, 674 cases were classified as normal (LR1), 342 as benign (LR2), and 1,196 as malignant, with 515 cases in LR4A and 681 cases in LR4B.

In addition to the primary dataset, the study incorporated the publicly available Lung cancer segmentation dataset with Lung-RADS class (D. Nam, 2024), which contains 1,721 cases with confirmed Lung-RADS classifications. These cases were distributed as follows: 749 cases in

LR1, 107 in LR2, 138 in LR3, 177 in LR4A, and 515 in LR4B. By strategically combining these two datasets, consisting of both standard-dose and low-dose CT imaging, the model is able to learn discriminative features that are robust across different radiation dose regimens, thereby improving its clinical applicability for screening populations.

Two board-certified radiologists with expertise in thoracic imaging independently performed the standardized Lung-RADS categorization for the Al-Makassed cohort to ensure diagnostic consistency. The classification process followed a structured protocol: First, each radiologist independently reviewed the complete imaging history for every patient to assess temporal progression and accurately characterize lesion dynamics. To minimize bias, the radiologists were blinded to all clinical information and evaluated only the imaging data. For each patient, all chest CT scans obtained within the preceding 12-month period were collected whenever available; if no prior scans existed, the available examination was treated as the baseline. This approach ensured consistent follow-up assessment. Lesion growth was identified based on measurable changes across this interval, with particular emphasis on nodules demonstrating progression over shorter timeframes, as such rapid growth is more clinically meaningful for predicting malignancy risk. Second, they identified the precise slice ranges containing suspicious nodules. Third, using the standardized Lung-RADS calculator (McKee et al., 2015), they assigned each case to one of the four risk categories: LR1 (benign findings), LR2 (low probability of malignancy), LR3 (intermediate concern), and LR4 (subdivided into LR4A and LR4B based on probability of malignancy). Following independent labeling, both radiologists participated in a consensus session to discuss discordant cases and achieve a unified classification scheme. This multi-reader approach reduces individual bias and enhances the reliability of the reference standard.

Lung-RADS v2022 (American College of Radiology, 2022; Christensen et al., 2024b) served as the reference clinical standard for label definition, model training, and performance evaluation across all machine-learning classifiers. The American College of Radiology (ACR) developed Lung-RADS as a standardized reporting and management system for lung cancer screening CT(American College of Radiology, 2022). In this study, Lung-RADS categories were employed as a unified risk-stratification labeling framework across both low-dose CT (LDCT) and standard-dose diagnostic CT (SDCT) to ensure consistent and clinically interpretable outputs, while acknowledging that the guideline is primarily intended for screening applications. Furthermore, Lung-RADS version 2022 explicitly recognizes that prior examinations used for category assignment may include either a previous screening CT or a diagnostic CT, supporting the inclusion of diagnostic CT in longitudinal comparison when assessing Lung-RADS categories. Notably, Lung-RADS was applied in this work as a standardized risk-stratification framework rather than as a screening protocol(American College of Radiology, 2022).

To address the limited number of cases, and enhance model generalization, we implemented data augmentation techniques (Z. Zhong et al., 2017) specifically designed for lung CT imaging. Our augmentation pipeline included random rotations (±7°), horizontal flips, and slight brightness adjustments (±10%)(Z. Zhong et al., 2017). These transformations effectively expanded our training dataset to a total of 5,856 CT cases, while maintaining the characteristic appearance of lung lesions in CT images, improving model robustness against variations in scanning angle, tissue orientation, and image acquisition settings.

4.4 Data Partitioning, Experimental Design, and Validation Strategy

To ensure robust model generalization across heterogeneous imaging conditions, the proposed framework firstly was trained using publicly low dose CT(LDCT) dataset(D. Nam, 2024) which is meticulously annotated by expert radiologists according to the Lung-RADS classification system. Significantly the dataset served as the core reference for model initialization and early feature learning. Thereafter a combined multi-source dataset was structured by integrating local standard-dose diagnostic chest CT (SDCT) images and the same publicly available low-dose chest CT (LDCT) images (D. Nam, 2024). This design enables the models to learn feature representations that are robust to variations in radiation dose, scanner type, and acquisition protocol.

To state an unbiased evaluation of real-world clinical data, a separate hold-out test set was reserved exclusively from the Al-Makassed Hospital diagnostic SDCT cohort. This subset consisted of 90 CT images and was completely withheld from all stages of model development, including training, internal testing, cross-validation, preprocessing optimization, and data augmentation. The hold-out dataset was used solely for final performance evaluation, thereby providing an independent assessment on real world local clinical data.

The remaining CT images drawn from both the local diagnostic SDCT data and public LDCT (D. Nam, 2024) datasets that were merged to form the development cohort. This combined dataset was subsequently divided into training and internal testing subsets using an 80:20 split, where 80% of the data was allocated for model training and 20% for internal testing. Data partitioning was performed at the patient level (case based) to prevent overlap between subsets, to preserve data independence and represents the real clinical data use. Significantly, case-based data partitioning helps prevent overfitting that reflecting true feature generalization not just memorizing, doesn’t exploit the intra patient similar anatomy.

Following data splitting, data augmentation was applied exclusively to the training subset to address the limited number of cases and enhance model generalization. Augmentation strategies specifically designed for lung CT imaging (Z. Zhong et al., 2017) including random rotations (±7°), horizontal flipping, and slight brightness adjustments (±10%). These transformations

effectively expanded our training dataset to a total of 5,856 CT cases, while maintaining the characteristic appearance of lung lesions in CT images, improving model robustness against variations in scanning angle, tissue orientation, and image acquisition settings .

This hierarchical data-splitting strategy is comprising an external local hold-out test set, a combined SDCT & LDCT development cohort, and a training only augmentation scheme, which ensures fair evaluation, prevents information leakage, and supports the development of dose-aware and clinically generalizable machine-learning models.

4.5 Clinical CT Acquisition Settings and Scanner Variability

All CT scans used in this study were obtained retrospectively from routine clinical practice; no examinations were acquired specifically for research purposes. As a result, the dataset naturally includes variability in scanner models, acquisition parameters, and reconstruction settings. This heterogeneity was intentionally preserved, as it reflects real-world diagnostic conditions and strengthens the generalizability of the proposed framework.

Chest CT examinations were collected from two major Palestinian tertiary hospitals: Al-Makassed Hospital and Augusta Victoria Hospital. All scans correspond to standard-dose diagnostic CT (SDCT) protocols. Low-dose lung cancer screening is not currently implemented in Palestine; therefore, the dataset represents routine diagnostic and oncologic imaging rather than screening-optimized acquisitions.

At Al-Makassed Hospital, chest CT scans were acquired using a Philips Incisive 128-slice CT scanner with automatic exposure control (DoseRight Smart mA), which dynamically adjusts tube current based on patient size and attenuation. HRCT examinations were performed at 100–120 kVp, with dose modulation resulting in variable mAs across slices. Typical acquisition settings included thin collimation, low pitch, and short rotation times, with radiation dose levels consistent with standard diagnostic chest CT. Reconstructions were tailored to the clinical task, using thicker slices for mediastinal evaluation and thin, overlapping slices for lung parenchyma assessment.

In addition to dedicated chest CT, combined chest–abdomen studies were routinely performed using standard diagnostic protocols, with lung-specific reconstructions generated when clinically indicated. These protocols reflect common thoracoabdominal imaging practice in oncologic care.

Biopsy-confirmed malignant cases were obtained from Augusta Victoria Hospital and acquired using a Siemens SOMATOM go.Top 128-slice CT scanner under a contrast-enhanced chest–abdomen–pelvis oncology protocol. Scans were performed at 120 kVp with moderate mAs

settings and reconstructed using overlapping axial slices and a medium-soft kernel, consistent with routine clinical oncology imaging.

The use of two different CT platforms introduces realistic variability in image noise, contrast, and reconstruction characteristics. Rather than being a limitation, this variability enhances the external validity of the study by exposing the model to multi-center, real-world imaging conditions. It also motivated the use of robust preprocessing strategies, including dose-aware normalization and contrast-limited adaptive histogram equalization (CLAHE), to ensure stable feature representation across scanners and acquisition protocols.

4.6 Image Preprocessing and Enhancement

Given the inherent challenges of low-dose CT imaging, which suffers from increased quantum noise and reduced contrast-to-noise ratio compared to standard-dose acquisitions, a preprocessing strategy was implemented to optimize image quality across both acquisition protocols. We systematically evaluated five preprocessing techniques: median filtering for noise reduction, histogram equalization for global contrast enhancement, unsharp masking for edge enhancement, contrast-limited adaptive histogram equalization (CLAHE), and a combined CLAHE with unsharp masking approach. Recent literature demonstrates that CLAHE followed by unsharp masking provides superior results compared to CLAHE alone for medical image analysis. Accordingly, this combined approach was selected for subsequent processing (Al-Areqi et al., 023).

CLAHE was carefully calibrated separately for standard-dose and low-dose images to address their distinct signal characteristics. For low-dose acquisitions, a higher clip limit was applied to aggressively enhance contrast in low-contrast regions while recovering subtle features necessary for malignancy discrimination. For standard-dose images, a more conservative clip limit was employed to increase conspicuity of clinically relevant structures while avoiding introduction of artifactual noise amplification that could degrade model performance. Following CLAHE enhancement, unsharp masking was applied to accentuate nodule boundaries and enhance textural details. This adaptive preprocessing strategy ensures that the feature extraction process captures clinically meaningful patterns from both imaging modalities while minimizing dose-related artifacts and noise-induced degradation.

To capture fine-grained spatial information, we included images reconstructed with thin-slice protocols (slice thickness ≤ 1.5 mm), which enhance the spatial resolution of small nodules and subtle texture variations critical for early-stage cancer recognition and precise three-dimensional morphological assessment (Jeong et al., 2024)

4.7 Feature Extraction Using Transfer Learning

To overcome the limitations of hand-crafted features and the computational burden of training deep networks from scratch on limited medical imaging datasets, we employed transfer learning with VGG16 as the feature extraction backbone (Simonyan & Zisserman, 2015). VGG16, a sixteen-layer convolutional neural network pre-trained on the ImageNet dataset, has demonstrated superior performance for lung imaging applications compared to contemporary architectures including VGG19, ResNet50, and DenseNet201 (Al-Areqi, Celebi, et al., 2023). This architecture excels at learning hierarchical spatial features through its sequence of 3×3 convolutional filters interspersed with max-pooling operations, enabling efficient extraction of multi-scale textural and morphological information from CT images. The model's relatively modest computational requirements, combined with its proven effectiveness in medical imaging domains, make it particularly suitable for this application.

For feature extraction, the pre-trained VGG16 network was utilized without fine-tuning; the convolutional layers retained their ImageNet-learned weights to preserve generalization capacity, and features were extracted from the penultimate fully connected layer (FC2), which has dimensionality of 4096. This approach yields a rich, abstract representation of image content that captures both low-level pixel-scale details (edges, textures) and high-level semantic patterns (morphological shapes, spatial relationships). By leveraging this pre-trained feature space, we avoid the data scarcity problem that often plagues direct application of deep networks to medical imaging tasks, while benefiting from representations optimized on millions of natural images that share fundamental visual principles with medical imaging (H. E. Kim et al., 2022).

4.8 Machine Learning Classification

In this study, pulmonary nodules were categorized according to the American College of Radiology Lung Imaging Reporting and Data System (Lung-RADS), version 2022(American College of Radiology, 2022; Christensen et al., 2024b), which served as the clinical reference standard for all machine learning experiments. The Lung-RADS system provides a structured and standardized framework for lung cancer risk stratification in CT imaging, assigning nodules to distinct categories that rely on nodule size, composition such as solid, part-solid, or ground-glass, and associated imaging features. Each category reflects an estimated malignancy risk and corresponding clinical management recommendation.

In this work, the radiologists' annotated CT images were reviewed and labeled in accordance with Lung-RADS v2022 guidelines. Nodules were assigned to one of the following categories: LR1 (negative), LR2 (benign appearance), LR3 (probably benign), LR4A (suspicious), and LR4B (very suspicious). These Lung-RADS categories were subsequently used as ground-truth class labels for training and evaluating the machine-learning classifiers. As a result, the developed models were explicitly designed to learn discriminative imaging patterns that correspond to clinically meaningful Lung-RADS risk levels, rather than arbitrary or purely data-driven class definitions.

By adopting Lung-RADS as the reference framework, the proposed machine-learning system ensures direct clinical interpretability, alignment with radiologist workflow, and applicability to real-world lung cancer screening and diagnostic settings.

In this study, pulmonary nodules were categorized according to the American College of Radiology Lung Imaging Reporting and Data System (Lung-RADS), version 2022(American College of Radiology, 2022; Christensen et al., 2024b), which served as the clinical reference standard for all machine learning experiments. The Lung-RADS system provides a structured and standardized framework for lung cancer risk stratification in CT imaging, assigning nodules to distinct categories that rely on nodule size, composition (solid, part-solid, or ground-glass), and associated imaging features. Each category reflects an estimated malignancy risk and corresponding clinical management recommendation.

Following feature extraction via VGG16, the resulting 4096-dimensional feature vectors were used to train Five distinct classical machine learning classifiers to comprehensively evaluate the discriminative capacity of the learned representations. The classifiers employed were logistic regression (LR)(Lou, 2024), random forest (RF)(Belgiu & Drăgu, 2016), gradient boosting (GB)(Lou, 2024), Support vector machine (SVM)(Sheth et al., 2022) and decision trees (DT)(Lou, 2024). Each algorithm operates on fundamentally different principles: logistic regression assumes linear separability in the feature space and provides probabilistic outputs; random forests aggregate predictions from multiple decision trees with bootstrap sampling to reduce overfitting and improve generalization; gradient boosting iteratively corrects residuals from sequential decision trees through gradient descent optimization; and decision trees partition the feature space hierarchically based on feature importance. By assessing performance across this diversity of algorithms, we can identify which learning paradigm best captures the relationship between extracted features and diagnostic categories, while also providing evidence that the feature representation itself drives model performance, rather than a specific classifier.

4.9 Model Validation and Testing

To ensure robust assessment of model generalization, we implemented a 10-fold cross-validation strategy applied to the combined dataset of standard-dose and low-dose CT images. This resampling approach partitions the data into ten equal-sized folds, iteratively trains the model on nine folds while evaluating on the held-out fold, and aggregates performance metrics across all iterations. Cross-validation provides robust estimates of model performance while maximizing the utilization of limited medical imaging data. Reported metrics represent the mean performance across all ten folds, establishing confidence in the model's ability to generalize to unseen data.

Additionally, a separate hold-out test set was reserved from the Al-Makassed Hospital cohort (withheld from all training and cross-validation procedures) to provide an independent external validation assessment on prospective institutional data. This held-out evaluation specifically assesses model performance on standard-dose diagnostic CT, quantifying how well features learned from the combined dataset translate to the original clinical acquisition protocol.

To further validate clinical applicability, a supplementary cohort of biopsy-confirmed malignant cases with Lung-RADS category LR4 designations was collected. These histologically verified positive cases provide high-confidence ground truth for assessing the model's ability to correctly identify truly malignant lesions, which is the most critical clinical objective for any lung cancer detection system.

4.10 Performance Evaluation Metrics

Model performance and image quality were assessed using a comprehensive set of quantitative metrics. For image quality assessment, three key measures were applied: Enhancement Measure Estimation (EME), Peak Signal-to-Noise Ratio (PSNR), and Structural Similarity Index (SSIM). The EME metric quantifies local contrast and brightness variations to evaluate the strength of enhancement, while PSNR measures the fidelity of the enhanced image relative to the original, with higher values indicating lower distortion. The SSIM metric assesses the perceptual similarity between enhanced and reference images, emphasizing the preservation of structural information and visual quality.

For model performance evaluation, standard classification metrics were used, including accuracy (ACC), positive predictive value (PPV), true positive rate (TPR or sensitivity), F1-score, and the area under the receiver operating characteristic curve (AUC). Accuracy represents the proportion of correctly classified samples, PPV reflects the fraction of correctly identified positive cases among all predicted positives, and TPR measures the ability of the model to correctly detect positive instances. The F1-score provides a balanced assessment of precision and recall, while the AUC offers a threshold-independent measure of discriminative performance and is particularly relevant in medical applications involving class imbalance. In addition, the training time was recorded to evaluate the computational efficiency of each classifier.

Chapter 5

Results

This chapter presents the study results of the proposed lung-nodule analysis framework, organized around its core components: contrast enhancement, deep feature–based classification, and multi-dataset validation. First, the performance of the enhancement methods (HE, USM, CLAHE, and CLAHE–USM) is evaluated using quantitative image-quality metrics, EME, PSNR, SSIM, and processing time, supplemented by visual comparisons across Lung-RADS categories. Second, the classification results of various machine-learning models are reported for LDCT, combined LDCT–SDCT, external institutional, and biopsy-confirmed clinical datasets using standard performance metrics. Finally, representative longitudinal cases are included to demonstrate temporal changes in nodule appearance and support the clinical interpretation of the findings.

5.3 Contrast Enhancement Evaluation

This section evaluates the applied contrast-enhancement techniques within the proposed dose-aware lung-nodule analysis framework.

5.3.1 Quantitative Image Quality Assessment of Enhancement Methods

The evaluation of contrast-enhancement techniques on chest CT images demonstrated clear performance variations across multiple quantitative metrics. **Table 5.1** provides a detailed comparison of Histogram Equalization (HE), Unsharp Masking (USM), Contrast-Limited Adaptive Histogram Equalization (CLAHE), and a hybrid CLAHE–USM method. The assessment includes Enhancement Measure Estimation (EME), Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index (SSIM), and processing time.

Table 5.1 Image Quality Assessment (IQA) of Different Contrast Enhancement Methods Across a Subset of 120 Malignant LDCT Lung Cancer Cases

IQA Metric	Original	HE	USM	CLAHE	CLAHE-USM
EME	10.543	17.862	14.095	17.214	20.648
PSNR	∞	20.898	29.415	21.596	19.711
SSIM	1.000	0.942	0.992	0.940	0.912
Time (ms)	0.000	0.511	0.880	0.519	0.601

- ***HE**: Histogram Equalization; **USM**: Unsharp Masking*

Results show that the hybrid CLAHE–USM approach produced the highest average EME value (20.648), indicating superior local contrast enhancement and improved visibility of malignant pulmonary nodules, especially in LDCT images, where subtle lesions require stronger contrast amplification. HE (17.862) and CLAHE (17.214) also outperformed the original CT image (EME = 10.543), demonstrating their effectiveness in enhancing parenchymal contrast. In comparison, USM achieved a lower EME value (14.095), reflecting more conservative enhancement but reduced noise amplification. From a noise-preservation perspective, USM achieved the highest PSNR (29.415), indicating the best maintenance of original image details with minimal noise. CLAHE (21.596) and HE (20.898) demonstrated moderate PSNR values, while the hybrid CLAHE–USM method produced the lowest PSNR (19.711) due to its aggressive local enhancement, which typically increases noise levels. A similar pattern was observed for SSIM: USM achieved the highest structural similarity (SSIM = 0.992), followed by HE (0.942) and CLAHE (0.940). The hybrid method showed the lowest SSIM (0.912), consistent with its stronger intensity redistribution that alters global structure despite improving lesion conspicuity. In terms of computational efficiency, HE (0.511 ms) and CLAHE (0.519 ms) were the fastest techniques. The hybrid CLAHE–USM method required slightly more processing time (0.601 ms), while USM was the slowest (0.880 ms) due to its convolution-based sharpening operations.

5.3.2 Visual Comparison Across Lung-RADS Categories

To complement the quantitative analysis, visual comparisons were performed on representative LDCT cases spanning different Lung-RADS categories.

Similarly Figure 5.1 & 5.2 presents a visual comparison of the same anatomical region processed with the Original, HE, USM, CLAHE, and CLAHE–USM techniques. This comparison demonstrates how each enhancement method reveals nodule details across different Lung-RADS malignancy risk levels (LR2, and LR4B), highlighting differences in contrast, boundary sharpness, and textural clarity. Corresponding EME scores are included to quantitatively support the visual improvements, demonstrating the superior contrast enhancement achieved by CLAHE and USM-based approaches while preserving structural integrity.

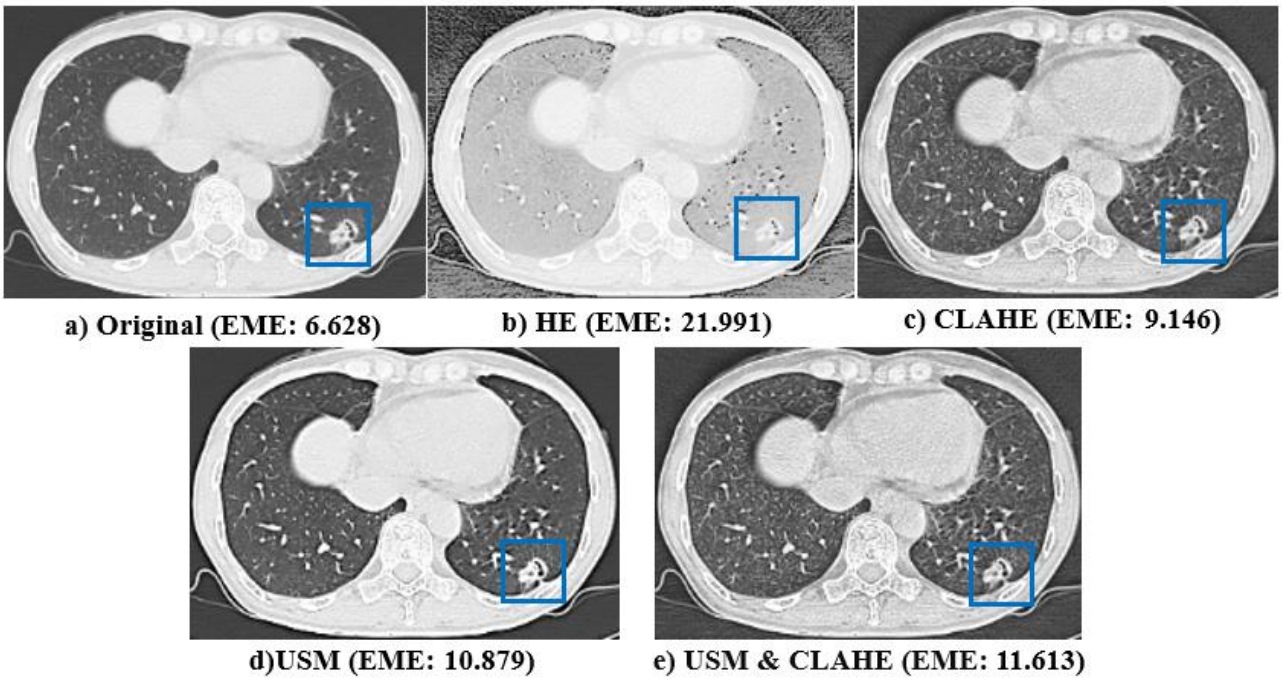


Figure 5.1 Visual Comparison of Enhancement Methods Across Lung-RAD 4B

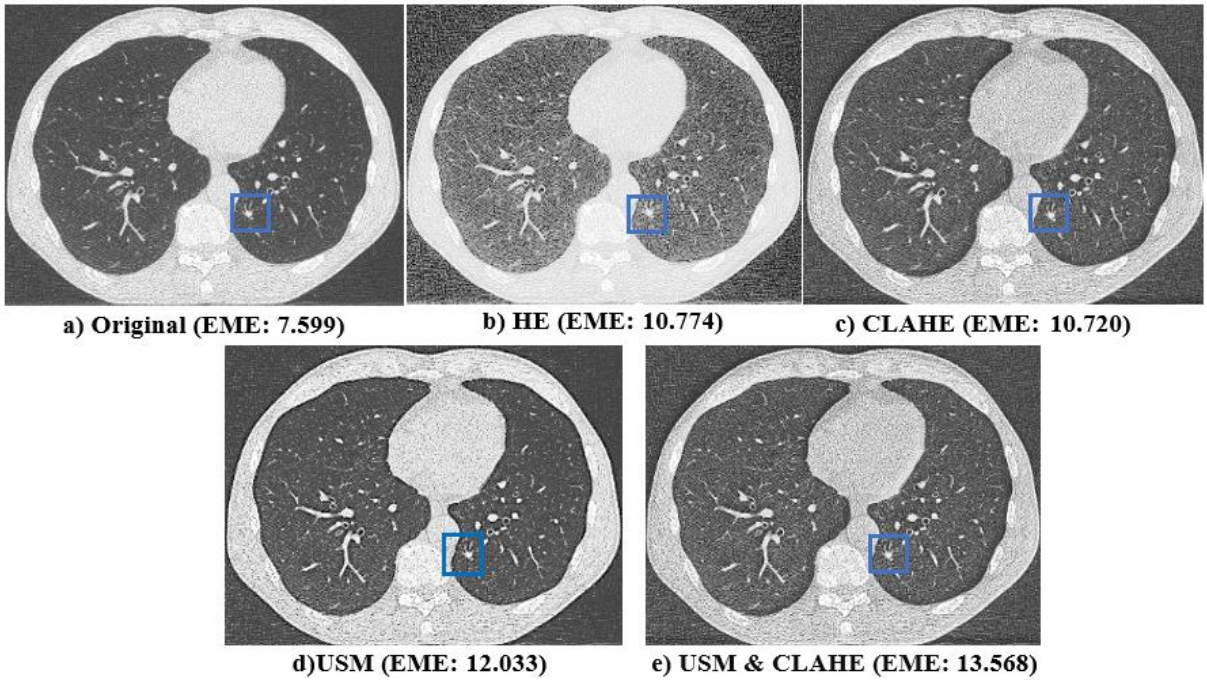


Figure 5.2 Visual Comparison of Enhancement Methods Across Lung-RAD 2

The visual results are consistent with the quantitative metrics, demonstrating how enhancement choices influence lesion conspicuity and structural appearance across risk categories.

5.3.3 Optimization of CLAHE Parameters Across Dose Levels

Table 5.2 presents the quantitative evaluation of CLAHE across different clip limits for both standard-dose and low-dose CT subsets. The table reports EME, PSNR, and SSIM values for each configuration, enabling direct comparison of how varying the clip limit affects contrast, noise visibility, and structural similarity. For the LDCT subset, a clip limit of 3 yielded the most balanced combination of EME (19.485), PSNR (20.896), and SSIM (0.929). For the standard-dose CT subset, a clip limit of 4 produced the highest EME value (20.648), with corresponding PSNR and SSIM values of 19.711 and 0.912, respectively.

Table 5.2 (a): Optimization of the CLAHE Clipping Parameter Across Subsets of 120 Malignant Cases from the LDCT and 120 cases from Al-Makassed Lung Cancer Datasets.

CT Type	Metric	CL=0.5	CL=1	CL=2	CL=3	CL=4
Regular	EME	14.280	14.592	15.416	16.332	20.648
	PSNR	39.581	36.261	31.368	28.049	19.711
	SSIM	0.998	0.997	0.994	0.988	0.912

Table 5.2 (b): Optimization of the CLAHE Clipping Parameter Across Subsets of 120 Malignant Cases from the LDCT and 120 cases from Al-Makassed Lung Cancer Datasets.

CT Type	Metric	CL=0.5	CL=1	CL=2	CL=3	CL=4
Low-dose	EME	12.952	14.127	16.260	19.485	23.116
	PSNR	33.510	28.365	23.579	20.896	17.878
	SSIM	0.996	0.988	0.961	0.929	0.867

Note: LDCT scans were processed using a grid size of 16×16 and an EME block size of 250, while standard-dose CT scans were processed using a grid size of 4×4 and an EME block size of 100. These settings reflect dataset-specific parameter configurations applied during CLAHE optimization.

Figure 5.3 & 5.4 displays the visual effect of increasing the CLAHE clip limit on both LDCT and standard-dose CT images. Higher clip limits correspond to stronger local contrast enhancement and increased noise visibility.

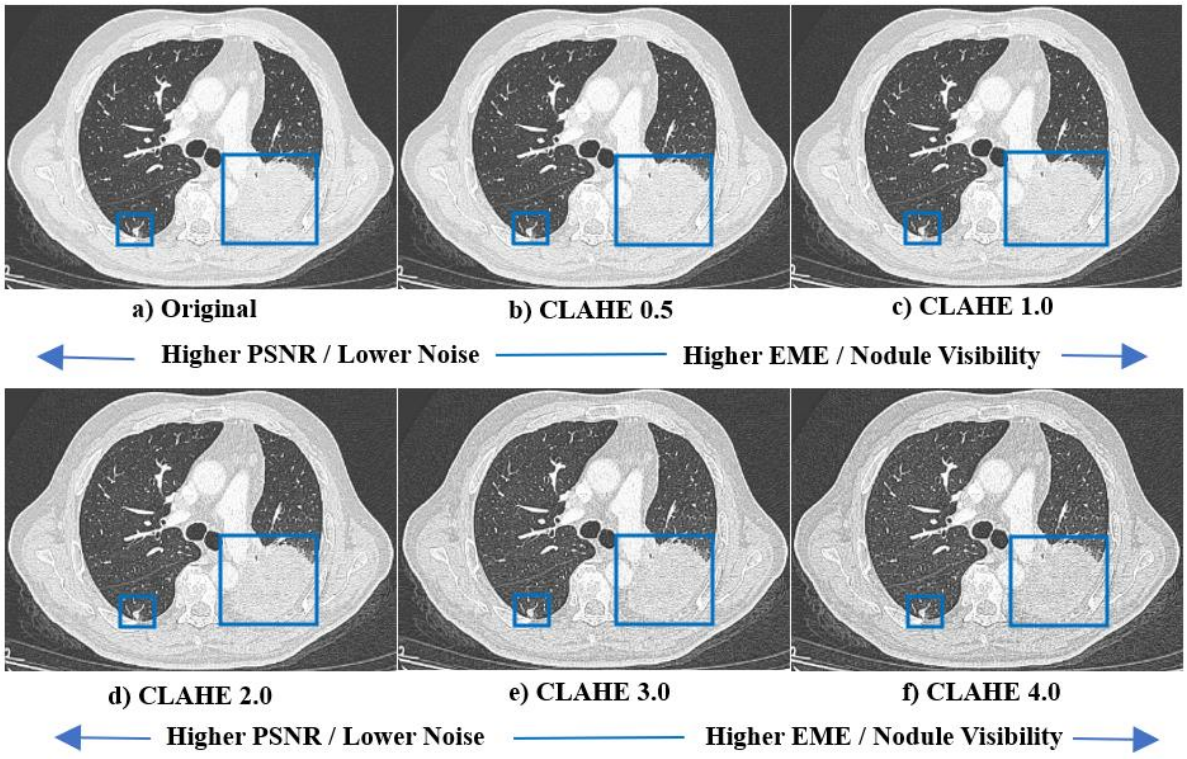


Figure 5.3 Impact of Clip Limits on CLAHE Performance on SDCT images

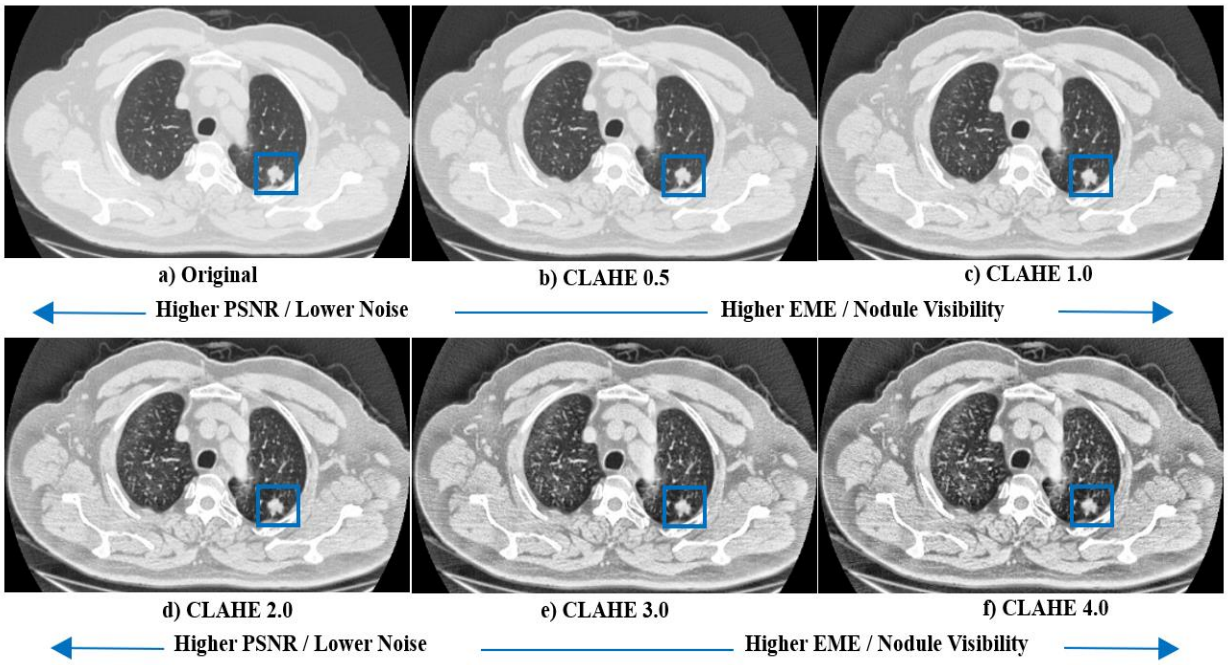


Figure 5.4 Impact of Clip Limits on CLAHE Performance on LDCT images

5.4 Classification Performance Results

Using images processed with the optimized enhancement settings, classification performance was evaluated across multiple Lung-RADS categories and datasets.

5.4.1 LDCT Dataset

Table 5.3 presents the comparative performance of the evaluated classification algorithms across different Lung-RADS stages on the LDCT dataset. The ensemble-based methods consistently achieved higher classification accuracy and F1-scores than linear and margin-based classifiers, indicating better discrimination of lung nodule categories. For the LR2 stage, gradient boosting and decision tree classifiers achieved the highest accuracy and F1-score (ACC = 0.8615, F1 = 0.8790), with gradient boosting incurring substantially higher computational time. Logistic regression and random forest showed moderate accuracy (ACC = 0.7462) but exceptionally high AUC values, suggesting strong ranking capability despite lower class-wise sensitivity. In contrast, support vector machines exhibited poor accuracy and sensitivity across stages, although AUC values remained high, reflecting limited practical utility under the selected decision thresholds.

At the LR3 stage, overall performance declined across all models, reflecting increased classification difficulty. Gradient boosting achieved the best balance between accuracy and F1-score (ACC = 0.7557, F1 = 0.7902), followed closely by the decision tree classifier, while logistic regression and random forest demonstrated comparable but lower performance. Similar trends were observed for the LR4A and LR4B stages, where gradient boosting and decision tree classifiers again delivered the strongest overall performance in terms of accuracy and F1-score, with logistic regression maintaining competitive results and consistently minimal execution time. Random forest models provided stable but slightly inferior performance, whereas support vector machines remained the weakest performers despite relatively high AUC values.

Across all stages, decision tree classifiers offered a favorable trade-off between predictive performance and computational efficiency, achieving accuracy and F1-scores comparable to gradient boosting at a fraction of the computational cost. These findings indicate that tree-based models are particularly well suited for LDCT-based Lung-RADS classification, especially in clinical settings where both accuracy and real-time applicability are critical.

Table 5.3 Comparative Performance of Classification Algorithms on the LDCT Dataset

Stage	Classifier	ACC	PPV	TPR	F1-Score	AUC	Time (S)
LR2	LR	0.7462	0.9120	0.7462	0.7859	0.9982	0.1
	SVM	0.4115	0.8904	0.4115	0.4618	0.9930	0.4
	RF	0.7462	0.9120	0.7462	0.7859	0.9865	1.8
	GB	0.8615	0.9317	0.8615	0.8790	0.9200	18.3
	DT	0.8615	0.9317	0.8615	0.8790	0.9200	0.1
LR3	LR	0.6947	0.8307	0.6947	0.7382	0.8142	0.1
	SVM	0.5038	0.8901	0.5038	0.5611	0.8854	0.3
	RF	0.6947	0.8307	0.6947	0.7382	0.8736	2.1
	GB	0.7557	0.8782	0.7557	0.7902	0.8992	19.5
	DT	0.7519	0.9100	0.7519	0.7890	0.8556	0.1
LR4A	LR	0.8357	0.9105	0.8357	0.8506	0.9950	0.1
	SVM	0.4821	0.8576	0.4821	0.5063	0.9811	0.3
	RF	0.7607	0.8921	0.7607	0.7851	0.9901	1.5
	GB	0.8500	0.8581	0.8500	0.8518	0.9455	22.5
	DT	0.8500	0.8581	0.8500	0.8518	0.9620	0.1
LR4B	LR	0.7906	0.8772	0.7906	0.7996	0.9915	0.1
	SVM	0.6594	0.8414	0.6594	0.6670	0.9818	0.4
	RF	0.7906	0.8772	0.7906	0.7996	0.9942	2.0
	GB	0.7969	0.8794	0.7969	0.8056	0.9899	23.5
	DT	0.7969	0.8794	0.7969	0.8056	0.8556	0.1

- ***ACC**: Accuracy; **PPV**: Precision; **TPR**: Sensitivity; **AUC**: Area Under the Curve*

These results reflect the baseline performance of each classifier under low-dose imaging conditions.

5.4.2 Combined Dataset

Table 5.4 reports the comparative performance of the classification algorithms on the combined lung dataset, which integrates LDCT and standard-dose CT images. The classification performance improved markedly across all Lung-RADS stages compared with the LDCT-only

setting, indicating that the inclusion of heterogeneous imaging data enhanced model robustness and feature discriminability.

For the LR2 stage, all classifiers achieved high accuracy, with logistic regression and decision tree models providing a favorable balance between accuracy and computational efficiency. Logistic regression attained the highest accuracy (ACC = 0.8961) and a strong F1-score, while random forest and gradient boosting delivered comparable but slightly lower performance at higher computational cost. Support vector machines showed competitive accuracy but lower AUC, suggesting reduced ranking stability relative to other methods.

At the LR3 stage, near-ceiling performance was observed for most models. Random forest and support vector machine classifiers achieved the highest accuracy and F1-scores (ACC $\geq$ 0.977), accompanied by excellent AUC values approaching unity, indicating highly reliable discrimination. Logistic regression and gradient boosting also performed strongly, while decision trees maintained high accuracy with negligible execution time, underscoring their practical suitability.

Similar trends were evident for the LR4A and LR4B stages, where support vector machines consistently delivered the best overall performance, achieving near-perfect accuracy, F1-scores, and AUC values. Logistic regression and gradient boosting remained highly competitive, while random forest and decision tree classifiers showed slightly reduced but still strong performance. Across all stages, execution times remained low, with decision trees and support vector machines offering the fastest inference.

Collectively, these results demonstrate that the combined lung dataset substantially improves classification performance and generalization across Lung-RADS categories. The findings also indicate that multiple classifiers can achieve clinically meaningful accuracy, with model selection guided by the desired balance between predictive performance and computational efficiency.

Table 5.4 (a): Comparative Performance of Classification Algorithms on the Combined Lung Dataset

Stage	Classifier	ACC	PPV	TPR	F1-Score	AUC	Time (S)
LR2	LR	0.8961	0.8858	0.8961	0.8826	0.9885	0.1
	SVM	0.8575	0.8795	0.8575	0.8298	0.8285	3.3
	RF	0.8704	0.8853	0.8704	0.8498	0.9225	0.7
	GB	0.8639	0.8802	0.8639	0.8406	0.8901	12.7
	DT	0.8812	0.8848	0.8812	0.8681	0.8510	0.4

Table 5.4 (b): Comparative Performance of Classification Algorithms on the Combined Lung Dataset

Stage	Classifier	ACC	PPV	TPR	F1-Score	AUC	Time (S)
LR3	LR	0.9085	0.9363	0.9085	0.9182	0.9924	3.1
	SVM	0.9773	0.9779	0.9773	0.9759	0.9998	0.2
	RF	0.9798	0.9803	0.9798	0.9787	0.9987	1.7
	GB	0.9572	0.9558	0.9572	0.9530	0.8945	0.1
	DT	0.9467	0.9468	0.9467	0.9967	0.9898	0.0
LR4A	LR	0.9700	0.9721	0.9700	0.9702	0.9989	2.7
	SVM	0.9983	0.9983	0.9983	0.9983	0.9992	0.2
	RF	0.9483	0.9483	0.9483	0.9483	0.9850	2.2
	GB	0.9717	0.9729	0.9717	0.9715	0.9818	0.3
	DT	0.9419	0.9449	0.9419	0.9414	0.9946	0.1
LR4B	LR	0.9435	0.9502	0.9435	0.9439	0.9997	2.3
	SVM	0.9726	0.9735	0.9726	0.9725	0.9992	0.4
	RF	0.9081	0.9196	0.9081	0.9059	0.9912	4.2
	GB	0.9113	0.9221	0.9113	0.9093	0.8946	0.3
	DT	0.8961	0.8858	0.8961	0.8826	0.9885	0.1

- ***ACC**: Accuracy; **PPV**: Precision; **TPR**: Sensitivity; **AUC**: Area Under the Curve*

Overall, combining LDCT and SDCT data resulted in more stable and consistently high performance across Lung-RADS categories.

5.2.3 External Validation

Table 5.5 provides a detailed assessment of external validation performance on an independent hold-out dataset from Al-Makassed Hospital, offering a stringent test of model generalizability under real-world clinical conditions. Across both evaluated Lung-RADS categories, the majority of classifiers preserved high accuracy, precision, and sensitivity, indicating that the learned feature representations and decision boundaries were not overfitted to the development dataset.

For the LR2 category, performance remained consistently strong across models, with accuracy ranging from 0.8419 to 0.8846. The support vector machine achieved the highest overall discrimination, combining the top accuracy (ACC = 0.8846), high precision (PPV = 0.8997), balanced sensitivity (TPR = 0.8846), and the highest AUC (0.9988), indicating excellent class

separability. Random forest closely followed, with comparable accuracy (ACC = 0.8825) and F1-score (0.8666), suggesting robust ensemble learning behavior under domain shift. Logistic regression and gradient boosting yielded slightly lower F1-scores (0.8147 and 0.8114, respectively), reflecting reduced sensitivity to subtle nodular patterns, while decision trees showed the lowest AUC (0.9134) but maintained rapid inference times, highlighting their efficiency-oriented trade-off.

In the LR4B category, all models demonstrated a marked improvement in performance, reflecting clearer radiological distinctions in higher-risk nodules. Random forest and decision tree classifiers achieved the highest accuracy, precision, and F1-scores (ACC = 0.9652, PPV = 0.9676, F1 = 0.9652), indicating strong and consistent identification of high-risk cases. Logistic regression and gradient boosting also performed reliably, both reaching an accuracy of 0.9233 with balanced precision and sensitivity, while maintaining very high AUC values (≥ 0.9775). Although support vector machines showed a relative reduction in accuracy (ACC = 0.8641), their AUC remained high (0.9633), suggesting preserved ranking capability despite suboptimal threshold-level classification.

Across both categories, execution times were minimal for most classifiers, with decision trees and logistic regression offering near-instantaneous inference, whereas gradient boosting incurred substantially higher computational cost without proportional performance gains. These results demonstrate that the proposed framework generalizes well to unseen clinical data, with tree-based ensemble models providing a strong balance between accuracy, robustness, and computational efficiency, reinforcing their suitability for deployment in routine clinical workflows.

Table 5.5 (a): External Validation of the Classification Algorithms on a Hold-Out Test Set from Al-Makassed Hospital

Base Model	Classifier	ACC	PPV	TPR	F1-Score	AUC	Time (S)
LR2	LR	0.8462	0.8718	0.8462	0.8147	0.9799	0.1
	SVM	0.8846	0.8997	0.8846	0.8694	0.9988	1.0
	RF	0.8825	0.8981	0.8825	0.8666	0.9733	0.7
	GB	0.8440	0.8703	0.8440	0.8114	0.9766	12.7
	DT	0.8419	0.8636	0.8419	0.8095	0.9134	0.3

Table 5.5 (b): External Validation of the Classification Algorithms on a Hold-Out Test Set from Al-Makassed Hospital

Base Model	Classifier	ACC	PPV	TPR	F1-Score	AUC	Time (S)
LR4B	LR	0.9233	0.9341	0.9233	0.9232	0.9910	0.0
	SVM	0.8641	0.8946	0.8641	0.8627	0.9633	0.2
	RF	0.9652	0.9676	0.9652	0.9652	0.9896	0.5
	GB	0.9233	0.9341	0.9233	0.9232	0.9775	11.6
	DT	0.9652	0.9676	0.9652	0.9652	0.9671	0.1

- ***ACC**: Accuracy; **PPV**: Precision; **TPR**: Sensitivity; **AUC**: Area Under the Curve*

The external validation confirms that the proposed framework maintains strong performance on unseen clinical data.

5.4.3 Clinical Validation

Table 5.6 shows the clinical validation results obtained on biopsy-confirmed malignant cases from Augusta Victoria Hospital, providing a high-confidence evaluation of model performance under definitive diagnostic conditions. Across all classifiers, performance remained consistently high, indicating reliable discrimination of malignant nodules within the LRAD 4B category.

Support vector machines achieved the strongest overall performance, attaining the highest accuracy (ACC = 0.9333), precision (PPV = 0.9412), F1-score (0.9330), and an AUC approaching unity (0.9996), reflecting excellent sensitivity to malignant patterns and robust class separability. Logistic regression also demonstrated strong performance, with high accuracy (ACC = 0.8833) and an AUC of 0.9980, suggesting effective risk stratification despite its simpler linear decision structure. Tree-based ensemble methods showed slightly reduced but still clinically acceptable performance. Random forest and gradient boosting achieved accuracies of 0.8500 and 0.8333, respectively, with corresponding AUC values exceeding 0.94, indicating stable ranking capability but comparatively lower sensitivity at the chosen thresholds. The decision tree classifier provided a balanced performance, combining moderate accuracy (ACC = 0.8422) with relatively high precision (PPV = 0.8885) and minimal computational overhead.

Execution times were negligible across all models, supporting their feasibility for real-time clinical use. Overall, the results demonstrate that the proposed classification framework maintains high diagnostic performance when validated against biopsy-confirmed malignant cases. The strong performance of support vector machines and logistic regression highlights their potential utility in clinical decision-support settings, particularly for the reliable identification of high-risk nodules within the Lung-RADS 4B category.

Table 5.6 : Clinical Validation of the Classification Algorithms on Biopsy-Confirmed Malignant Cases from Augusta Victoria Hospital

Stage	Classifier	ACC	PPV	TPR	F1-Score	AUC	Time (S)
LRAD 4B	LR	0.8833	0.9054	0.8833	0.8817	0.9980	0.05
	SVM	0.9333	0.9412	0.9333	0.9330	0.9996	0.21
	RF	0.8500	0.8535	0.8500	0.8496	0.9456	0.14
	GB	0.8333	0.8394	0.8333	0.8326	0.9522	2.3
	DT	0.8422	0.8885	0.8422	0.8664	0.9333	0.08

- ***ACC**: Accuracy; **PPV**: Precision; **TPR**: Sensitivity; **AUC**: Area Under the Curve*

Performance on pathologically verified cases demonstrates the diagnostic reliability of the proposed classification framework.

5.5 Longitudinal Case Results

Figure 5.5 illustrates a representative longitudinal case demonstrating the appearance of a new pulmonary nodule during routine follow-up imaging. The baseline CT shows no detectable lesion, while the subsequent examination performed after 3 months reveals a newly visible solid nodule. This example highlights how interval imaging can capture early nodule development and supports the evaluation of temporal changes in lesion visibility and Lung-RADS categorization.

When LDCT and SDCT cases were merged into a single training and evaluation set, classifier performance improved and became more consistent across Lung-RADS categories. In LR2, all models achieved high accuracy (ACC = 0.86–0.896), with LR, RF, and DT surpassing 0.87 and maintaining strong PPV and F1-scores. This suggests that integrating both dose levels enables the models to learn more generalizable feature representations that are not tied to a specific noise or contrast profile.

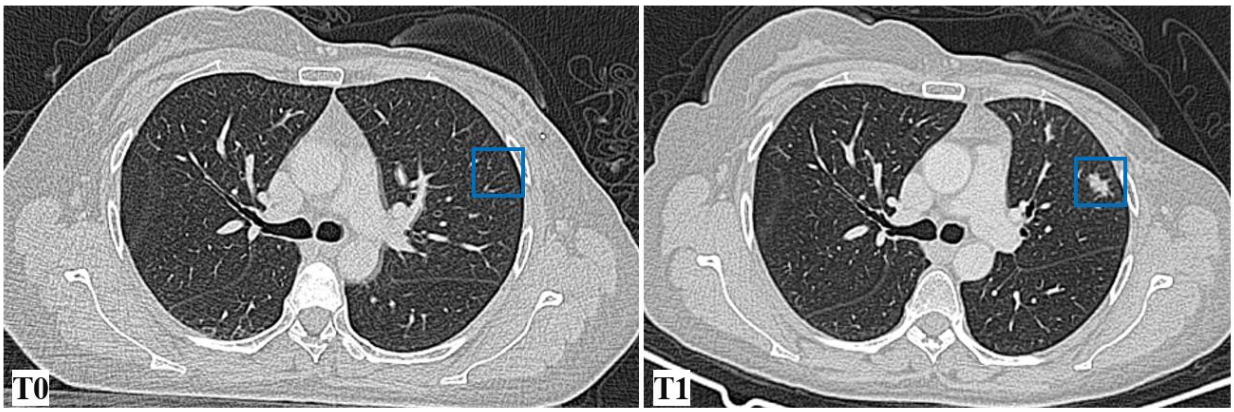


Figure 5.5 Longitudinal Case Showing a Newly Visible Pulmonary Nodule**.**

Baseline CT (T0) shows no detectable nodule (LR1). A 3-month follow-up scan (T1) demonstrates a newly visible solid nodule measuring 15 × 6 mm (LR4B). Images processed with CLAHE–USM for consistency.

A representative progression case is shown in **Figure 5.5,** highlighting interval changes in a left upper lobe lesion between baseline and a 3-month follow-up scan.

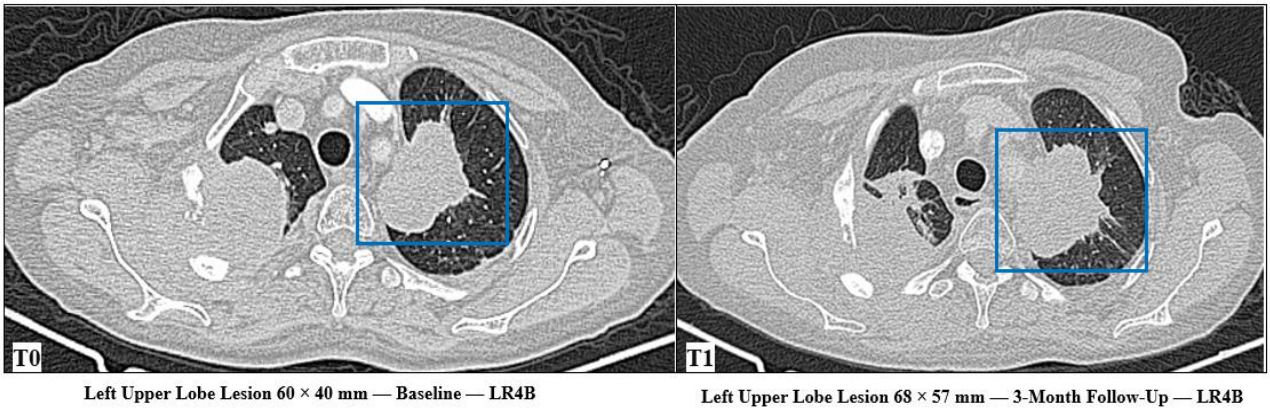


Figure 5.6 Longitudinal Case Showing Interval Growth of a Left Upper Lobe Pulmonary Lesion.

Baseline CT (T0) shows a left upper lobe lesion measuring 60 × 40 mm. A follow-up scan at 3 months (T1) demonstrates interval growth to 68 × 57 mm. All images are processed using the CLAHE–USM enhancement method for consistency

These cases visually demonstrate disease progression and support the quantitative findings reported above.

Chapter 6

Discussion and Conclusion

This chapter summarizes the key findings of the study and explains their technical and clinical significance for lung cancer diagnosis in Palestine. It reviews the performance of the proposed framework and highlights how contrast enhancement, VGG16-based features, and machine-learning models jointly improve nodule visibility and diagnostic accuracy.

The discussion then addresses clinical relevance in resource-limited settings with high radiologist workload, considers the role of combining LDCT and SDCT data in improving robustness, and concludes with key challenges, opportunities for clinical adoption, and future research directions.

6.1 Discussion of Findings

This section interprets the findings of the proposed dose-aware lung-nodule analysis framework in relation to the study objectives and existing literature. The discussion focuses on three main components: the impact of image preprocessing and contrast enhancement, the role of VGG16-based feature extraction, and the overall classification performance across multiple datasets. Particular attention is given to the consistency of results across LDCT and SDCT images, as well as the clinical validation findings.

6.1.1 Image Preprocessing and Contrast Enhancement Optimization

The analysis of contrast-enhancement techniques showed clear differences in their ability to improve local contrast, preserve structural fidelity, and maintain diagnostic quality across both LDCT and SDCT images. These observations align with limitations in prior literature (Jin et al.,

2001b; Toet & Wu, 2014; S. Zhong & Chen, 2016), where enhancement methods were typically applied with fixed parameters and rarely optimized for different dose levels. By contrast, the present study introduces a dose-aware evaluation and demonstrates the value of tailoring preprocessing to the underlying acquisition protocol.

Within this framework, the systematic assessment performed on a subset of 120 malignant LDCT cases confirmed that the hybrid CLAHE–USM technique delivered the most favorable performance across all quantitative metrics. Specifically, it achieved the highest Enhancement Measure Estimation (EME: 20.648), superior Peak Signal-to-Noise Ratio (PSNR: 19.711), and a strong Structural Similarity Index (SSIM: 0.912), while maintaining minimal computational overhead (0.601 ms). These results highlight the effectiveness of CLAHE–USM in improving chest CT image quality and preserving clinically relevant information, particularly in challenging low-dose imaging scenarios. This great performance reflects the method's enhanced ability to increase local contrast and delineate nodule boundaries more clearly. These findings align with recent studies suggesting that combining local contrast enhancement with edge sharpening can improve nodule conspicuity (Jin et al., 2001b; Pizer et al., n.d.; Toet & Wu, 2014); however, unlike previous work, the present study optimized enhancement parameters separately for LDCT and SDCT, yielding more stable results across heterogeneous imaging protocols. The improved visibility is particularly important for subtle lesions and Lung-RADS 4 nodules, where faint margins or low-attenuation centers may be easily overlooked in standard reconstructions.

Unsharp masking (USM) produced the highest PSNR (29.415) and SSIM (0.992), showing that it sharpens lung CT images while largely preserving their original structure. In practical terms, USM enhances anatomical edges and nodule boundaries without noticeably distorting tissue intensity patterns. This behavior is well documented in the medical imaging literature (Deng, 2011; Shukla et al., 2022b; S. Zhong & Chen, 2016), where edge-based sharpening has been shown to improve visual clarity while maintaining structural similarity, particularly in CT-based lung analysis.

In contrast, histogram equalization (HE) and CLAHE showed slightly lower PSNR and SSIM values, reflecting a familiar trade-off between boosting contrast and preserving fine structural detail. Global HE improves overall contrast but can over-enhance homogeneous lung regions, sometimes obscuring subtle textures (Toet & Wu, 2014). CLAHE addresses this issue by enhancing contrast locally, which improves visibility of small structures, but when applied alone it may still alter intensity distributions enough to affect structural similarity. These trends are consistent with previous low-dose CT enhancement studies, which report improved local visibility accompanied by modest reductions in similarity metrics (Demir et al., 2023).

The strength of the proposed CLAHE–USM hybrid approach becomes clear when examining the EME results in Figure 5.1 and Figure 5.2. This method achieved the highest contrast enhancement among all techniques, indicating that it effectively combines the local contrast benefits of CLAHE with the edge-preserving properties of USM. Similar hybrid strategies have been reported to outperform single-method enhancements in pulmonary CT and other medical imaging applications, particularly for improving the visibility of subtle lesions without introducing excessive noise or artifacts(Nanglia et al., 2021; Yim' et al., n.d.) .

The visual comparisons in Figures 5.1 and 5.2 reinforce these quantitative findings across both LR4B and LR2 categories. In the original images, nodules appear poorly defined, with blurred margins and limited contrast against surrounding lung parenchyma, a challenge widely reported in lung nodule detection studies(Armato et al., 2015b). After applying the CLAHE–USM enhancement, nodule edges become clearer and parenchymal textures more distinct, improving both visual interpretation and suitability for automated analysis. Our study results align well with existing studies and support the proposed hybrid enhancement as a practical and effective preprocessing step for lung nodule detection and classification (Saha et al., 2024c; Wankhade & S., 2023).

The differences observed in CLAHE performance between LDCT and SDCT images reflect the fundamental imaging characteristics of each modality. LDCT scans inherently contain higher photon noise and reduced contrast-to-noise ratio, which makes soft-tissue distinctions less apparent and produces less sharply defined nodule boundaries. Under these conditions, enhancement strategies must carefully balance noise amplification with retrieval of clinically relevant details. This explains why a clip limit of 3 produced the most favorable combination of EME, PSNR, and SSIM in the LDCT subset: stronger clip limits introduced excessive noise that ultimately degraded structural fidelity and reduced the clinical interpretability of the enhanced images. In our dataset, CLAHE with clip limit 3 achieved an optimal balance, yielding EME = 19.485, PSNR = 20.896, and SSIM = 0.929, demonstrating its effectiveness in enhancing low-dose CT images while preserving diagnostic information.

In contrast, SDCT images possess higher signal-to-noise ratio and more stable global contrast, making soft-tissue differences more distinguishable and fine anatomical structures more visible. These characteristics allow SDCT scans to tolerate, and benefit from, stronger local contrast enhancement. Accordingly, a clip limit of 4 produced the highest EME values in SDCT cases while maintaining acceptable PSNR and SSIM performance. In our evaluation, CLAHE with clip limit 4 achieved EME = 20.648, PSNR = 19.711, and SSIM = 0.912, confirming that moderate-to-strong local enhancement is optimal for standard-dose CT.

Additionally, the optimization of CLAHE's grid size and EME block size further supports these findings. LDCT images benefited from a larger grid size (16×16) and a larger EME block size (250), which reduced sensitivity to noise and enabled the extraction of more stable local contrast patterns. In contrast, SDCT images achieved higher enhancement with a smaller grid size (4×4)

and an EME block size of 100, allowing the algorithm to emphasize fine anatomical details without being influenced by noise fluctuations. These dose-aware parameter adjustments directly explain the superior quantitative results and improved visual clarity observed in the enhanced images, reinforcing the importance of tailoring preprocessing strategies to the specific characteristics of each acquisition protocol.

Figures 5.3 and 5.4 illustrate the effect of progressively increasing the CLAHE clip limit on contrast, noise, and nodule visibility in SDCT and LDCT images, highlighting a clear trade-off between contrast enhancement and noise amplification. As the clip limit increases, local contrast improves in both dose conditions; however, the tolerance to enhancement differs markedly between SDCT and LDCT.

In SDCT images (Figure 5.3), the higher intrinsic signal-to-noise ratio allows for stronger contrast enhancement with minimal quality degradation. Increasing the clip limit progressively enhances parenchymal texture and vessel–nodule boundaries, with values of 3.0–4.0 providing the most conspicuous nodules. This visual trend is consistent with the quantitative results, where a clip limit of 4 achieved the highest EME while maintaining acceptable PSNR and SSIM, justifying the use of stronger enhancement in standard-dose scans.

In contrast, LDCT images (Figure 5.4) are substantially more sensitive to noise amplification due to lower photon statistics. While moderate enhancement improves nodule visibility, clip limits above 3.0 introduce noticeable graininess and intensity heterogeneity that obscure fine diagnostic details. This observation aligns with the quantitative analysis, which identified a clip limit of 3 as the optimal balance between contrast improvement and noise control for LDCT. These results confirm that maximizing contrast alone is insufficient and that noise behavior must be explicitly considered, particularly in low-dose settings.

The study findings demonstrate that optimal CLAHE performance is dose dependent. LDCT benefits most from conservative enhancement, whereas SDCT can tolerate and benefit from stronger contrast amplification. This conclusion is visually reinforced by the inverse relationship between PSNR and EME highlighted in Figures 5.3 and 5.4 and supports the need for dose-specific parameter tuning to preserve diagnostic quality.

Compared with previous studies, this work advances current practice by explicitly accounting for dose variability in preprocessing. For example, (D. Kim et al., 2024) improved nodule visibility using synthetic thin-slice reconstruction but relied on single-center data and artificially generated slices. In contrast, the present study demonstrates that an optimized CLAHE–USM pipeline applied to real LDCT and SDCT data achieves comparable gains while preserving structural fidelity (Hashemi et al., 2014). Moreover, many prior enhancement and CNN-based approaches applied fixed preprocessing parameters across all scans (Jin et al., 2001b; Toet & Wu, 2014; Yasaka et al., 2018), implicitly assuming uniform image characteristics. The results

here show that such assumptions can lead to suboptimal enhancement, particularly for LDCT. By demonstrating that contrast enhancement must be tailored to dose conditions, this study fills an important gap in the literature and strengthens the robustness of subsequent feature extraction and classification.

6.1.2 Interpretation of VGG16 Feature Extraction

The dose-aware preprocessing strategy developed in this study plays a central role in determining the quality of the learned representations. By optimizing enhancement parameters separately for LDCT and SDCT scans, the preprocessing pipeline ensures that VGG16 receives images with improved local contrast, reduced noise interference, and clearer anatomical boundaries. This is especially important for thin-slice CT images (≤ 1.5 mm), which capture fine-grained spatial information and subtle textural variations that are essential for early-stage nodule recognition and accurate three-dimensional morphological assessment (Jeong et al., 2024).

To overcome the limitations of handcrafted radiomic features and the computational constraints associated with training deep networks from scratch, the proposed framework employs transfer learning with VGG16 as the feature-extraction backbone. VGG16, a sixteen-layer CNN pretrained on the ImageNet dataset, has demonstrated strong performance in thoracic imaging and has been shown to outperform architectures such as VGG19, ResNet50, and DenseNet201 in similar applications (Al-Areqi, Çelebi, et al., 2023). The model's layered convolutional structure enables extraction of both low-level visual cues (edges, textures) and high-level semantic patterns (shape, morphology, spatial relationships), producing a rich 4096-dimensional representation well suited for nodule characterization.

This approach directly addresses the data-scarcity challenges common in medical imaging, where limited case numbers and class imbalance often hinder the training of deep models. By leveraging the pretrained VGG16 feature space, the model avoids overfitting while maintaining the ability to capture complex, clinically meaningful image characteristics.

Compared with prior work such as (Y. Liu et al., 2024), which relied exclusively on handcrafted radiomic features, the transfer-learning approach used here provides several advantages. Radiomic features in Liu's study were sensitive to LDCT noise, required extensive feature-selection steps, and were extracted without dose-aware enhancement, factors that reduced stability and generalizability, particularly across heterogeneous datasets. In contrast, the present study integrates optimized enhancement with deep feature extraction, yielding representations that are more robust to noise, more consistent across varying acquisition protocols, and better suited for downstream classification. These improvements are reflected in the higher AUC values, stronger cross-modality performance, and consistent results across external SDCT validation and biopsy-confirmed clinical cases.

Limitations of handcrafted features extracted from synthetic LDCT images were highlighted in the radiomics-based model by (J. Liu et al., 2024). Although their framework showed that simple shape and size descriptors can be predictive, performance was restricted by the absence of real LDCT variability, the lack of enhancement-aware preprocessing, and the omission of external validation. The present study addresses these issues by using real LDCT and SDCT data from multiple scanners, applying dose-optimized enhancement, and extracting high-level deep features that generalize more reliably across heterogeneous imaging environments.

These findings also build on prior literature that relied either on handcrafted radiomics or on transfer-learning approaches that were not enhancement- or dose-aware. Radiomic features, as shown in (Y. Liu et al., 2024), can capture malignancy-related patterns but exhibit instability in LDCT due to noise sensitivity and limited preprocessing control. Similarly, transfer-learning studies such as (S. Zhang et al., 2019b), (Raza et al., 2023), and (Kumaran S et al., 2024b) used VGG16 or EfficientNet but applied uniform preprocessing to single-dose datasets, limiting feature robustness across varying CT protocols.

In contrast, the present framework integrates dose-specific enhancement with VGG16 feature extraction, producing more stable and noise-tolerant representations across LDCT and SDCT. This directly addresses the feature-instability problem noted in previous studies and provides empirical evidence that preprocessing quality and dose-awareness significantly influence the discriminative power of deep features in lung nodule analysis.

6.1.3 Interpretation of Classification Performance

The classification results obtained across the LDCT, combined, external validation, and clinical validation datasets demonstrate that the proposed VGG16–machine learning framework provides robust and consistent performance for Lung-RADS–based nodule risk stratification. Across all evaluation settings, the models achieved high accuracy, F1-scores, and AUC values for categories LR2 through LR4B, with only modest performance variation between internal and external datasets. These findings indicate that the deep features extracted from dose-aware, enhanced CT images are both discriminative and transferable, supporting reliable generalization across heterogeneous imaging conditions and clinical environments.

On the LDCT dataset (Table 5.3), all classifiers achieved clinically meaningful performance, although variability was observed across Lung-RADS categories. Logistic regression and random forest showed comparable accuracy in LR2 and LR3, accompanied by high precision and excellent AUC values. This suggests that, following VGG16 feature extraction, the separation between lower-risk categories is close to linear, allowing simple linear models to perform effectively. Random forest achieved similar results through ensemble aggregation, providing increased robustness to feature variability. Gradient boosting and decision tree classifiers demonstrated particularly strong performance in LR2 and higher-risk categories, achieving accuracies above 0.85. However, gradient boosting incurred substantially higher

computational cost, whereas decision trees achieved comparable accuracy with negligible inference time, highlighting their practical advantage in time- and resource-constrained clinical settings. Support vector machines showed lower accuracy across LDCT categories, likely due to sensitivity to high-dimensional feature spaces and residual noise in low-dose scans. Nevertheless, their consistently high AUC values indicate preserved ranking capability, even when threshold-based classification was less stable.

A clear performance improvement was observed with increasing malignancy risk. For LR4A and LR4B, most classifiers achieved higher accuracy than in lower-risk categories, reflecting the greater structural distinctiveness of high-risk nodules. This trend is clinically important, as accurate identification of higher-risk Lung-RADS categories directly supports appropriate triage and follow-up decisions in lung cancer screening workflows.

The most important performance gains were observed on the combined LDCT–SDCT dataset (Table 5.4), particularly for LR3 and LR4A–LR4B. In this setting, support vector machines and random forest achieved near-ceiling performance, with accuracies exceeding 0.97 and AUC values approaching unity. Logistic regression also performed strongly, while gradient boosting and decision trees maintained high accuracy with differing computational profiles. These results indicate that increased dataset diversity enhances the ability of certain classifiers, particularly SVM and RF, to exploit the VGG16 feature space, especially in intermediate- and high-risk categories where subtle structural cues are critical. These findings demonstrate that joint training on LDCT and SDCT data, combined with dose-aware preprocessing, produces more stable and generalizable classifiers than training on LDCT alone.

External validation on an independent test set from Al-Makassed Hospital further confirmed the robustness of the proposed framework. For LR2, all major classifiers achieved accuracies above 0.84, with support vector machines and random forest showing the strongest performance. In LR4B, random forest and decision tree classifiers achieved the highest accuracy, while logistic regression and gradient boosting remained competitive. Although support vector machines exhibited slightly lower accuracy in this category, their AUC values remained high, indicating reliable ranking performance.

Clinical validation on biopsy-confirmed LR4B cases from Augusta Victoria Hospital provided the most stringent assessment of diagnostic performance. In this setting, support vector machines achieved the highest accuracy and AUC, followed closely by logistic regression. Random forest, gradient boosting, and decision tree models also maintained strong accuracy and discrimination, with AUC values exceeding 0.93. Collectively, these results demonstrate that the proposed VGG16–machine learning framework maintains high diagnostic performance under real-world clinical conditions, supporting its potential utility as a reliable decision-support tool for lung nodule risk stratification.

Figures 5.5 and 5.6 show longitudinal case illustrations that demonstrate the clinical utility of monitoring suspicious lesions through disease-progression tracking. Figure 5.5 illustrates the critical role of interval imaging in lung cancer risk evaluation. The baseline CT examination (T0) showed no detectable pulmonary nodules and was classified as Lung-RADS 1. However, a follow-up scan acquired three months later (T1) revealed a newly visible solid nodule measuring 15 × 6 mm, meeting the criteria for Lung-RADS 4B. The transition from a negative baseline examination to a high-risk category over a short surveillance interval highlights the dynamic nature of nodule evolution and the importance of timely follow-up imaging. Importantly, the newly developed nodule demonstrated sufficient structural conspicuity following dose-aware contrast enhancement (CLAHE-USM), which supports the reliable feature extraction for machine-learning–based classification. This example reinforces the clinical value of combining longitudinal assessment with automated risk stratification, particularly for detecting emerging high-risk lesions that may not be present at initial screening.

Another progression case is illustrated in Figure 5.6, in which a left upper-lobe pulmonary lesion demonstrates over a short follow-up period, a measurable interval growth. At baseline (T0), the lesion measured 60 × 40 mm, consistent with a Lung-RADS 4B classification. On the follow-up scan performed three months later (T1), the lesion had enlarged to 68 × 57 mm. This interval progression highlights the clinical value of serial CT imaging for monitoring changes in lesion size and morphology, as such temporal information provides important indicators of aggressive pathology. All images were processed using the CLAHE–USM enhancement pipeline to ensure consistent visualization and reliable comparison across time points.

6.2 Conclusion

This study developed and validated a dose-aware deep learning framework for multi-class Lung-RADS classification using heterogeneous LDCT and SDCT datasets collected from multiple clinical institutions. By combining optimized contrast-enhancement techniques, VGG16-based deep feature extraction, and classical machine-learning classifiers, the proposed system demonstrated strong and consistent performance across internal testing, external validation, and biopsy-confirmed malignant cases. The findings highlight the importance of dose-specific preprocessing and cross-modality training in improving generalizability, particularly for intermediate Lung-RADS categories where diagnostic uncertainty is greatest.

The framework's stability across different scanners, reconstruction settings, and clinical environments underscores its potential as a practical decision-support tool, especially in resource-limited healthcare settings. Although several opportunities for future enhancement remain, such as incorporating volumetric growth modeling, expanding longitudinal datasets, and developing adaptive preprocessing pipelines, the present work establishes a robust

foundation for AI systems designed to improve early lung cancer detection and streamline radiologist workflow. Importantly, in the context of Palestine, where no formal lung cancer screening program currently exists, the proposed framework offers a valuable opportunity to strengthen early detection and risk prediction efforts, helping radiologists identify clinically significant nodules at earlier, more treatable stages and ultimately bridging critical diagnostic gaps.

6.3 Future Directions

Several directions emerge for extending this work. First, incorporating true three-dimensional and longitudinal modeling is essential to capture volumetric morphology, internal heterogeneity, and growth dynamics that underpin Lung-RADS progression. The integration of 3D CNNs, volumetric transformer architectures, and automated volume doubling time estimation would enable more accurate assessment of temporal nodule evolution, which cannot be fully addressed by the current 2D framework.

Second, expanding the dataset to larger, multi-institutional cohorts with longer follow-up and greater morphological diversity is critical for improving generalizability. Increased representation of aggressive, part-solid, and post-treatment nodules would strengthen model calibration for clinically complex cases and support earlier detection of high-risk progression patterns.

Third, future studies should evaluate more advanced feature-extraction backbones and hybrid approaches that combine deep learning with radiomics and clinical variables. Vision Transformers, hybrid 2D–3D models, and multi-modal fusion frameworks may better capture subtle textural and contextual cues, particularly in intermediate Lung-RADS categories.

Fourth, further refinement of dose-aware preprocessing is warranted. Adaptive enhancement strategies that account for contrast use, reconstruction kernels, and scanner-specific parameters could reduce cross-institution variability and enhance robustness.

Finally, validation in clinically enriched settings incorporating longitudinal imaging for essential disease progression tracking, patient history, and radiologist feedback would allow more realistic assessment of performance under diagnostic uncertainty and support the development of decision-support tools that augment, rather than replace, clinical judgment.

6.4 Considerations and Challenges

Despite the strong performance of the proposed dose-aware framework across internal, external, and clinical validation cohorts, several limitations should be considered. First, the analysis is based on 2D slice-level feature extraction, which, while computationally efficient, cannot fully capture three-dimensional nodule morphology, volumetric heterogeneity, or growth dynamics. The lack of consistent longitudinal imaging further restricts assessment of temporal progression, a key component of Lung-RADS categorization.

Second, the reliability of reference labels may be affected by incomplete clinical context and imaging history. Lung-RADS assignments were made using single time-point CT scans, often without prior studies or clinical information. In treated cases, post-therapeutic changes such as fibrosis or scarring may mimic or obscure malignancy, introducing label uncertainty. In addition, incomplete or non-diagnostic scans, motion artifacts, low-dose noise, and protocol variability (e.g., reconstruction kernels, slice thickness, contrast use) further complicated interpretation, particularly for intermediate-risk categories (LR3 and LR4A).

Third, although data were collected from multiple scanners and institutions, the overall dataset remains modest compared with large-scale screening cohorts. Rare or complex nodule phenotypes were under-represented, which may limit calibration for less common presentations. Moreover, reliance on VGG16, while effective and computationally practical, constrains the ability to model fine-grained texture patterns that could be better captured by more advanced architectures such as 3D CNNs or transformer-based models.

Finally, although CLAHE and USM were optimized for LDCT and SDCT, residual variability arising from heterogeneous acquisition protocols and dose modulation strategies may not be fully accounted for.

References

ACR. (n.d.). ACR PRACTICE PARAMETER FOR COMMUNICATION OF DIAGNOSTIC IMAGING FINDINGS. Retrieved November 30, 2025, from https://gravitas.acr.org/PPTS/GetDocumentView?docId=74

ACR-SABI-SPR-STR. (n.d.). ACR-SABI-SPR-STR PRACTICE PARAMETER FOR THE PERFORMANCE OF THORACIC COMPUTED TOMOGRAPHY (CT). Retrieved November 30, 2025, from https://gravitas.acr.org/PPTS/DownloadPreviewDocument?DocId=13

ACR-SPR-STR. (n.d.). ACR-SPR-STR PRACTICE PARAMETER FOR THE PERFORMANCE OF HIGH-RESOLUTION COMPUTED TOMOGRAPHY (HRCT) OF THE LUNGS. Retrieved November 30, 2025, from https://gravitas.acr.org/PPTS/DownloadPreviewDocument?DocId=83

ACR–STR. (n.d.). ACR-STR PRACTICE PARAMETER FOR THE PERFORMANCE AND REPORTING OF LUNG CANCER SCREENING THORACIC COMPUTED TOMOGRAPHY (CT). Retrieved November 30, 2025, from https://www.acr.org/Clinical-Resources/Clinical-Tools-and-Reference/Practice-Parameters-and-Technical-Standards

Al-Areqi, F., Celebi, A. T., & Konyar, M. Z. (2023). Lung Cancer Classification Using Image Enhancement and CNN-Based Pre-Trained Models. 2023 Innovations in Intelligent Systems and Applications Conference, ASYU 2023. https://doi.org/10.1109/ASYU58738.2023.10296677

Al-Areqi, F., Çelebi, A. T., & Konyar, M. Z. (2023). Lung Cancer Classification Using Image Enhancement and CNN-Based Pre-Trained Models. 2023 Innovations in Intelligent Systems and Applications Conference (ASYU), 1–7. https://doi.org/10.1109/ASYU58738.2023.10296677

Al Mohammad, B., Hillis, S. L., Reed, W., Alakhras, M., & Brennan, P. C. (2019). Radiologist performance in the detection of lung cancer using CT. Clinical Radiology, 74(1), 67–75. https://doi.org/10.1016/j.crad.2018.10.008

alyasriy, hamdalla; A.-H. M. (2023). The IQ-OTH/NCCD lung cancer dataset. https://doi.org/doi:10.17632/bhmdr45bh2.4

American College of Radiology, 2022. (2022). Lung-RADS Category Descriptor Findings Management 0. https://www.acr.org/Clinical-Resources/Reporting-and-Data-Systems/Lung-Rads

Armato, S. G., Hadjiiski, L., Tourassi, G. D., Drukker, K., Giger, M. L., Li, F., Redmond, G., Farahani, K., Kirby, J. S., & Clarke, L. P. (2015a). LUNGx Challenge for computerized lung nodule classification: reflections and lessons learned.

Armato, S. G., Hadjiiski, L., Tourassi, G. D., Drukker, K., Giger, M. L., Li, F., Redmond, G., Farahani, K., Kirby, J. S., & Clarke, L. P. (2015b). LUNGx Challenge for computerized lung nodule classification: reflections and lessons learned.

Armato, S. G., McLennan, G., Bidaut, L., McNitt-Gray, M. F., Meyer, C. R., Reeves, A. P., Zhao, B., Aberle, D. R., Henschke, C. I., Hoffman, E. A., Kazerooni, E. A., MacMahon, H., van Beek, E. J. R., Yankelevitz, D., Biancardi, A. M., Bland, P. H., Brown, M. S., Engelmann, R. M., Laderach, G. E., … Clarke, L. P. (2011). The Lung Image Database Consortium (LIDC) and Image Database Resource Initiative (IDRI): A Completed Reference Database of Lung Nodules on CT Scans. Medical Physics, 38(2), 915–931. https://doi.org/10.1118/1.3528204

Armato, S. G., McLennan, G., Bidaut, L., McNitt-Gray, M. F., Meyer, C. R., Reeves, A. P., Zhao, B., Aberle, D. R., Henschke, C. I., Hoffman, E. A., Kazerooni, E. A., MacMahon, H., Van Beek, E. J. R., Yankelevitz, D., Biancardi, A. M., Bland, P. H., Brown, M. S., Engelmann, R. M., Laderach, G. E., … Clarke, L. P. (2011). The Lung Image Database Consortium (LIDC) and Image Database Resource Initiative (IDRI): A completed reference database of lung nodules on CT scans. Medical Physics, 38(2), 915–931. https://doi.org/10.1118/1.3528204

Belgiu, M., & Drăgu, L. (2016). Random forest in remote sensing: A review of applications and future directions. In ISPRS Journal of Photogrammetry and Remote Sensing (Vol. 114, pp. 24–31). Elsevier B.V. https://doi.org/10.1016/j.isprsjprs.2016.01.011

Berlin, N. I. (2000). Overview of the NCI Cooperative Early Lung Cancer Detection Program. Cancer, 89(11 SUPPL.), 2349–2351. https://doi.org/10.1002/1097-0142(20001201)89:11+<2349::aid-cncr6>3.0.co;2-4

Bray, F., Laversanne, M., Sung, H., Ferlay, J., Siegel, R. L., Soerjomataram, I., & Jemal, A. (2024). Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: A Cancer Journal for Clinicians, 74(3), 229–263. https://doi.org/10.3322/caac.21834

Canayaz, M., Şehribanoğlu, S., Özgökçe, M., & Akıncı, M. B. (2024). A comprehensive exploration of deep learning approaches for pulmonary nodule classification and segmentation in chest CT images. Neural Computing and Applications, 36(13), 7245–7264. https://doi.org/10.1007/s00521-024-09457-9

Chand Bansal, J., Kusum, ·, Nagar, A. K., Jeena, I., Selvanayaki, J., Shanmugam, K., & Bestak, R. (n.d.). Algorithms for Intelligent Systems Series Editors. https://link.springer.com/bookseries/16171

Christensen, J., Prosper, A. E., Wu, C. C., Chung, J., Lee, E., Elicker, B., Hunsaker, A. R., Petranovic, M., Sandler, K. L., Stiles, B., Mazzone, P., Yankelevitz, D., Aberle, D., Chiles, C., & Kazerooni, E. (2024a). ACR Lung-RADS v2022: Assessment Categories and Management Recommendations. Journal of the American College of Radiology, 21(3), 473–488. https://doi.org/10.1016/j.jacr.2023.09.009

Christensen, J., Prosper, A. E., Wu, C. C., Chung, J., Lee, E., Elicker, B., Hunsaker, A. R., Petranovic, M., Sandler, K. L., Stiles, B., Mazzone, P., Yankelevitz, D., Aberle, D., Chiles, C., & Kazerooni, E. (2024b). ACR Lung-RADS v2022: Assessment Categories and Management Recommendations. Journal of the American College of Radiology, 21(3), 473–488. https://doi.org/10.1016/j.jacr.2023.09.009

Ciofiac, C. M., Mămuleanu, M., Florescu, L. M., & Gheonea, I. A. (2024). CT Imaging Patterns in Major Histological Types of Lung Cancer. Life, 14(4). https://doi.org/10.3390/life14040462

Corrales, L., Rosell, R., Cardona, A. F., Martín, C., Zatarain-Barrón, Z. L., & Arrieta, O. (2020). Lung cancer in never smokers: The role of different risk factors other than tobacco smoking. In Critical Reviews in Oncology/Hematology (Vol. 148). Elsevier Ireland Ltd. https://doi.org/10.1016/j.critrevonc.2020.102895

Croswell, J. M., Baker, S. G., Marcus, P. M., Clapp, J. D., & Kramer, B. S. (n.d.). Cumulative Incidence of False-Positive Test Results in Lung Cancer Screening A Randomized Trial. www.annals.org

de Koning, H. J., van der Aalst, C. M., de Jong, P. A., Scholten, E. T., Nackaerts, K., Heuvelmans, M. A., Lammers, J.-W. J., Weenink, C., Yousaf-Khan, U., Horeweg, N., van 't Westeinde, S., Prokop, M., Mali, W. P., Mohamed Hoesein, F. A. A., van Ooijen, P. M. A., Aerts, J. G. J. V., den Bakker, M. A., Thunnissen, E., Verschakelen, J., … Oudkerk, M. (2020). Reduced Lung-Cancer Mortality with Volume CT Screening in a Randomized Trial. New England Journal of Medicine, 382(6), 503–513. https://doi.org/10.1056/nejmoa1911793

Demir, A., Shames, M. M. A., Gerek, O. N., Ergin, S., Fidan, M., Koc, M., Gulmezoglu, M. B., Barkana, A., & Calisir, C. (2023). Low-Dose CT Image Enhancement Using Deep Learning. ArXiv Preprint ArXiv:2310.20265.

Deng, G. (2011). A generalized unsharp masking algorithm. IEEE Transactions on Image Processing, 20(5), 1249–1261. https://doi.org/10.1109/TIP.2010.2092441

de Wever, W., Coolen, J., & Verschakelen, J. A. (2011). Imaging techniques in lung cancer. In Breathe (Vol. 7, Issue 4, pp. 338–346). https://doi.org/10.1183/20734735.022110

Ferlay J, E. M. L. F. L. M. C. M. M. L. P. M. Z. A. S. I. B. F. (2024). (2024, February 8). Global Cancer Observatory: Cancer Today. Lyon, France: International Agency for Research on Cancer. https://gco.iarc.who.int/today/en

Gong, J., Liu, J. yu, Wang, L. jia, Sun, X. wen, Zheng, B., & Nie, S. dong. (2018). Automatic detection of pulmonary nodules in CT images by incorporating 3D tensor filtering with local image feature analysis. Physica Medica, 46, 124–133. https://doi.org/10.1016/j.ejmp.2018.01.019

Gong, L., Jiang, S., Yang, Z., Zhang, G., & Wang, L. (2019). Automated pulmonary nodule detection in CT images using 3D deep squeeze-and-excitation networks. International Journal of Computer Assisted Radiology and Surgery, 14(11), 1969–1979. https://doi.org/10.1007/s11548-019-01979-1

Hashemi, S., Mehrez, H., Cobbold, R. S. C., & Paul, N. S. (2014). Optimal image reconstruction for detection and characterization of small pulmonary nodules during low-dose CT. European Radiology, 24(6), 1239–1250.

Hejbari Zargar, H., Hejbari Zargar, S., Mehri, R., & Tajdini, F. (n.d.). Using VGG16 Algorithms for classification of lung cancer in CT scans Image.

Horeweg, N., Scholten, E. T., de Jong, P. A., van der Aalst, C. M., Weenink, C., Lammers, J. W. J., Nackaerts, K., Vliegenthart, R., ten Haaf, K., Yousaf-Khan, U. A., Heuvelmans, M. A., Thunnissen, E., Oudkerk, M., Mali, W., & de Koning, H. J. (2014). Detection of lung cancer through low-dose CT screening (NELSON): A prespecified analysis of screening test performance and interval cancers. The Lancet Oncology, 15(12), 1342–1350. https://doi.org/10.1016/S1470-2045(14)70387-0

Inamura, K. (2017). Lung cancer: understanding its molecular pathology and the 2015 wHO classification. Frontiers in Oncology, 7(AUG). https://doi.org/10.3389/fonc.2017.00193

Jeong, J., Park, D., Kang, J. H., Kim, M., Kim, H. Y., Choi, W., & Ham, S. Y. (2024). Deep Learning-Based Slice Thickness Reduction for Computer-Aided Detection of Lung Nodules in Thick-Slice CT. Diagnostics, 14(22). https://doi.org/10.3390/diagnostics14222558

Jian, M., Chen, H., Zhang, Z., Yang, N., Zhang, H., Ma, L., Xu, W., & Zhi, H. (2024). A Lung Nodule Dataset with Histopathology-based Cancer Type Annotation. Scientific Data, 11(1). https://doi.org/10.1038/s41597-024-03658-6

Jin, Y., Fayad, L., & Laine, A. (2001a). Contrast Enhancement by Multi-scale Adaptive Histogram Equalization.

Jin, Y., Fayad, L., & Laine, A. (2001b). Contrast Enhancement by Multi-scale Adaptive Histogram Equalization.

Kazerooni, E. A., Armstrong, M. R., Amorosa, J. K., Hernandez, D., Liebscher, L. A., Nath, H., McNitt-Gray, M. F., Stern, E. J., & Wilcox, P. A. (2016). ACR CT Accreditation Program and the Lung Cancer Screening Program Designation. Journal of the American College of Radiology, 13(2), R30–R34. https://doi.org/10.1016/j.jacr.2015.12.010

Kim, D., Park, J. H., Lee, C. H., Kim, Y.-J., & Kim, J. H. (2024). Improved Consistency of Lung Nodule Categorization in CT Scans with Heterogeneous Slice Thickness by Deep Learning-Based 3D Super-Resolution. Diagnostics, 15(1), 50.

Kim, H. E., Cosa-Linan, A., Santhanam, N., Jannesari, M., Maros, M. E., & Ganslandt, T. (2022). Transfer learning for medical image classification: a literature review. BMC Medical Imaging, 22(1), 1–13. https://doi.org/10.1186/S12880-022-00793-7/FIGURES/6

Klangbunrueang, R., Pookduang, P., Chansanam, W., & Lunrasri, T. (2025). AI-Powered Lung Cancer Detection: Assessing VGG16 and CNN Architectures for CT Scan Image Classification. Informatics, 12(1). https://doi.org/10.3390/informatics12010018

Kozuka, T., Matsukubo, Y., Kadoba, T., Oda, T., Suzuki, A., Hyodo, T., Im, S. W., Kaida, H., Yagyu, Y., Tsurusaki, M., Matsuki, M., & Ishii, K. (2020). Efficiency of a computer-aided diagnosis (CAD) system with deep learning in detection of pulmonary nodules on 1-mm-thick images of computed tomography. Japanese Journal of Radiology, 38(11), 1052–1061. https://doi.org/10.1007/s11604-020-01009-0

Kumaran S, Y., Jeya, J. J., Mahesh T, R., Khan, S. B., Alzahrani, S., & Alojail, M. (2024a). Explainable lung cancer classification with ensemble transfer learning of VGG16, Resnet50 and InceptionV3 using grad-cam. BMC Medical Imaging, 24(1). https://doi.org/10.1186/s12880-024-01345-x

Kumaran S, Y., Jeya, J. J., Mahesh T, R., Khan, S. B., Alzahrani, S., & Alojail, M. (2024b). Explainable lung cancer classification with ensemble transfer learning of VGG16, Resnet50 and InceptionV3 using grad-cam. BMC Medical Imaging, 24(1). https://doi.org/10.1186/s12880-024-01345-x

Kumar, V., Prabha, C., Sharma, P., Mittal, N., Askar, S. S., & Abouhawwash, M. (2024). Unified deep learning models for enhanced lung cancer prediction with ResNet-50–101 and EfficientNet-B3 using DICOM images. BMC Medical Imaging, 24(1). https://doi.org/10.1186/s12880-024-01241-4

Lin, C. Y., Guo, S. M., Lien, J. J. J., Lin, W. T., Liu, Y. S., Lai, C. H., Hsu, I. L., Chang, C. C., & Tseng, Y. L. (2024). Combined model integrating deep learning, radiomics, and clinical data to classify lung nodules at chest CT. Radiologia Medica, 129(1), 56–69. https://doi.org/10.1007/s11547-023-01730-6

Li, R., Zhou, L., Wang, Y., Shan, F., Chen, X., & Liu, L. (2023). A graph neural network model for the diagnosis of lung adenocarcinoma based on multimodal features and an edge-generation network. Quantitative Imaging in Medicine and Surgery, 13(8), 5333–5348. https://doi.org/10.21037/qims-23-2

Liu, D., Zhao, Y., & Liu, B. (2024). The effectiveness of deep learning model in differentiating benign and malignant pulmonary nodules on spiral CT. Technology and Health Care, 32(6), 5129–5140. https://doi.org/10.3233/THC-241079

Liu, J., Corti, A., Corino, V. D. A., & Mainardi, L. (2024). Lung nodule classification using radiomics model trained on degraded SDCT images. Computer Methods and Programs in Biomedicine, 257. https://doi.org/10.1016/j.cmpb.2024.108474

Liu, Y., Hsu, H. Y., Lin, T., Peng, B., Saqi, A., Salvatore, M. M., & Jambawalikar, S. (2024). Lung nodule malignancy classification with associated pulmonary fibrosis using 3D attention-gated convolutional network with CT scans. Journal of Translational Medicine, 22(1). https://doi.org/10.1186/s12967-023-04798-w

Lou, J. (2024). Comparative Analysis of Logistic Regression, Random Forest, and XGBoost for Click-Through Rate Prediction in Digital Advertising (pp. 462–470). https://doi.org/10.2991/978-94-6463-542-3_54

Lundberg, S. M., Allen, P. G., & Lee, S.-I. (n.d.). A Unified Approach to Interpreting Model Predictions. https://github.com/slundberg/shap

MacMahon, H., Naidich, D. P., Goo, J. M., Lee, K. S., Leung, A. N. C., Mayo, J. R., Mehta, A. C., Ohno, Y., Powell, C. A., Prokop, M., Rubin, G. D., Schaefer-Prokop, C. M., Travis, W. D., Van Schil, P. E., & Bankier, A. A. (2017a). Guidelines for management of incidental pulmonary nodules detected on CT images: From the Fleischner Society 2017. In Radiology (Vol. 284, Issue 1, pp. 228–243). Radiological Society of North America Inc. https://doi.org/10.1148/radiol.2017161659

MacMahon, H., Naidich, D. P., Goo, J. M., Lee, K. S., Leung, A. N. C., Mayo, J. R., Mehta, A. C., Ohno, Y., Powell, C. A., Prokop, M., Rubin, G. D., Schaefer-Prokop, C. M., Travis, W. D., Van Schil, P. E., & Bankier, A. A. (2017b). Guidelines for management of incidental pulmonary nodules detected on CT images: From the Fleischner Society 2017. In Radiology (Vol. 284, Issue 1, pp. 228–243). Radiological Society of North America Inc. https://doi.org/10.1148/radiol.2017161659

Mahmoud, M., Wen, Y., Pan, X., Liufu, Y., & Guan, Y. (2025). Evaluation of recent lightweight deep learning architectures for lung cancer CT classification. Frontiers in Oncology, 15. https://doi.org/10.3389/fonc.2025.1647701

Majumder, S., Gautam, N., Basu, A., Sau, A., Geem, Z. W., & Sarkar, R. (2024). MENet: A Mitscherlich function based ensemble of CNN models to classify lung cancer using CT scans. PLoS ONE, 19(3 March). https://doi.org/10.1371/journal.pone.0298527

Malhotra, J., Malvezzi, M., Negri, E., La Vecchia, C., & Boffetta, P. (2016). Risk factors for lung cancer worldwide. European Respiratory Journal, 48(3), 889–902. https://doi.org/10.1183/13993003.00359-2016

Manser, R., Lethaby, A., Irving, L. B., Stone, C., Byrnes, G., Abramson, M. J., & Campbell, D. (2013). Screening for lung cancer. In Cochrane Database of Systematic Reviews (Vol. 2013, Issue 6). John Wiley and Sons Ltd. https://doi.org/10.1002/14651858.CD001991.pub3

Marinakis, I., Karampidis, K., & Papadourakis, G. (2024). Pulmonary Nodule Detection, Segmentation and Classification Using Deep Learning: A Comprehensive Literature Review. In BioMedInformatics (Vol. 4, Issue 3, pp. 2043–2106). Multidisciplinary Digital Publishing Institute (MDPI). https://doi.org/10.3390/biomedinformatics4030111

McKee, B. J., Regis, S. M., McKee, A. B., Flacke, S., & Wald, C. (2015). Performance of ACR lung-RADS in a clinical CT lung screening program. Journal of the American College of Radiology, 12(3), 273–276. https://doi.org/10.1016/j.jacr.2014.08.004

Nam, D. (2024). Lung cancer segmentation dataset with Lung-RADS class. Mendeley Data. https://doi.org/10.17632/5RR22HGZWR

Nam, J. G., & Goo, J. M. (2022). Evaluation and Management of Indeterminate Pulmonary Nodules on Chest Computed Tomography in Asymptomatic Subjects: The Principles of Nodule

Guidelines. Seminars in Respiratory and Critical Care Medicine, 43(6), 851–861. https://doi.org/10.1055/s-0042-1753474

Nanglia, P., Kumar, S., Mahajan, A. N., Singh, P., & Rathee, D. (2021). A hybrid algorithm for lung cancer classification using SVM and Neural Networks. ICT Express, 7(3), 335–341. https://doi.org/10.1016/j.icte.2020.06.007

Palestinian Health Information Center. (2025). Annual Health Report, Palestine 2024- MOH. https://site.moh.ps/index/Books/BookType/2/Language/ar

Pang, S., Meng, F., Wang, X., Wang, J., Song, T., Wang, X., & Cheng, X. (2020). VGG16-T: A novel deep convolutional neural network with boosting to identify pathological type of lung cancer in early stage by ct images. International Journal of Computational Intelligence Systems, 13(1), 771–780. https://doi.org/10.2991/ijcis.d.200608.001

Peters, A. A., Wiescholek, N., Müller, M., Klaus, J., Strodka, F., Macek, A., Primetis, E., Drakopulos, D., Huber, A. T., Obmann, V. C., Ruder, T. D., Roos, J. E., Heverhagen, J. T., Christe, A., & Ebner, L. (2024). Impact of artificial intelligence assistance on pulmonary nodule detection and localization in chest CT: a comparative study among radiologists of varying experience levels. Scientific Reports, 14(1). https://doi.org/10.1038/s41598-024-73435-3

Pizer, S. M., Johnston, R. E., Ericksen, J. P., Yankaskas, B. C., & Muller, K. E. (n.d.). Contrast-Limited Adaptive Histogram Equalization: Speed and Effectiveness.

Qiu, Z., Wu, Q., Wang, S., Chen, Z., Lin, F., Zhou, Y., Jin, J., Xian, J., Tian, J., & Li, W. (2022). Development of a deep learning-based method to diagnose pulmonary ground-glass nodules by sequential computed tomography imaging. Thoracic Cancer, 13(4), 602–612. https://doi.org/10.1111/1759-7714.14305

Raza, R., Zulfiqar, F., Khan, M. O., Arif, M., Alvi, A., Iftikhar, M. A., & Alam, T. (2023). Lung-EffNet: Lung cancer classification using EfficientNet from CT-scan images. Engineering Applications of Artificial Intelligence, 126. https://doi.org/10.1016/j.engappai.2023.106902

Reduced Lung-Cancer Mortality with Low-Dose Computed Tomographic Screening. (2011). New England Journal of Medicine, 365(5), 395–409. https://doi.org/10.1056/NEJMoa1102873

Saha, A., Ganie, S. M., Pramanik, P. K. D., Yadav, R. K., Mallik, S., & Zhao, Z. (2024a). VER-Net: a hybrid transfer learning model for lung cancer detection using CT scan images. BMC Medical Imaging, 24(1). https://doi.org/10.1186/s12880-024-01238-z

Saha, A., Ganie, S. M., Pramanik, P. K. D., Yadav, R. K., Mallik, S., & Zhao, Z. (2024b). VER-Net: a hybrid transfer learning model for lung cancer detection using CT scan images. BMC Medical Imaging, 24(1). https://doi.org/10.1186/s12880-024-01238-z

Saha, A., Ganie, S. M., Pramanik, P. K. D., Yadav, R. K., Mallik, S., & Zhao, Z. (2024c). VER-Net: a hybrid transfer learning model for lung cancer detection using CT scan images. BMC Medical Imaging, 24(1). https://doi.org/10.1186/s12880-024-01238-z

Satish, P., Chowdary, R., Anguera, J., Chandra, S., Vikrant, S., & Editors, B. (n.d.). Lecture Notes in Electrical Engineering 839 Evolution in Signal Processing and Telecommunication Networks (Vol. 2). https://link.springer.com/bookseries/7818

Setio, A. A. A., Traverso, A., de Bel, T., Berens, M. S. N., Bogaard, C. van den, Cerello, P., Chen, H., Dou, Q., Fantacci, M. E., Geurts, B., Gugten, R. van der, Heng, P. A., Jansen, B., de Kaste, M. M. J., Kotov, V., Lin, J. Y. H., Manders, J. T. M. C., Sóñora-Mengana, A., García-Naranjo, J. C., … Jacobs, C. (2017a). Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images: The LUNA16 challenge. Medical Image Analysis, 42, 1–13. https://doi.org/10.1016/j.media.2017.06.015

Setio, A. A. A., Traverso, A., de Bel, T., Berens, M. S. N., Bogaard, C. van den, Cerello, P., Chen, H., Dou, Q., Fantacci, M. E., Geurts, B., Gugten, R. van der, Heng, P. A., Jansen, B., de Kaste, M. M. J., Kotov, V., Lin, J. Y.-H., Manders, J. T. M. C., Sóñora-Mengana, A., García-Naranjo, J. C., … Jacobs, C. (2017b). Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images: The LUNA16 challenge. Medical Image Analysis, 42, 1–13. https://doi.org/10.1016/j.media.2017.06.015

Sheth, V., Tripathi, U., & Sharma, A. (2022). A Comparative Analysis of Machine Learning Algorithms for Classification Purpose. Procedia Computer Science, 215, 422–431. https://doi.org/10.1016/j.procs.2022.12.044

Shukla, A., Bhateja, V., Singh, A., & Rathore, A. S. (2022a). An Improved Unsharp Masking (UM) Filter with GL Mask (pp. 225–231). https://doi.org/10.1007/978-981-16-8554-5_22

Shukla, A., Bhateja, V., Singh, A., & Rathore, A. S. (2022b). An Improved Unsharp Masking (UM) Filter with GL Mask (pp. 225–231). https://doi.org/10.1007/978-981-16-8554-5_22

Simonyan, K., & Zisserman, A. (2014). Very Deep Convolutional Networks for Large-Scale Image Recognition. 3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings. https://arxiv.org/pdf/1409.1556

The Annual Health Report for 2024. (n.d.).

The National Lung Screening Trial Research Team. (2011). Reduced Lung-Cancer Mortality with Low-Dose Computed Tomographic Screening. New England Journal of Medicine, 365(5), 395–409. https://doi.org/10.1056/NEJMoa1102873

Toet, A., & Wu, T. (2014). Efficient contrast enhancement through log-power histogram modification. Journal of Electronic Imaging, 23(6), 63017. https://doi.org/10.1117/1.JEI.23.6.063017

UrRehman, Z., Qiang, Y., Wang, L., Shi, Y., Yang, Q., Khattak, S. U., Aftab, R., & Zhao, J. (2024). Effective lung nodule detection using deep CNN with dual attention mechanisms. Scientific Reports, 14(1). https://doi.org/10.1038/s41598-024-51833-x

van Ginneken, B., Armato, S. G., de Hoop, B., van Amelsvoort-van de Vorst, S., Duindam, T., Niemeijer, M., Murphy, K., Schilham, A., Retico, A., Fantacci, M. E., Camarlinghi, N., Bagagli,

F., Gori, I., Hara, T., Fujita, H., Gargano, G., Bellotti, R., Tangaro, S., Bolaos, L., … Prokop, M. (2010). Comparing and combining algorithms for computer-aided detection of pulmonary nodules in computed tomography scans: The ANODE09 study. Medical Image Analysis, 14(6), 707–722. https://doi.org/10.1016/j.media.2010.05.005

van Ginneken, B. S. M. P. M. et al. (2010). Automatic Nodule Detection 2009 (ANODE09): A benchmark for pulmonary nodule detection. https://anode09.grand-challenge.org/

Veronesi et al. (2014). Diagnostic performance of low-dose computed tomography screening for lung cancer over five years. Journal of Thoracic Oncology, 9. https://doi.org/https://doi.org/10.1097/jto.0000000000000200

Wang, K., Huang, X., Huang, X., Mu, X., Liu, L., & Jin, G. (2025). Fusion of 2.5D deep transfer learning and radiomics for predicting benign and malignant Lung Imaging Reporting and Data System (Lung-RADS) 3 and 4A nodules. Journal of Thoracic Disease, 17(10), 8360–8373. https://doi.org/10.21037/jtd-2025-584

Wang, Y., Zhang, H., Chae, K. J., Choi, Y., Jin, G. Y., & Ko, S. B. (2020). Novel convolutional neural network architecture for improved pulmonary nodule classification on computed tomography. Multidimensional Systems and Signal Processing, 31(3), 1163–1183. https://doi.org/10.1007/s11045-020-00703-6

Wankhade, S., & S., V. (2023). A novel hybrid deep learning method for early detection of lung cancer using neural networks. Healthcare Analytics, 3. https://doi.org/10.1016/j.health.2023.100195

Wender, R., Fontham, E. T. H., Barrera, E., Colditz, G. A., Church, T. R., Ettinger, D. S., Etzioni, R., Flowers, C. R., Scott Gazelle, G., Kelsey, D. K., LaMonte, S. J., Michaelson, J. S., Oeffinger, K. C., Shih, Y. T., Sullivan, D. C., Travis, W., Walter, L., Wolf, A. M. D., Brawley, O. W., & Smith, R. A. (2013). American Cancer Society lung cancer screening guidelines. CA: A Cancer Journal for Clinicians, 63(2), 106–117. https://doi.org/10.3322/caac.21172

WHO. (2023, June 26). World Health Organization (WHO), “Lung cancer,” (WHO) Fact sheets/Lung cancer. . https://www.who.int/news-room/fact-sheets/detail/lung-cancer

Wolf, A. M. D., Oeffinger, K. C., Shih, T. Y., Walter, L. C., Church, T. R., Fontham, E. T. H., Elkin, E. B., Etzioni, R. D., Guerra, C. E., Perkins, R. B., Kondo, K. K., Kratzer, T. B., Manassaram-Baptiste, D., Dahut, W. L., & Smith, R. A. (2024). Screening for lung cancer: 2023 guideline update from the American Cancer Society. CA: A Cancer Journal for Clinicians, 74(1), 50–81. https://doi.org/10.3322/caac.21811

Xiang, D., Zhang, B., Doll, D., Shen, K., Kloecker, G., & Freter, C. (2013). Lung cancer screening: from imaging to biomarker. In Biomarker Research (Vol. 1). http://www.biomarkerres.org/content/1/x1/4

Yasaka, K., Akai, H., Abe, O., & Kiryu, S. (2018). Deep learning with convolutional neural network for differentiation of liver masses at dynamic contrast-enhanced CT: A preliminary study. Radiology, 286(3), 887–896. https://doi.org/10.1148/radiol.2017170706

Yim', Y., Hong, H., & Shin', Y. G. (n.d.). Hybrid Lung Segmentation in Chest CT Images for Computer-aided Diagnosis.

Zeng, D., Huang, J., Bian, Z., Niu, S., Zhang, H., Feng, Q., Liang, Z., & Ma, J. (2015). A simple low-dose X-Ray CT simulation from high-dose scan. IEEE Transactions on Nuclear Science, 62(5), 2226–2233. https://doi.org/10.1109/TNS.2015.2467219

Zhang, S., Sun, F., Wang, N., Zhang, C., Yu, Q., Zhang, M., Babyn, P., & Zhong, H. (2019a). Computer-Aided Diagnosis (CAD) of Pulmonary Nodule of Thoracic CT Image Using Transfer Learning. Journal of Digital Imaging, 32(6), 995–1007. https://doi.org/10.1007/s10278-019-00204-4

Zhang, S., Sun, F., Wang, N., Zhang, C., Yu, Q., Zhang, M., Babyn, P., & Zhong, H. (2019b). Computer-Aided Diagnosis (CAD) of Pulmonary Nodule of Thoracic CT Image Using Transfer Learning. Journal of Digital Imaging, 32(6), 995–1007. https://doi.org/10.1007/s10278-019-00204-4

Zhang, Y., Jiang, B., Zhang, L., Greuter, M. J. W., de Bock, G. H., Zhang, H., & Xie, X. (2021). Lung Nodule Detectability of Artificial Intelligence-assisted CT Image Reading in Lung Cancer Screening. Current Medical Imaging Formerly Current Medical Imaging Reviews, 18(3), 327–334. https://doi.org/10.2174/1573405617666210806125953

Zheng, M. (2016). Classification and Pathology of Lung Cancer. In Surgical Oncology Clinics of North America (Vol. 25, Issue 3, pp. 447–468). W.B. Saunders. https://doi.org/10.1016/j.soc.2016.02.003

Zhong, S., & Chen, S. (2016). An improved unsharp masking model for intensive care unit chest radiograph enhancement. 2016 17th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD), 655–661. https://doi.org/10.1109/SNPD.2016.7515974

Zhong, Z., Zheng, L., Kang, G., Li, S., & Yang, Y. (2017). Random Erasing Data Augmentation. http://arxiv.org/abs/1708.04896

التعلّم الانتقالي متعدد الأنماط لتصنيف موثوق لعُقيدات سرطان الرئة في التصوير المقطعي منخفض الجرعة

إعداد: سارة عصفور

المشرف الرئيسي: د. رضوان قصراوي

المشرف المشارك: د. حسين المصري

الملخص

يُعد سرطان الرئة من أبرز التحديات الصحية عالميًا نظرًا لارتفاع معدلات الإصابة والوفيات المرتبطة به، كما يُمثل في فلسطين عبئًا صحيًا ملحوظًا حيث يُسجَّل ضمن أعلى أسباب الوفيات السرطانية، وغالبًا ما يتم تشخيصه في مراحل متقدمة تقل فيها فرص العلاج الشافي. ويؤكد هذا الواقع الحاجة إلى حلول فعّالة تدعم الكشف المبكر وتحسّن دقة تصنيف العقيدات الرئوية وتُقلّل الأخطاء التشخيصية، خاصة في البيئات محدودة الموارد.

يرتبط خطر الإصابة بسرطان الرئة بعدة عوامل، يأتي في مقدمتها التدخين، إضافة إلى التعرضات المهنية والبيئية والعوامل الوراثية. وعلى الرغم من توفر وسائل متعددة للتشخيص، فإن الكشف المبكر يظل محورًا أساسيًا لتحسين البقاء على قيد الحياة. وقد أثبتت الأدلة من التجارب السريرية الكبرى فاعلية التصوير المقطعي منخفض الجرعة (LDCT) في خفض وفيات سرطان الرئة لدى الفئات عالية الخطورة مقارنةً بأشعة الصدر، إلا أن تطبيقه العملي يترافق مع تحديات عديدة.

ولتقليل هذه التحديات، طُوّرت أنظمة معيارية للتقرير والتصنيف مثل Lung–RADS التي تهدف إلى توحيد المصطلحات والتوصيات وخفض الإيجابيات الكاذبة وتعزيز اتخاذ القرار السريري. غير أن تطبيق Lung–RADS في الممارسة اليومية قد يكون معقدًا بسبب الحاجة إلى تقييم دقيق لخصائص العقيدة وتغيرها عبر الزمن، إلى جانب اختلاف البروتوكولات والأجهزة وجودة الصور. وفي السياق الفلسطيني تحديدًا، تبرز فجوة إضافية تتمثل في غياب برامج وطنية منظمة لفحص (التصوير المقطعي منخفض الجرعة)LDCT ، إضافةً إلى أن معظم نماذج الذكاء الاصطناعي المنشورة جرى تطويرها وتقييمها على بيانات دولية قد لا تعكس خصائص البروتوكولات المحلية أو تنوع الأجهزة وخصائص السكان، مما يحد من قابليتها للتعميم على البيئة السريرية المحلية.

انطلاقًا من هذه الفجوات، تهدف هذه الدراسة إلى تطوير وتقييم إطار ذكاء اصطناعي لدعم تصنيف العقيدات الرئوية وفق معيار Lung-RADS v2022، مع مراعاة اختلاف الجرعة بين التصوير منخفض الجرعة (LDCT) والتصوير التشخيصي القياسي (SDCT) وما يترتب عليه من اختلافات في الضجيج والتباين وجودة الصورة. يعتمد الإطار المقترح على دمج تحسين الصور ومعالجة التباين بصورة مراعية للجرعة، ثم استخراج ميزات عميقة باستخدام VGG16، تليها مرحلة تصنيف عبر خوارزميات تعلم آلي كلاسيكية متعددة بهدف مقارنة الأداء واختيار الأكثر كفاءة ضمن قيود الموارد المتاحة.

اعتمدت الدراسة بيانات متعددة المصادر تشمل صورًا تشخيصية محلية (SDCT) من مستشفيات فلسطينية، إضافة إلى حالات خبيثة مؤكدة بخزعات لتوفير معيار مرجعي عالي الثقة في التحقق السريري، إلى جانب دمج بيانات عامة متاحة من فحوصات LDCT مصنفة وفق Lung-RADS. وتم تبني استراتيجية تحقق متدرجة لضمان تقييم موضوعي لقدرة النموذج على التعميم، تضمنت: (1) مجموعة اختبار محجوزة (Hold-out) من بيانات SDCT المحلية لم تُستخدم في أي مرحلة من التدريب أو تحسين المعالجة، (2) تقييم داخلي على مجموعة التطوير المدمجة LDCT+SDCT، و(3) تحقق سريري إضافي باستخدام حالات مثبتة بالخزعة، بما يعزز موثوقية الاستنتاجات ويدعم قابلية التطبيق العملي.

ركزت الدراسة كذلك على تحسين جودة الصور قبل الاستخلاص الآلي للميزات، حيث تم تقييم عدة تقنيات لتعزيز التباين والحد من الضجيج، وجرى اعتماد نهج يجمع بين CLAHE وUnsharp Masking بعد ضبط معاملات CLAHE بشكل منفصل لصور LDCT وSDCT، بما يحقق توازنًا بين إبراز التفاصيل الدقيقة وحدود العقيدات من جهة، وتجنب تضخيم الضجيج في صور الجرعة المنخفضة من جهة أخرى. وقد مكّن ذلك من توفير مدخلات أكثر ثباتًا لمرحلة استخراج الميزات العميقة، وأتاح تحسينًا في قابلية تمييز الفئات المختلفة وفق Lung-RADS.

ورغم الأداء القوي الذي أظهره الإطار المقترح عبر مراحل التقييم المختلفة، تظل هناك اعتبارات منهجية وعملية ينبغي أخذها بالحسبان، مثل محدودية نمذجة المعلومات الحجمية ثلاثية الأبعاد والنمو الزمني في حال الاعتماد على مدخلات ثنائية الأبعاد، إضافةً إلى تأثير عدم انتظام المتابعات السابقة لبعض الحالات، وتباين البروتوكولات السريرية (السماكة، نوى إعادة البناء، المادة الظليلة، أنظمة التحكم بالجرعة)، وهي عوامل قد تؤثر على ثبات الأداء خاصة في الفئات المتوسطة مثل LR3 وLR4A التي تتسم بطبيعتها بالغموض التشخيصي.

تلخص الدراسة إلى أن الإطار المقترح يمثل خطوة عملية نحو تطوير نظام دعم قرار سريري قابل للتفسير والاستخدام ومُوائم للواقع المحلي، حيث يجمع بين المعالجة المسبقة المراعية للجرعة، واستخلاص الميزات العميقة، والتصنيف وفق Lung-RADS v2022، مع استراتيجية تحقق تشمل اختبارًا خارجيًا على بيانات محلية وتحققًا سريريًا بحالات مؤكدة بالخزعة. ومن المتوقع أن يسهم هذا النهج في تقليل التباين بين القراءات، وتحسين دقة التصنيف وتقليل الإيجابيات الكاذبة، ودعم الكشف المبكر عن سرطان الرئة في فلسطين، خصوصًا في ظل غياب برامج وطنية منظمة للفحص وارتفاع عبء العمل على اختصاصيي الأشعة.

الكلمات المفتاحية: الرئة, العقيدات الرئوية, التعلّم العميق، تعلّم الآلة، التصوير المقطعي المحوسب، الجرعة المنخفضة، الجرعة القياسية.