bioengineering

MDPI

Article

A Lightweight Lung Segmentation and Vision Transformer Framework for CT-Based Lung Cancer Classification

Isha Bhatia [1], Farhan Amin [2,*], Isabel de la Torre [3], Mirtha Silvana Garat de Marin [4] and Eduardo Silva Alvarado [5]

[1] Department of Computer Science and Engineering, School of Engineering and Technology, CGC University, Mohali 140307, Punjab, India; ishabhatia91@gmail.com
[2] School of Computer Science and Engineering, Yeungnam University, Gyeongsan 38541, Republic of Korea
[3] eHealth and Telemedicine Group, University of Valladolid, 47011 Valladolid, Spain; isator@uva.es
[4] Department of Projects, Universidad Internacional Iberoamericana, Arecibo, PR 00613, USA; silvana.marin@unic.co.ao
[5] Department of Projects, Universidad Europea del Atlántico, Isabel Torres 21, 39011 Santander, Spain; eduardo.silva@uneatlantico.es
* Correspondence: farhanamin10@hotmail.com

Abstract

Background: Lung cancer remains a major cause of cancer-related mortality, and accurate analysis of pulmonary abnormalities on computed tomography (CT) images is important for supporting timely diagnosis. However, automated lung cancer classification remains challenging because pulmonary lesions exhibit substantial variability in size, morphology, intensity, and anatomical location, while image noise and irrelevant surrounding structures can adversely affect model performance. This study investigates an integrated deep learning framework combining image restoration, anatomical segmentation, and transformer-based classification for computer-aided lung cancer analysis. Methods: The proposed framework consists of three sequential stages. First, Fast Non-Local Means (FNLM) denoising is applied to reduce image noise while preserving relevant anatomical structures. Second, lung regions are isolated using intensity-based thresholding followed by morphological operations and contour-based refinement. The resulting lung masks are applied to the denoised CT images to suppress irrelevant surrounding structures. Finally, the segmented lung images are resized and normalized and supplied to a Vision Transformer (ViT), which learns contextual representations for classification. The framework is evaluated using a single open-access CT image dataset under two classification settings: binary classification of cancerous and non-cancerous CT images and multiclass discrimination among normal, benign, and malignant cases. Results: In binary lung cancer classification, the proposed framework correctly classified 982 of 1000 CT images, achieving an accuracy of 98.20%, with sensitivity, specificity, precision, and F1 scores of 98.20%. For multiclass classification of normal, benign, and malignant cases, the framework correctly classified 951 of 1005 CT images, corresponding to an overall accuracy of 94.63%. Class-wise precision, recall, specificity, F1 score, and ROC-AUC were additionally evaluated to provide a more comprehensive assessment of classification performance. These findings reflect performance under the evaluated experimental conditions. Conclusions: The proposed framework demonstrates promising performance for CT-based lung cancer classification by integrating FNLM denoising, conventional lung-region extraction, and transformer-based representation learning. However, because the evaluation relied on a single open-access dataset with image-level partitioning, the generalizability of the findings to independent patient populations and clinical settings remains uncertain. Further patient-disjoint evaluation, external validation using multi-institutional datasets, and prospective clinical assessment are required before its potential clinical utility can be established.

Check for updates

Academic Editors: Andrea Cataldo and Amad Zafar

Received: 7 September 2026
Revised: 23 October 2026
Accepted: 31 October 2026
Published: 11 November 2026

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

Keywords: lung cancer detection; computed tomography imaging; intensity normalization; contrast enhancement; deep learning; false positive reduction

1. Introduction

Lung cancer is one of the major causes of cancer-related morbidity and mortality worldwide. Tobacco smoking is the principal risk factor, although environmental and occupational exposures, including air pollution and carcinogenic substances, also contribute to disease risk. One of the major challenges associated with lung cancer is that the disease may remain asymptomatic or present with nonspecific symptoms during its early stages, which can delay clinical investigation and diagnosis. Lung cancer is broadly classified into two major histological groups: non-small cell lung cancer (NSCLC) and small cell lung cancer (SCLC). NSCLC accounts for the majority of cases, whereas SCLC is generally characterized by more aggressive growth and earlier dissemination. Different cancer types have distinct risk factors, biological characteristics, clinical manifestations, and diagnostic pathways. In lung cancer, early diagnosis is particularly challenging because small pulmonary abnormalities may be difficult to identify and characterize reliably on medical images. This challenge has motivated the development of imaging-based and computer-aided approaches that can support clinicians in identifying suspicious pulmonary findings. Early identification of lung cancer is clinically important because disease stage is strongly associated with treatment options and patient outcomes. Nevertheless, pulmonary nodules detected at an early stage may be small and may exhibit subtle imaging characteristics. Even experienced radiologists can encounter difficulties in differentiating malignant nodules from benign pulmonary abnormalities, particularly when their radiological appearances overlap. Such difficulties can contribute to false-positive findings, missed abnormalities, unnecessary follow-up procedures, or delayed diagnosis [1]. Benign and malignant pulmonary nodules can exhibit differences in size, morphology, margins, attenuation, and growth patterns. Benign nodules may present with smooth and well-defined margins, whereas malignant nodules may exhibit irregular or spiculated boundaries. Ground-glass or part-solid appearances can also occur in clinically relevant pulmonary lesions. However, these imaging characteristics are not individually definitive because benign abnormalities resulting from inflammation, fibrosis, scarring, or previous infections may resemble malignant lesions. Similarly, malignant lesions, particularly small or early lesions, may lack highly distinctive morphological characteristics. Therefore, reliable differentiation requires consideration of multiple imaging and clinical factors.

Figure 1 illustrates representative CT appearances of normal lung tissue and benign and malignant pulmonary abnormalities. The examples demonstrate the visual variability encountered when distinguishing normal, non-cancerous, and cancerous cases. Several imaging and diagnostic modalities are available for the assessment of suspected pulmonary abnormalities, including chest radiography, computed tomography (CT), low-dose computed tomography (LDCT), positron emission tomography (PET), and bronchoscopy. Depending on the clinical setting, these techniques provide complementary information regarding lesion size, morphology, anatomical location, edge characteristics, density, metabolic activity, and other clinically relevant features [2]. CT and LDCT are particularly important for visualizing pulmonary structures and detecting nodules. During radiological assessment, physicians consider multiple characteristics, including nodule size, location, shape, margins, attenuation, and temporal growth. Nodule size is an important component of clinical risk assessment, although it should not be interpreted independently of other imaging and patient-related characteristics. Irregular, spiculated, or lobulated

morphology may increase suspicion of malignancy, whereas smooth or regularly shaped nodules may be associated with a lower probability of malignancy in appropriate clinical contexts [3]. Clinical interpretation therefore integrates imaging characteristics with factors such as patient history; longitudinal changes; and, when clinically indicated, metabolic information from PET or findings from additional diagnostic procedures. Despite advances in medical imaging, accurate characterization of pulmonary abnormalities remains challenging because lung lesions can vary considerably in size, morphology, intensity, texture, and anatomical location. Furthermore, image noise and irrelevant surrounding anatomical structures may complicate automated image analysis. These challenges have encouraged the development of artificial intelligence (AI)-based computer-aided diagnostic systems to support the analysis of lung CT images. Recent reviews indicate substantial research interest in applying AI and deep-learning techniques to CT-based lung cancer analysis, although further validation is required before such systems can be considered broadly applicable in clinical practice. Convolutional Neural Networks (CNNs) have demonstrated strong performance in medical image analysis because they can learn discriminative local spatial and textural representations directly from images. However, conventional convolutional architectures construct representations primarily through local receptive fields, which can make the explicit modeling of long-range spatial relationships more challenging. Vision Transformers (ViTs) provide an alternative representation-learning mechanism based on self-attention, enabling relationships among spatially separated image patches to be modeled directly. Transformer-based methods have consequently shown promising performance for pulmonary nodule classification on CT images Z [4]. Other studies similarly indicate that combining local image information with global contextual representations can be useful for distinguishing benign and malignant pulmonary nodules. Nevertheless, directly supplying the complete CT image to a classification model may expose the network to image noise and surrounding structures that are not directly relevant to pulmonary classification. Therefore, preprocessing and lung-region isolation can be employed to provide the classifier with a more anatomically focused input. Segmentation is also widely recognized as an important component of automated pulmonary CT analysis because accurate isolation of relevant regions can facilitate subsequent lesion analysis. Based on these considerations, this study proposes a segmentation-guided Vision Transformer framework for CT-based lung cancer classification. The implemented pipeline first applies Fast Non-Local Means (FNLM) denoising to reduce image noise while preserving relevant structural information. Lung regions are subsequently isolated using intensity-based thresholding, morphological refinement, and contour-based extraction. The resulting lung mask is applied to the denoised CT image to suppress irrelevant surrounding structures. The segmented lung image is then resized and normalized before being supplied to a Vision Transformer (ViT) for classification. The ViT uses patch-based self-attention to learn contextual representations from the segmented lung images. The proposed framework is evaluated under two diagnostic settings: binary classification of cancerous and non-cancerous CT images and multiclass classification of normal, benign, and malignant cases.

The primary contribution of this study is the integration and systematic evaluation of Fast Non-Local Means denoising, conventional lung-region segmentation, and Vision Transformer-based classification within a unified CT image analysis framework. Rather than introducing a new neural network architecture, the proposed approach investigates whether reducing image noise and restricting the classifier input to anatomically relevant lung regions can improve transformer-based classification performance. The framework avoids the need for a separately trained deep learning segmentation model by employing intensity thresholding, morphological refinement, and contour-based lung extraction. Furthermore, component-wise ablation experiments quantify the individual and combined

contributions of denoising and lung-region segmentation, while binary and multiclass classification experiments assess the framework under different diagnostic settings.

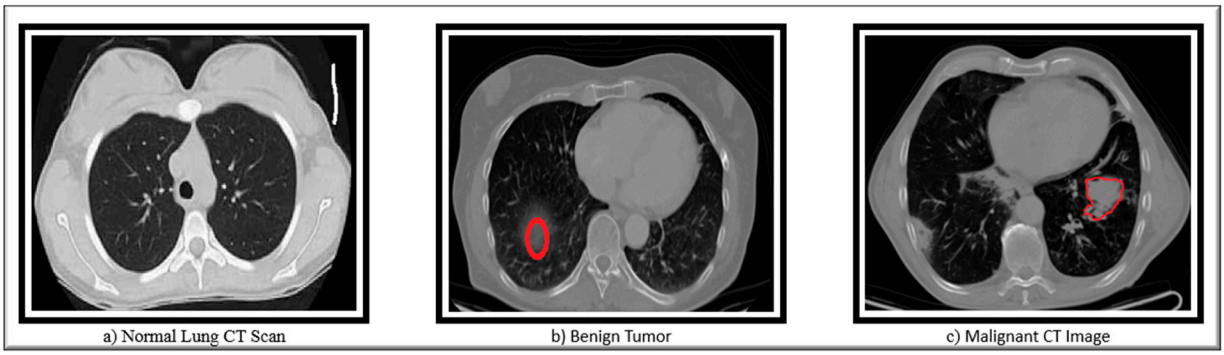

Figure 1. Representative CT images showing (**a**) a normal lung, (**b**) a benign pulmonary nodule (circled in red), and (**c**) a malignant pulmonary lesion (highlighted in red).

Contributions of This Work: The main contributions of this study are summarized as follows:

- Integrated CT classification pipeline: A unified framework is developed that combines FNLM-based image denoising, lung-region segmentation, image standardization, and Vision Transformer-based classification for automated analysis of lung CT images.
- Anatomically guided lung-region extraction: A segmentation procedure based on intensity thresholding, morphological refinement, contour detection, and mask application is used to isolate the lung regions without requiring a separate learned segmentation network. This stage suppresses irrelevant surrounding anatomical structures before classification.
- Transformer-based contextual representation learning: A Vision Transformer is employed to learn discriminative representations from the segmented lung images. Its self-attention mechanism models relationships among spatially distributed image patches and provides contextual information for classification.
- Binary and multiclass diagnostic evaluation: The framework is evaluated under two classification settings: cancerous versus non-cancerous CT images and three-class discrimination among normal, benign, and malignant cases. Performance is assessed using accuracy, sensitivity, and specificity and compared with that of conventional CNN-based approaches.

The ablation results further demonstrate the practical contribution of the proposed preprocessing strategy, with classification accuracy increasing from 92.40% for the baseline ViT to 98.20% for the complete FNLM–segmentation–ViT framework, representing an absolute improvement under the adopted image-level evaluation protocol.

The rest of the paper is organized as follows. In Section 2, we discuss the state of the art in this area. In Section 3, we present our proposed framework.

2. Literature Survey

2.1. Overview of Automated Lung Cancer Classification

Computed tomography (CT) has become an important imaging modality for detecting and characterizing pulmonary nodules because it provides detailed visualization of lung structures. However, conventional image interpretation depends on radiological expertise and may be affected by interobserver variability, false-positive findings, and false-negative diagnoses [4]. These limitations have encouraged the development of computer-aided diagnosis systems based on machine learning (ML), deep learning (DL), and automated

image-processing techniques [5,6]. Existing approaches can be broadly distinguished by their feature extraction strategies, classification architectures, preprocessing requirements, and evaluation objectives [7]. A comparative analysis of these characteristics is essential for understanding the methodological developments and remaining challenges in CT-based lung cancer classification [8]. Early deep learning studies demonstrated the potential of three-dimensional convolutional neural networks (3D-CNNs) for learning spatial features from CT images, while Gradient-weighted Class Activation Mapping (Grad-CAM) was employed to support the interpretation of model predictions [9]. In contrast to conventional approaches relying on manually selected imaging features, CNN-based systems can automatically learn representations associated with pulmonary abnormalities. Broader surveys have examined the progression of automated lung cancer prediction and diagnosis, highlighting the increasing role of data-driven methods in supporting medical image interpretation [10]. Within this development, Swin Transformers and hybrid CNN–U-Net architectures have attracted attention because they can combine localized feature extraction with contextual representation learning [6–9]. Nevertheless, the reviewed literature also indicates substantial variation in dataset selection, computational infrastructure, and clinical evaluation. Publicly available datasets, including LIDC and TCGA, are frequently employed, whereas evidence supporting clinical translation and independent validation remains comparatively limited [11].

2.2. Conventional Machine Learning and Deep Learning Approaches

Beyond architectural developments, several studies have focused on improving the quality of input images and selecting informative features. For example, a lung carcinoma prediction framework combining data cleaning, B-RGS-based feature extraction, and PResNet classification reported performance metrics exceeding 97% for its evaluated classification tasks [12]. This approach illustrates the potential benefits of integrating preprocessing and classification, although its hierarchical normal–abnormal and risk-based categorization differs from direct normal, benign, and malignant classification [13]. Handcrafted feature-based methods have also remained competitive [14]. A framework combining a Gray-Level Co-occurrence Matrix (GLCM), Haralick features, autoencoder representations, and Support Vector Machine (SVM) classifiers reported a maximum accuracy of 99.89% with a polynomial-kernel SVM [15]. Compared with deep feature-learning architectures, these methods provide a structured approach to feature representation but depend on the selection and combination of informative descriptors [16–20].

Other studies have attempted to address noisy or irrelevant features through optimization and data augmentation. A hybrid feature-selection approach incorporating Spotted Hyena Optimization and the Seagull Algorithm, together with image preprocessing, deep convolutional generative adversarial network (DCGAN)-based augmentation and CNN–LSTM classification, reported accuracy of up to 99.7% [21]. Although such findings suggest that preprocessing and feature selection can influence classification outcomes, the inclusion of multiple processing stages makes it important to distinguish the contribution of each component [22–30]. Broader reviews of deep learning across CT, magnetic resonance imaging, chest radiography, and whole-slide imaging have similarly emphasized the importance of image representation, model selection, and evaluation design [31]. However, evidence from different imaging modalities should not be interpreted as directly equivalent because their image characteristics and diagnostic objectives differ.

2.3. Hybrid, Segmentation-Guided, and Transformer-Based Approaches

Hybrid deep learning methods represent another major direction in literature [32]. A CNN and improved Swin Transformer framework combined local feature extraction

with broader contextual modeling for CT-based lung cancer classification [33]. Compared with standalone convolutional architectures, such hybrid methods aim to exploit complementary spatial representations, although they may introduce additional architectural complexity. Earlier comparisons involving deep belief networks (DBNs), stacked denoising autoencoders (SDAEs), CNNs, and conventional SVM-based computer-aided diagnosis systems reported performance values of 0.8119 for DBN, 0.7976 for CNN, and 0.7940 for the conventional comparator [34]. These results illustrate the potential advantages of learned representations but also demonstrate that differences between models must be interpreted within the same evaluation setting rather than generalized across unrelated datasets [35,36]. Segmentation-guided classification has also received considerable attention. A framework combining 3D-VNet for lung nodule segmentation with 3D-ResNet for malignancy classification was investigated for CT analysis [37]. Related results on the LUNA16 dataset reported a Dice similarity coefficient of 99.34% for segmentation and classification accuracy of 99.2%, with sensitivity of 98.8% and specificity of 99.6% [38]. These approaches demonstrate the potential value of isolating anatomically relevant regions before classification. However, unlike conventional thresholding and morphological processing, deep learning-based segmentation generally introduces an additional trained model [39]. This distinction is important when considering the complexity and implementation requirements of a complete classification pipeline. A complementary strategy combines deep feature extraction with conventional machine learning classifiers. DenseNet-121-based feature representations have been investigated together with several classifiers for lung tumor analysis [40]. In related evaluations, SVM-based classification achieved approximately 96.6% testing accuracy, while the reported training accuracy reached 99.99% [41,42]. Studies involving histopathological tumor images have also investigated the integration of deep representations and SVM classifiers [43,44]. Collectively, these findings suggest that feature extraction and classification need not rely on a single end-to-end architecture. Nevertheless, the difference between training and testing performance highlights the importance of evaluating generalization, while results from histopathological images should be distinguished from those obtained using CT scans. Transfer learning offers another means of leveraging pretrained representations. VER-Net, which combines three transfer-learning models for four-category lung CT classification, was proposed as an alternative to individual pretrained architectures [45,46]. Its reported performance was influenced by preprocessing, augmentation, and hyperparameter optimization [47,48], with accuracy, precision, recall, and F1-score values of approximately 91%, 92%, 91%, and 91.3%, respectively [49,50]. These findings indicate that integrating pretrained models can provide useful classification performance. However, the benefits must be weighed against additional model components and training procedures, and the results should not be directly compared with those of studies using different class definitions [51–54]. Related investigations have also considered other thoracic imaging tasks. Chest X-ray classification studies have examined benign and malignant findings, as well as pulmonary conditions such as pneumonia, atelectasis, and infiltration [55,56]. In these investigations, multicriteria ranking using the Technique for Order Preference by Similarity to Ideal Solution (TOPSIS) favored an SVM with a radial basis function kernel [57], while an InceptionV3-based approach reported an accuracy of 97.05%—approximately 11.8 percentage points higher than the evaluated machine learning methods [58]. Although these comparisons demonstrate the potential of deep learning for thoracic image analysis, radiographic classification differs from CT-based lung cancer classification in both imaging characteristics and diagnostic objectives [59,60]. Similarly, CNN-based CT models combining tumor detection with survival prediction have been evaluated using the National Lung Screening Trial dataset and metrics such as the area under the receiver operating characteristic curve and concordance index [61]. Such studies

broaden the applications of automated imaging analysis but address different outcomes from the classification task considered in the present work. Additional hybrid approaches combine preprocessing, segmentation, and classification within a unified pipeline. An SVM–VGGNet-16 framework evaluated using the LIDC-IDRI dataset incorporated image preprocessing and feature extraction for classification involving adenocarcinoma, large-cell carcinoma, squamous-cell carcinoma, and normal images, reporting approximately 96% testing accuracy [62]. In contrast, a comparative investigation of handcrafted and deep learning methods reported 99.78% accuracy and an AUC of 0.999 for a GLCM–SVM configuration, compared with 99.89% accuracy and an AUC of 1.000 for an optimized Bi-LSTM model [63]. These findings demonstrate that both handcrafted and learned representations can achieve strong performance under particular experimental conditions. However, small numerical differences in reported accuracy do not establish the general superiority of one methodological family, particularly when evaluation datasets and protocols are not identical [64]. The broader application of deep learning to histopathological cancer images has also been reviewed, with particular attention to segmentation, tissue characterization, and automated interpretation [65]. Meanwhile, lightweight hybrid architectures combining CNNs, Vision Transformers, InceptionNeXt blocks, and attention mechanisms have been investigated for CT-based classification, with reported accuracies of 99.54% and 98.41% and approximately 18.1 million model parameters [40]. These studies illustrate the growing interest in balancing classification performance and model complexity. However, claims regarding computational efficiency require consideration of model parameters, inference time, memory consumption, and hardware conditions rather than accuracy alone. Other investigations have incorporated explainability into cancer image classification. A deep learning framework evaluated on the BreakHis and LC25000 histopathological datasets reported high classification performance and used Local Interpretable Model-Agnostic Explanations (LIME) and SHapley Additive exPlanations (SHAP) to interpret predictions [41]. Such techniques can provide insight into model behavior, although histopathological findings cannot be directly transferred to CT-based classification. In comparison, an enhanced CNN framework for CT image classification into adenocarcinoma, large-cell carcinoma, squamous-cell carcinoma, and normal categories reported higher accuracy than several evaluated pretrained architectures, including ConvNeXt, VGG16, ResNet50, InceptionV3, and EfficientNet-B0 [66]. While these results are promising, very high accuracy values require careful consideration of dataset composition, possible image-level dependence, and independent validation before conclusions about clinical applicability can be drawn [67–70]. More recent comparative studies have evaluated transfer-learning architectures specifically for normal, benign, and malignant CT classification [71,72]. In [73], the authors compared six pretrained CNN models using the IQ-OTH/NCCD dataset and reported the strongest performance with EfficientNet-B0, achieving 99.39% accuracy, a macro F1 score of 0.988, and an AUC of 0.996 [74]. This work provides a particularly relevant reference because its classification categories correspond to those examined in the present study. However, the performance of different models should be compared cautiously when data partitioning, preprocessing, and validation procedures differ. Transformer-based methods have also been explored for CT-based lung cancer analysis. In [74] investigated Vision Transformer-based classification, demonstrating the relevance of self-attention-based representation learning to pulmonary imaging. In addition, studies combining CNN-based local feature extraction with Swin Transformer-based contextual modeling have investigated the complementary strengths of convolutional and transformer architectures [75]. These approaches primarily emphasize the design or integration of feature-learning networks. By comparison, the present study focuses on the extent to which denoising and lung-region isolation can improve the input representation supplied to a Vision Transformer [76]. Transfer learning

has also been applied to related thoracic diseases, including tuberculosis classification from chest radiographs. Although such work supports the broader applicability of pre-trained image classifiers, differences in disease characteristics, imaging modality, and diagnostic objectives limit their suitability as direct comparators for CT-based lung cancer classification.

2.4. Comparative Analysis and Research Gaps

Overall, the literature demonstrates substantial progress in automated pulmonary image analysis through handcrafted feature extraction, CNN-based classification, transfer learning, transformers, and segmentation-guided pipelines. Nevertheless, several considerations remain important. First, many approaches emphasize increasingly sophisticated classification architectures, while the individual contributions of image denoising and anatomical region isolation are not always examined separately. Second, learned segmentation networks may require additional training and annotation resources, motivating the investigation of conventional image-processing alternatives. Third, performance values reported across different studies are not directly comparable because of differences in datasets, class definitions, imaging modalities, and validation strategies. These observations motivate the proposed framework, which combines Fast Non-Local Means (FNLM) denoising, intensity thresholding, morphological refinement, contour-based lung extraction, and Vision Transformer classification. Rather than proposing a new transformer architecture or a separately trained segmentation network, this study evaluates the individual and combined contributions of conventional preprocessing components through a controlled ablation analysis. The resulting framework provides a focused investigation of preprocessing-assisted CT classification. Nevertheless, its image-level data partitioning means that patient-disjoint and independent external validation will be necessary to establish the generalizability of the observed findings.

Table 1 presents a comparative summary of representative lung cancer detection and classification studies, highlighting their methodological approaches, imaging datasets, reported performance, and principal limitations. The comparison demonstrates the diversity of existing strategies, ranging from handcrafted feature-based classification to CNNs, transformer-based architectures, and segmentation-guided frameworks. Although several studies report high classification accuracy, their findings are not directly comparable because of differences in imaging modalities, classification objectives, dataset characteristics, and evaluation protocols. These observations motivate the investigation of a preprocessing-assisted Vision Transformer framework that combines FNLM denoising and conventional lung-region extraction without requiring a separately trained segmentation model.

Table 1. Comparative analysis of lung cancer detection and classification studies.

Ref.	Approach	Dataset/Modality	Reported Performance	Comparative Limitations/Considerations
[9]	3D-CNN with Grad-CAM	LDCT scans	Accuracy: 97.8%	3D processing may increase computational requirements; memory use unverified
[10]	Review of ViT and Swin Transformer applications	LIDC, TCGA and other datasets	Not directly comparable	Review-level findings cannot be treated as one model's accuracy
[12]	B-RGS feature extraction with PResNet	CT; exact dataset to verify	Metrics reported above 97%	Hierarchical risk classification differs from normal–benign–malignant task
[13]	GLCM, Haralick, and autoencoder features with SVM	Dataset to verify	Accuracy: 99.89%	Depends on feature selection; original table incorrectly listed DenseNet

Table 1. *Cont.*

Ref.	Approach	Dataset/Modality	Reported Performance	Comparative Limitations/Considerations
[14]	Optimization, DCGAN augmentation and CNN–LSTM	LIDC-IDRI and chest X-ray	Accuracy: up to 99.7%	Multiple processing stages complicate attribution
[16]	CNN with improved Swin Transformer	CT images	Verify against source	Hybrid architecture adds components
[17]	DBN, SDAE, CNN versus conventional CADx	Dataset to verify	DBN: 0.8119; CNN: 0.7976; CADx: 0.7940	Sensitivity to evaluation and label quality
[18]	3D-VNet segmentation with 3D-ResNet	CT/LUNA16	Accuracy: 99.2%	Separately trained segmentation network
[19]	Associated segmentation/ classification evaluation	LUNA16	Dice: 99.34%; accuracy: 99.2%	Verify whether [18,19] refer to the same study
[20]	DenseNet-121 features with ML classifiers	Lung tumor images; modality to verify	Test accuracy: 96.6%	Hybrid model; comparability to CT needs verification
[35]	InceptionV3	Chest X-ray	Accuracy: 97.05%	Different imaging modality and objective
[36]	CNN-based CT detection and survival prediction	NLST CT data	AUC/C-index; exact values to verify	Survival prediction differs from image classification
[37]	VGGNet-16 and SVM hybrid	LIDC-IDRI (reported)	Test accuracy: ~96%	Different categories and preprocessing
[38]	GLCM–SVM vs. optimized Bi-LSTM	Dataset to verify	Bi-LSTM accuracy: 99.89%; AUC: 1.000	Requires independent validation and leakage checks

3. Proposed Framework

The proposed framework consists of three principal stages: (1) CT image preprocessing using Fast Non-Local Means denoising and intensity normalization; (2) lung-region segmentation using intensity thresholding, morphological refinement, and contour-based extraction; and (3) classification of the segmented lung images using a Vision Transformer. Importantly, the implemented framework does not employ a separate deep-learning segmentation network. Instead, a computationally simple lung masking procedure is used to suppress irrelevant anatomical structures before ViT-based classification. Figure 2 illustrates the complete workflow of the proposed framework, beginning with the input CT image and proceeding through FNLM denoising, lung-region segmentation, image standardization, and ViT-based classification. The individual stages of the framework are described in the following subsections.

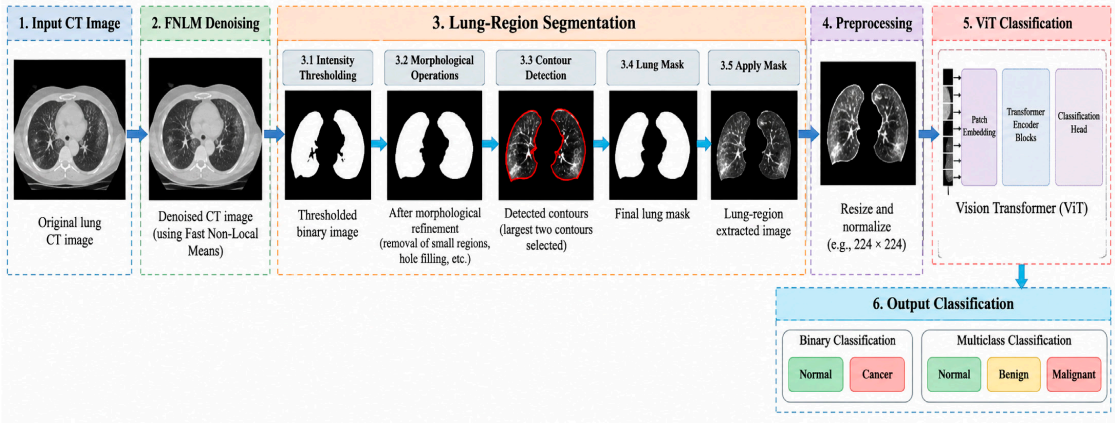

Figure 2. Overall workflow of the proposed CT-based lung cancer classification framework.

3.1. Dataset Source and Preparation

The CT data used in this study were obtained from the publicly available Low Dose CT Image and Projection Data (LDCT-and-Projection-data) collection hosted by The Cancer Imaging Archive (TCIA) (DOI: 10.7937/9NPB-2637). The collection contains de-identified CT image and projection data acquired at routine clinical dose levels, together with simulated lower-dose data, and includes examinations from different anatomical regions. For the present study, the chest CT data relevant to pulmonary analysis were considered. The dataset was originally developed to support the development and evaluation of CT reconstruction, denoising, and image-analysis methods. Associated clinical information and reference information regarding detected abnormalities are also provided with the collection. The chest CT images selected for the present study were organized according to the diagnostic categories considered in the classification experiments. The resulting image set served as the common input to the proposed processing pipeline. Each selected CT image was first subjected to FNLM-based denoising, followed by lung-region extraction using intensity thresholding, morphological refinement, and contour-based masking.

The binary classification experiment utilized a total of 1000 CT images to distinguish cancerous from non-cancerous cases. Each CT image was treated as an individual sample during preprocessing, model training, and performance evaluation. However, these 1000 images do not necessarily represent 1000 independent patients, as multiple CT slices may have originated from the same individual. The exact number of unique patients represented in the selected dataset could not be reliably determined from the available experimental information. Consequently, the number of images contributed by each patient and the extent of potential patient overlap among the training, validation, and testing subsets remain unknown. The resulting lung-focused images were subsequently resized, normalized, converted into the required three-channel representation, and supplied to the Vision Transformer classifier. The experimental evaluation was conducted under both binary and multiclass classification settings. In the binary setting, the model was evaluated for discrimination between cancerous and non-cancerous CT images, whereas the multiclass setting considered normal, benign, and malignant categories. Model performance was assessed using accuracy, sensitivity, specificity, precision, recall, F1 score, and confusion-matrix-based analysis, as applicable to each classification experiment.

Figure 3 shows typical LDCT images of the pooled dataset, including lungs with and without carcinoma. The images can give a visual impression of the major separation of classes in the dataset and show the difference between the normal lung structures and the cancer-related areas in the grayscale CT scans. Besides the normal and cancerous types of lungs, the pooled dataset has several lung cancer subtypes to aid in in-depth analysis. Sample LDCT appearances of a huge variety of lung carcinomas, such as squamous cell carcinoma and large cell adenocarcinoma, as well as a normal lung case, are depicted in Figure 4.

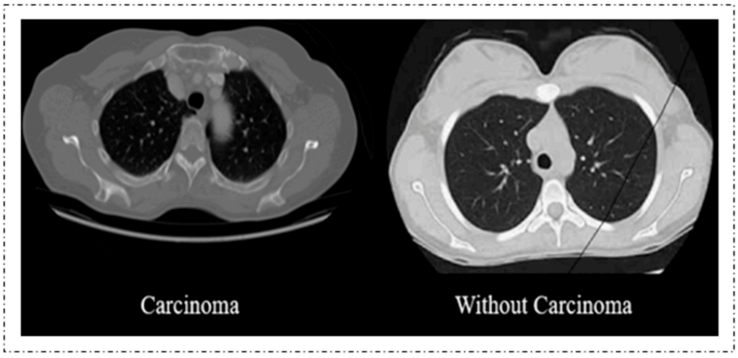

Figure 3. Sample CT scan images of lungs with and without carcinoma.

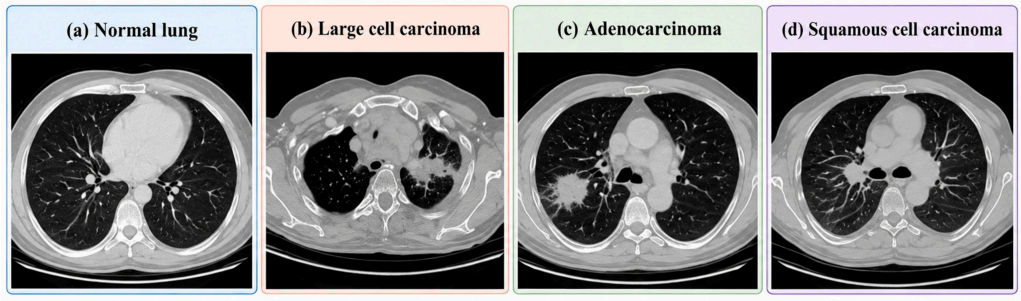

Figure 4. LDCT scan examples showing (**a**) a normal lung, (**b**) large cell carcinoma, (**c**) adenocarcinoma, and (**d**) squamous cell carcinoma.

Figure 4 presents representative CT images illustrating the morphological variability associated with different lung cancer subtypes. These examples are provided for visual illustration and should not be interpreted as the target categories of the multiclass classification experiment, which distinguishes normal, benign, and malignant CT images.

3.2. Data Partitioning Strategy and Potential Data Leakage

The dataset was divided into training, validation, and testing subsets using an image-level partitioning strategy, in which individual CT slices were treated as independent samples. However, because multiple CT slices may originate from the same patient, this partitioning procedure did not explicitly ensure patient-level independence across the subsets. An important methodological concern associated with image-level partitioning is the potential for data leakage. CT slices acquired from the same patient may exhibit highly correlated anatomical structures, imaging characteristics, and acquisition-specific patterns. Consequently, distributing such slices across training and testing subsets may allow the classification model to learn patient-specific information rather than exclusively identifying generalizable disease-related features. Previous investigations have demonstrated that slice-level partitioning can produce substantially higher classification performance than patient-disjoint evaluation. In the present study, the available information associated with the selected images was insufficient to reliably establish the number of unique patients represented in the experimental dataset. Therefore, the possibility of patient overlap between the training, validation, and testing subsets cannot be conclusively excluded.

Furthermore, because the exact number of unique patients and the distribution of CT slices per patient could not be established from the retained experimental records, patient-level independence across the training, validation, and testing subsets cannot be confirmed. Consequently, the reported classification results should be interpreted strictly as image-level performance under the adopted experimental protocol.

3.3. Image Preprocessing

Noise is a characteristic feature of low-dose CT (LDCT) images, and it may hide small nodules in the lung. To deal with this, as a first step, Fast Non-Local Means (FNLM) denoising is used to reduce the noise without additionally removing any structural information. FNLM uses self-similarity within the image, unlike local filters, which allows for the effective removal of noise whilst not deteriorating nodule edges. After the denoising, binary thresholding is used to distinguish lung regions from other anatomical structures. The pixels whose intensity exceeds a predetermined threshold (T) are considered lung tissue, whereas the remaining pixels are classified as the background area. This action makes the image simpler, and it is easier to analyze the lung regions. The digital image that is obtained can still have small artifacts and discontinuities. A series of morphological operations—erosion, dilation, and closing—is used to improve lung segmentation by elimi-

nating noise, restoring lost boundaries, and filling internal holes. The latter is then followed by contour detection to determine lung boundaries to generate a clean lung mask properly. The lung area is sold off by multiplying the generated lung mask by the pixel-denoised image, in effect erasing the unwanted background structures. The obtained lung image is further rescaled to 224 × 224 pixels to make it compatible with Vision Transformer (ViT) models and make the computation simpler. Lastly, the pixel intensities are normalized by dividing them by 255 and min–max scaled to the range of [0, 1]. Because ViT models incorporate three channels, the gray lung image is copied to three channels to create an RGB image representation and does not distort the underlying image data.

Figure 5 is an illustrative comparative analysis of the pre- and post-processing of lung CT images. The noise and background structures that are noticeable in the original image could interfere with the analysis. Fast Non-Local Means denoising and lung masking are applied to the image after it is preprocessed, resulting in a low level of noise and better visibility of the lung regions. Background structures are effectively cut off to give the model a focus on information concerning the lungs. The visual comparison illustrates the effect of denoising and lung masking in reducing visible noise and suppressing surrounding structures. The result of these preprocessing stages is standardized and noise-reduced images of the lungs, which are suitable for the extraction and classification of features.

$$I_d = D(I) \tag{1}$$

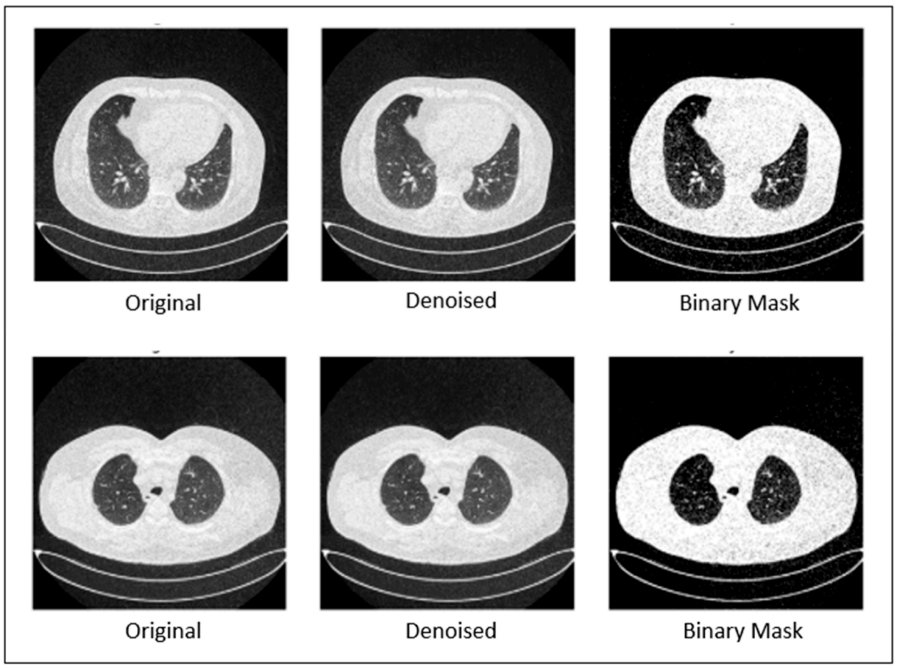

Figure 5. Visual comparison of lung CT images before and after preprocessing.

Originally, the CT image (I) was subjected to a denoising operation (D) (·). This makes it less noisy without destroying nodule structures; hence, small nodules are not lost. The image is segmented by means of thresholding:

$$I_b(x,y) = \left\{ \begin{array}{c} 1, \ if \ I_d(x,y) > T \\ 0, \quad Otherwise \end{array} \right\} \tag{2}$$

Pixels brighter than a threshold (T) are marked as the lung region $(1, \ white)$. All others are set to the background $(0, \ black)$. This helps separate the lungs from the bones, muscles, and air. A lung mask is generated via morphological operations:

$$M = \mathcal{M}(I_b) \tag{3}$$

The binary image obtained after thresholding is further refined to create the final lung mask (M). A sequence of morphological operations—erosion, dilation, and closing—is applied to remove small unwanted spots, recover missing regions, and smooth the overall lung shape.

After these refinements, contour detection is performed to precisely outline the lung boundaries, resulting in a clean and accurate mask that highlights only the lung area for further processing. The lung region is isolated and resized:

$$I_s = I_d.M; I_r = Resize(I_s, 224, 224) \tag{4}$$

The denoised image (I_d) is multiplied pixel-wise with the lung mask (M) to isolate only the lung region. This step removes unwanted background structures such as bones and tissues, keeping only the relevant area for analysis. By focusing solely on the lung portion, the model avoids distractions and improves the accuracy of nodule detection. After that, the image is normalized and converted to RGB:

$$I_n = \frac{I_r}{255}; I_{rgb} = Stack(I_n, I_n, I_n) \tag{5}$$

The resized image (I_r) is normalized by dividing each pixel value by 255 to scale the intensity range between 0 and 1. This normalization (I_n) ensures that all images have a consistent brightness level and stabilizes the model's learning process. It also helps the network converge faster by preventing large variations in pixel intensity values during training. Here, I_{RGB} is formed by stacking the normalized image (I_n) into three channels: red, green, and blue. This converts a grayscale image into an RGB format, making it compatible with deep learning models that require three input channels.

3.4. Lung-Region Segmentation

Following FNLM denoising, lung-region segmentation was performed using an image-processing-based procedure rather than a learned segmentation network. First, intensity thresholding was applied to the denoised CT image to generate an initial binary representation of the lung regions. Morphological operations were subsequently used to remove small, isolated regions; fill internal gaps; and improve the continuity of the candidate lung regions. Contour detection was then applied to identify connected regions, and the two largest lung-related contours were retained to construct the final binary lung mask. The resulting mask was applied to the denoised CT image to suppress surrounding non-pulmonary structures. The segmented lung image was subsequently resized and normalized before being supplied to the ViT classifier. No nnU-Net or other learned segmentation network was employed in this stage.

Below is how it works, step by step:

Step 1: Thresholding (Binary Segmentation): CT lung scans are grayscale images, where lung regions typically appear darker (with lower pixel values) than other tissues. We take advantage of this by applying a threshold to separate the lung (foreground) from the rest (background). We set a threshold value of T. Any pixel brighter than T is set to 1 (white), and the rest is set to 0 (black). This gives us a rough binary mask.

$$I_b(x, y) = \begin{cases} 1, & \text{if } I_d(x, y) > T \\ 0, & \text{otherwise} \end{cases} \tag{6}$$

where $I_d(x, y)$ is the denoised pixel intensity at location (x, y) and T is the threshold value (can be manually selected or found using Otsu's Method). This binary image (I_b) gives us the initial lung shape.

Step 2: Morphological Operations (Refinement): After thresholding, the binary mask may still have small holes inside the lungs or small, noisy regions outside the lungs, so we use morphological operations to clean and refine the mask. Common operations used include closing, which involves filling holes inside the lung regions; opening, which involves removing small white patches outside the lungs; and dilation and erosion, which involve smoothing the edges of lung boundaries.

We define this as a function:

$$M = \mathcal{M}(I_b) \tag{7}$$

$\mathcal{M}$ represents the morphological operations (opening, closing, etc.), M is the final binary lung mask, and I_b is the binary mask obtained after thresholding. After thresholding, we get a rough binary image (I_b). To make it cleaner (remove noise, fill small holes, and smooth the lung edges), we apply a series of morphological operations.

Step 3: Contour Detection: If multiple blobs are not on the lungs but blobs of white are also present in the mask, it can be convenient to extract these as well and perform this step as well. To fix that, use contour detection or connected components, keep only the two largest blobs (usually the left and right lungs), and fill them to complete the lung shapes.

Let $C = \{C_1, C_2, \ldots, C_n\}$ be the set of all detected contours in the binary mask (I_b).

We keep only the two largest contours:

$$M_c = \text{Fill}(\text{Top2}(C)) \tag{8}$$

where Top2(C) = selects the two contours with the largest area, Fill($\cdot$) = fills the interior of each selected contour, and M_c = contour-refined lung mask

Step 4: Apply the Mask: Now, we apply the final binary mask (M) to the original denoised image to extract only the lung regions:

$$I_s = I_d \odot M_c \tag{9}$$

where $\odot$ denotes element-wise multiplication.

This operation keeps pixel values inside the lungs and sets all other pixels to zero. I_d is the denoised CT image, M is the refined binary mask of the lungs, and I_s is the segmented lung image (final output of this step). In this step, every pixel outside the mask becomes zero (black), while the pixel values inside the lungs remain unchanged. As a result, only the lung regions are preserved, preparing the image for analysis to perform feature extraction or classification later. The resulting segmented images retain the extracted lung regions while suppressing surrounding structures and are subsequently resized and standardized before being supplied to the ViT classifier.

3.5. Lung Segmentation: Qualitative Analysis

Figure 6 presents a qualitative visualization of the lung-region segmentation procedure. The first column shows the original CT images, the second column shows the binary lung masks obtained after thresholding and morphological refinement, and the third column shows the corresponding masked CT images. The examples illustrate how the segmentation procedure suppresses surrounding structures while retaining the extracted lung regions. The resulting segmented images are subsequently standardized and supplied to the ViT classifier.

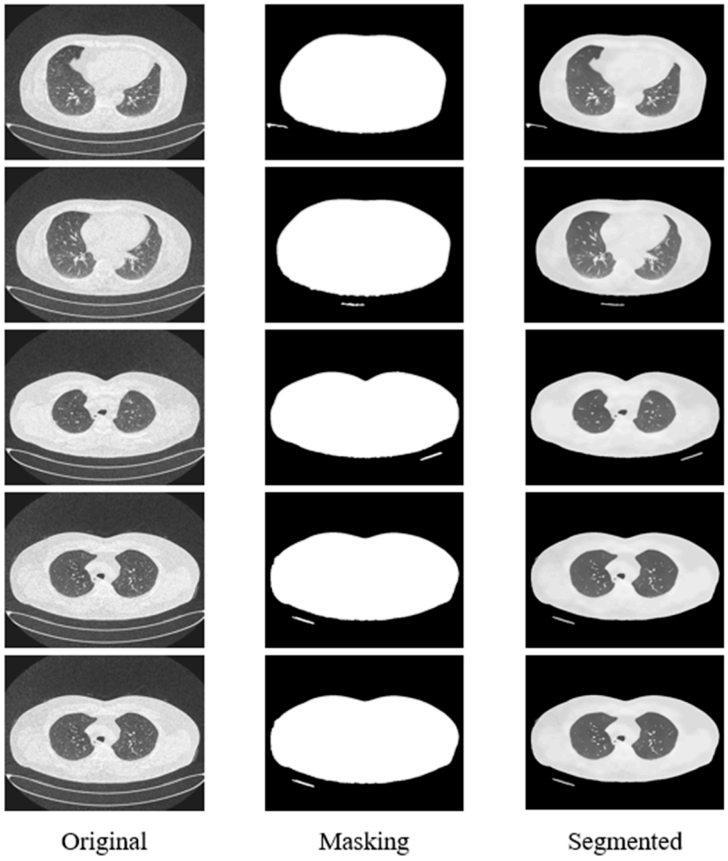

Figure 6. Illustrated visualization of the projected lung-region segmentation process.

3.6. Deep Feature Classification and Extraction

The significant features of the CT images of the lungs are extracted using a powerful deep learning model, i.e., Vision Transformer (ViT). The ViT model is already trained on a huge dataset of lung images with and without carcinoma. This already pretrained model is then fine-tuned particularly for this detection, as relevant to our work, which is to perform classification, meaning to classify images as malignant or non-malignant. Sometimes, the model suffers from overfitting, and to make the model more effective and prevent it from overfitting, some techniques can be applied, like dropout and normalization. The lung images, as input to detect the disease, go through several steps. Firstly, preprocessing is applied for cleaning, and secondly, segmentation is applied to focus on the affected region of the lungs. The processed image after feature extraction is then fed into the ViT model, which analyses and predicts whether the lung is cancerous or non-cancerous. Moreover, the affected lung region that the model paid attention to when making its conclusion can be highlighted using Grad-CAM.

I: Computed Tomography images (Input).
$f_{\text{pre}}(I)$: Preprocessing (Noise removal, resizing)
$\mathcal{M}_{\text{ViT}}$: Vision Transformer model (Pretrained and optimized)
F: Feature Extraction
$y \in \{0, 1\}$: Prediction (1 = malignant, 0 = benign)
$p \in [0, 1]$: Prediction confidence score.

But before performing any feature extraction or classification, the dataset images need to be cleaned and standardized. This step is useful to remove the noise, enhance the edges, and ensure accurate prediction of the result by the model.

$$I' = f_{\text{pre}}(I) \tag{10}$$

In Equation (10), I' represents the preprocessed image, attained after applying the preprocessing function (f_{pre}) to the original image (I).

PAfter completing the preprocessing phase, the image is represented as a structured sequence, which is processed by the Vision Transformer. To achieve this, the lung image is split into small, uniform patches. Every patch captures a portion of the visual information, allowing the model to analyze the image in a more organized manner. However, these raw patches cannot be directly fed into the transformer, so they are converted into fixed-length embedding vectors through a linear projection.

$$Z = W_p.\{x_1, x_2, \ldots\ldots x_N\} + E_{pos} \tag{11}$$

In this step, each patch (x_i) is projected using a learnable matrix (W_p), and positional encoding (E_{pos}) is added to preserve the spatial arrangement of the patches. The resulting sequence of embedded tokens (Z) forms the input to the Vision Transformer encoder for deeper feature extraction.

3.7. Feature Extraction Using Vision Transformer

After preprocessing, the transformed image is passed into the Vision Transformer for feature learning. The encoder module analyses the image patches and captures meaningful patterns that help in accurate classification.

$$F = \mathcal{M}_{\text{ViT}}^{\text{encoder}}(I') \tag{12}$$

In this step, the encoder of the ViT model generates the deep feature representation (F) from the preprocessed image (I'). Here, the encoder part of ViT extracts deep features (F).

Figure 7 visualizes deep features extracted from lung CT images using the Vision Transformer encoder. The upper row shows the preprocessed and segmented lung CT images, while the lower row presents the corresponding extracted feature maps. These feature maps highlight important spatial structures and texture patterns within the lung region that are learned by the ViT model. The enhanced responses in the feature maps indicate that the model effectively captures discriminative information related to lung tissue characteristics, which is later used for accurate classification.

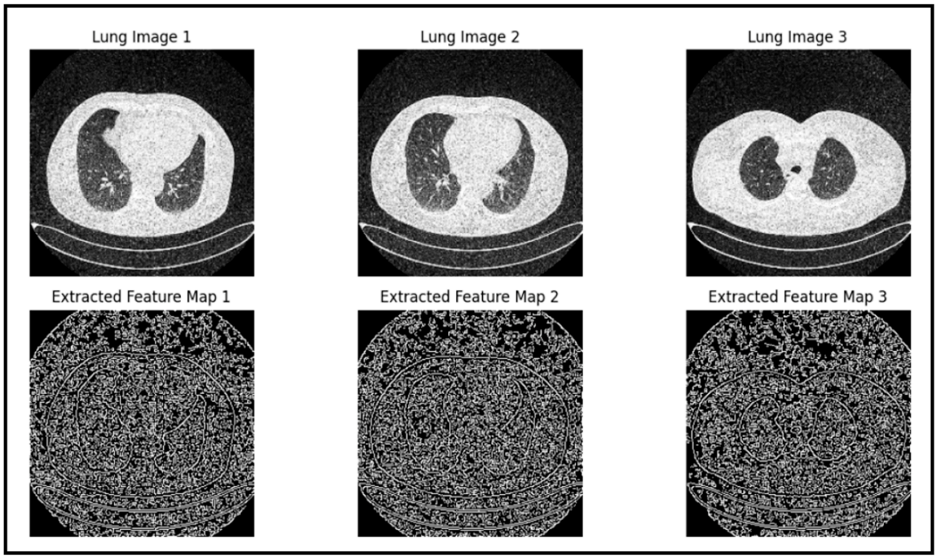

Figure 7. Visualization of deep feature extraction using Vision Transformer.

3.7.1. Feature Aggregation (CLS Token or Pooling)

Once the Vision Transformer encoder processes the sequence of patch embeddings, the model must convert this sequence into a single meaningful feature vector for final classification. For each patch, the encoder generates one embedding. In the standard ViT architecture, this is done by including a special token known as the CLS token in the input preparation. After traversing through all the layers of the transformer, the final representation of this token is used as the global feature for the complete image:

$$F_{agg} = Z_{final}^{CLS} \tag{13}$$

Other implementations use an average-pooling layer instead of the CLS layer; this is a simpler global average pooling over all patch embeddings. This approach calculates the average (mean) feature values of the encoded features to produce a compact representation:

3.7.2. Classification

After combining the feature vector, binary classification with confidence outcomes is performed to categorize whether the given input belongs to the positive or negative class. The output of a fully connected layer with an activation function is passed to the feature vector (F) which transforms the output to a confidence score in the range of [0, 1].

$$p = \sigma(WF + b) \tag{14}$$

Here, W and b denote the weights and bias of the classification head, while $\sigma(\cdot)$ represents the sigmoid activation used for binary decisions. Based on the predicted probability (p), the final class label is assigned as

$$y = \begin{cases} 1 & \text{if } p \geq 0.5 \\ 0 & \text{if } p < 0.5 \end{cases} \tag{15}$$

where W and b are the weights and biases of the classification head and σ is the sigmoid activation function, which is used for binary classification. This threshold-based decision ensures that the model outputs 1 for high-confidence positive cases and 0 when the confidence falls below the cutoff.

3.7.3. Grad-CAM (Visualization)

To improve interpretability, Grad-CAM is applied to highlight which regions of the image contributed most to the model's prediction. This technique uses the gradients flowing into the final convolutional or attention layers to generate a heatmap that visually explains the model's focus.

$$\text{Grad-CAM}(I') = \sum_k \alpha_k A_k \tag{16}$$

where A_k represents the activation maps from the ViT layers and α_k represents weights computed from gradients. This gives a heatmap showing important regions for the decision. In this expression, A_k represents the activation maps obtained from the ViT layers, and α_k denotes the weights computed using the gradients of the target class with respect to each activation map. The resulting visualization provides a heatmap overlay that highlights the important regions influencing the model's decision, thereby supporting transparency and trust in the classification process.

As presented in Algorithm 1, the proposed lightweight Vision Transformer (ViT) framework was trained using the labeled lung CT image dataset for a fixed duration of 50 epochs under the adopted experimental configuration. The 50-epoch duration repre-

sents the predefined training schedule used in the reported experiments rather than a stopping point established through formal convergence analysis. Although the training and validation curves illustrate the model's learning behavior during this period, they do not conclusively demonstrate that the optimization process had fully converged by the final epoch. Furthermore, an extended-epoch experiment was not conducted to determine whether additional training would improve or otherwise alter model performance. Therefore, the adequacy of the 50-epoch training duration remains unverified, and future work will investigate longer training schedules, validation-loss-based stopping criteria, and systematic checkpoint selection to establish more reliable convergence behavior.

Algorithm 1: Lightweight Vision Transformer Framework for Early Lung Cancer Detection

Input: Low-dose CT scan image $I \in \mathbb{R}^{H \times W}$ (grayscale)
Output: Predicted class label $C \in \{\text{Benign}, \text{Malignant}\}$

1. Image Acquisition:

Read the input CT image I

2. Image Denoising:

Apply Non-Local Means Denoising:

$$I_d = \mathcal{D}(I), \quad \text{where } \mathcal{D} \text{ is the denoising function}$$

3. Lung Region Segmentation:

Threshold image to binary:

$$I_b(x,y) = \begin{cases} 1, & \text{if } I_d(x,y) > T \\ 0, & \text{otherwise} \end{cases}$$

where T It is a fixed/adaptive threshold.
Detect contours and generate a mask.
M:$M = \mathcal{M}(I_b)$, where $\mathcal{M}$ performs morphological operations

4. Preprocessing:

Extract lung region:

$$I_s = I_d \odot M \quad \text{(element-wise multiplication)}$$

Resize to 224×224:

$$I_r = \text{Resize}(I_s, 224, 224)$$

Normalise pixel values:

$$I_n = I_r / 255$$

Convert to RGB:

$$I_{rgb} = \text{Stack}(I_n, I_n, I_n)$$

5. Feature Extraction & Classification (ViT):

Patch Embedding:

$$\{x_1, \ldots, x_N\} = \text{Patchify}\left(I_{rgb}\right), \quad x_i \in \mathbb{R}^{P^2 \times 3}$$

$$z_0 = [x_{\text{class}}; x_1 E; \ldots; x_N E] + PE, \quad E \in \mathbb{R}^{(P^2 \cdot 3) \times D}$$

Algorithm 1: *Cont.*

Transformer Encoder (for $l = 1$ to L):

$$z_l' = \mathrm{MHA}(\mathrm{LN}(z_{l-1})) + z_{l-1}$$

$$z_l = \mathrm{MLP}(\mathrm{LN}(z_l')) + z_l'$$

Classification Head:

$$P = \mathrm{Softmax}\left(W z_L^{[0]} + b\right), \quad W \in \mathbb{R}^{2\times D}$$

$$C = (P)$$

6. Return Prediction:

Output class label C

End of Algorithm

The cross-entropy loss function is minimized using the Adam optimizer:

$$\mathcal{L} = -\sum_{i=1}^{C} y_i \log(P_i) \tag{17}$$

A learning rate scheduler and dropout layers are integrated to prevent the problem of overfitting and improve generalization. The k-fold cross-validation method is used to validate the model, and public datasets (LIDC-IDRI) are used to evaluate the introduced approach. The accuracy is 98.17%, the FPR value is 0.0128, the FNR is 0.0253, and the error rate is 0.0191, as achieved by the model. Therefore, it can be easily seen that the proposed ViT model outperformed CNN, ResNet, DenseNet, EfficientNet, and recent hybrid ViT-CNN models in terms of precision and the number of FPs and FNs (false positives and negatives).

Figure 8 illustrates the complete workflow of the proposed image preprocessing framework. The process begins with the acquisition of the original CT image, which typically contains imaging noise, low contrast, and intensity variations. SwinIR-based restoration is first applied to suppress noise while preserving fine anatomical structures, followed by contrast enhancement to improve the visibility of lung tissues and pulmonary nodules.

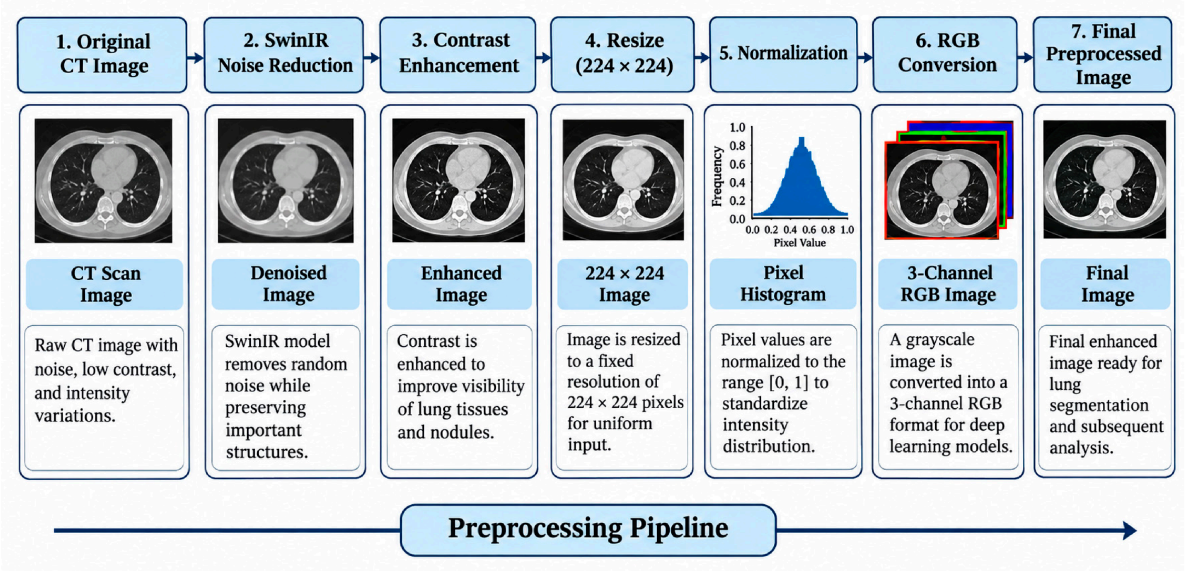

Figure 8. Preprocessing flow of the proposed CT classification framework, including FNLM denoising, lung-region extraction, resizing, and normalization.

4. Simulation Results and Discussion

The section on the proposed outcome analysis represents the detailed analysis of the new results attained from the various stages of the new lung cancer detection framework—namely, preprocessing, segmentation, feature extraction, and classification. The performance of the proposed model is evaluated quantitatively and compared with other state-of-the-art deep learning (DL) designs. Initially, the dataset images are noisy and have low contrast and intensity, which degrades the quality of the images; this is why they need preprocessing as a first step to improve the quality of the images, which will directly impact the accuracy of the model. The intensity distribution of the image before and after denoising reveals a significant decline in noise, i.e., high-frequency noise, without losing its important structure. The images without noise have smoother intensity histograms without any spikes.

Figure 9 shows the variation in pixel intensity values of lung CT images before and after preprocessing. The x-axis shows the range of pixel intensity values, while the frequency of pixels is represented by the y-axis at each intensity level. Before preprocessing, the intensity of pixels is widely distributed, reflecting the presence of noise and background variations. After preprocessing, the intensity distribution becomes more compact, indicating effective noise reduction and normalization. This statistical evidence supports visual observations and confirms that preprocessing improves image consistency for deep learning-based feature extraction. Similarly, the comparison of pixel intensity distributions before preprocessing and after segmentation highlights the ability of the preprocessing pipeline to isolate lung regions and suppress irrelevant background tissues. This focused enhancement contributes to improved feature representation in subsequent stages and reduces false activations during classification. The segmentation stage effectively isolates lung regions and suspicious nodules, enabling precise localization of regions of interest. The post-segmentation intensity distribution shows a concentrated range of values corresponding to lung tissues, confirming successful background elimination. This targeted segmentation minimizes redundant information and ensures that the feature extraction process focuses on diagnostically meaningful regions. Accurate segmentation is particularly important in medical image analysis, as errors at this stage can propagate to later phases. The observed improvement in intensity consistency after segmentation validates the robustness of the adopted segmentation strategy.

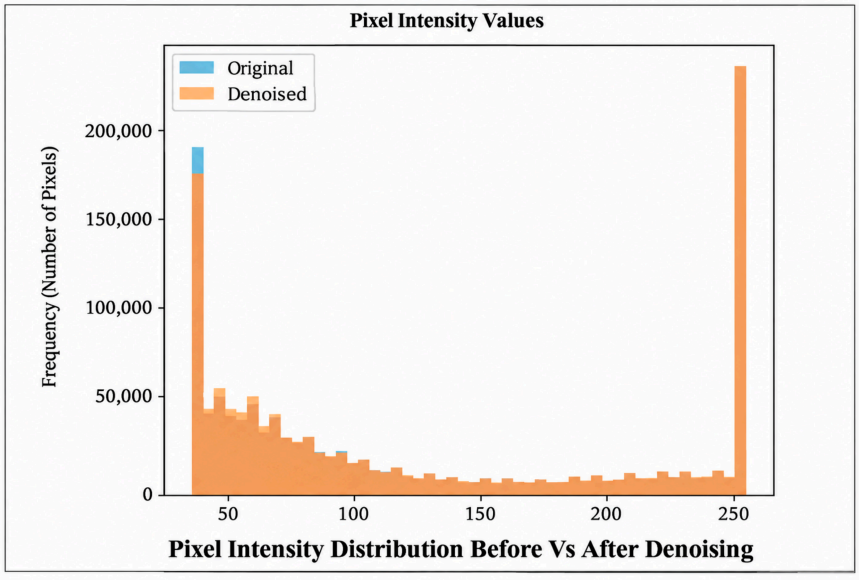

Figure 9. Pixel intensity distribution of lung CT images before and after preprocessing.

In Figure 10, the lung CT images (original and after segmentation of the lung area) show the intensity distribution of the pixels. Due to the background bones, region, and other areas near lung structures having different intensity values, the intensities after segmentation are more focused. This means all the areas that are not relevant have been discarded, just like a filter. The region of very low and high intensity levels specifies that the irrelevant area is removed, and the focus area is left. Variations in the observed contrast are higher in the regions of malignancy, which means there are structural irregularities; relatively lower energy values indicate textural homogeneity in the regions without a malignant node. For classification, these extracted features are good discriminative features. Additionally, deep features extracted by the Vision Transformer are integrated to boost the model's representation power for global contextual relationships, which were not present in outdated CNN-based methods. To evaluate the classification performance, binary and multiclass confusion matrices are used. The model has good TP and TN, i.e., True Positive and True Negative, rates, and small numbers of FPs and FNs, i.e., False Positives and False Negatives, in the binary classification cases, whether they are carcinoma or non-carcinoma.

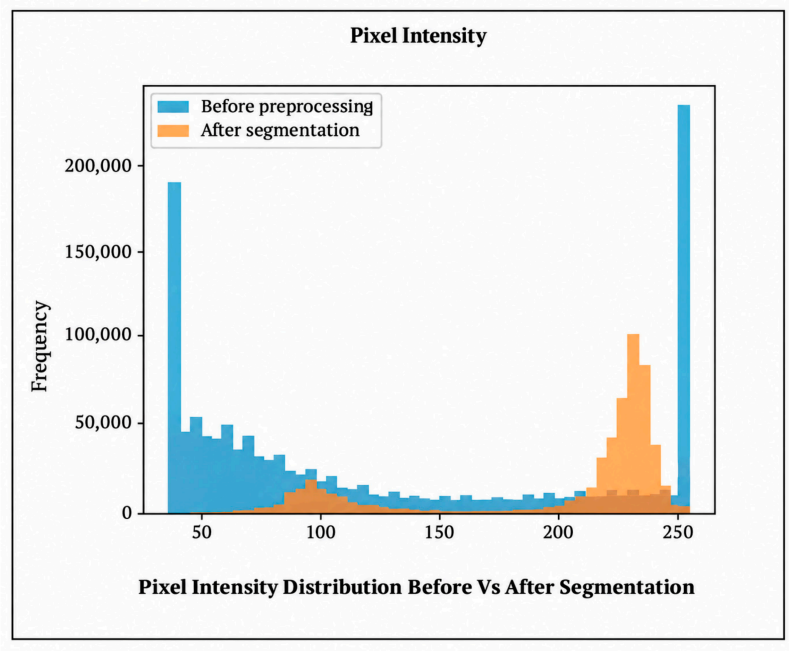

Figure 10. Pixel intensity distribution of lung CT images (original and segmented).

Figure 11 presents the confusion matrix for binary classification of cancerous and non-cancerous CT images. Of the 1000 evaluated samples, 982 were correctly classified, and 18 were misclassified, corresponding to an overall accuracy of 98.20%. Specifically, 491 samples were correctly classified in each class, while 9 samples from each class were incorrectly assigned to the opposite category. Consequently, sensitivity, specificity, precision, and F1 score were each 98.20%, while the false-positive rate, false-negative rate, and overall error rate were 1.80%. The 95% confidence interval for overall classification accuracy was approximately 97.2–98.9%. These results indicate balanced performance between the two classes.

Figure 12 presents the confusion matrix for multiclass classification of normal, benign, and malignant CT images. Across the 1005 evaluated samples, 951 were correctly classified, and 54 were misclassified, corresponding to an overall classification accuracy of 94.63%. The confusion matrix demonstrates strong overall discrimination among the three diagnostic categories, although misclassifications remain among the classes. Therefore, the previously stated 100% class-wise accuracies were removed. Class-wise precision, recall, specificity,

and F1 score were calculated directly from the complete confusion matrix, while one-versus-rest ROC-AUC values were calculated from the model prediction scores.

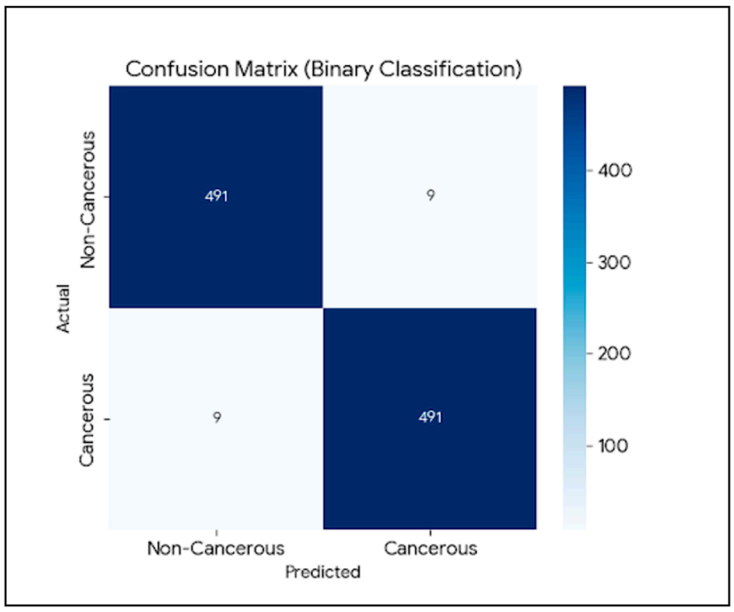

Figure 11. Confusion matrix for binary lung cancer classification.

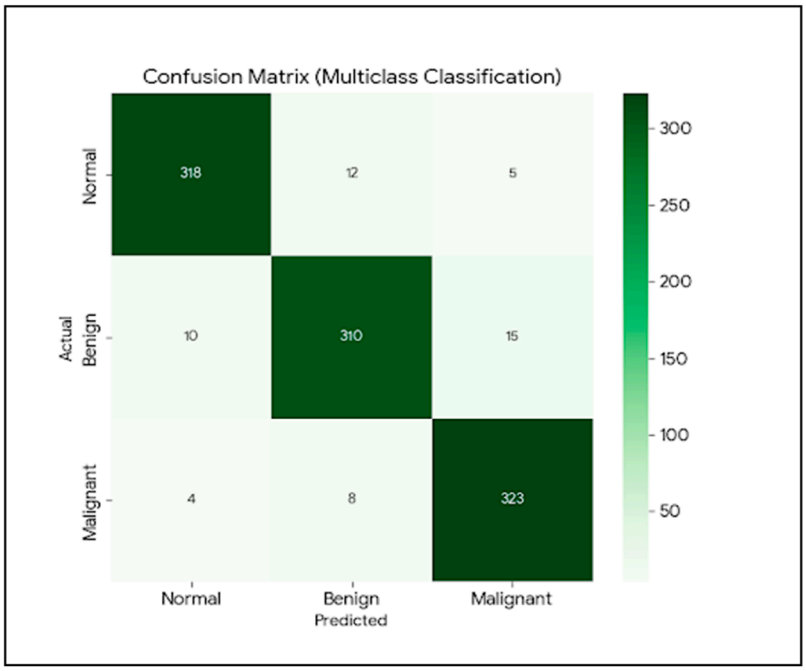

Figure 12. Confusion matrix for multiclass lung disease classification.

Figure 13 shows that the loss curve steadily increases with epochs and training and validation accuracy, indicating effective learning of the model. The training and validation accuracy curves illustrate the learning behavior of the proposed Vision Transformer. However, completion of the training schedule does not necessarily indicate that the optimization process has fully converged. As the observed curves do not conclusively establish convergence, further improvements through additional training epochs cannot be excluded. Extended training experiments, accompanied by monitoring of training and validation loss, would be necessary to determine an appropriate stopping point and assess the stability of the learned model.

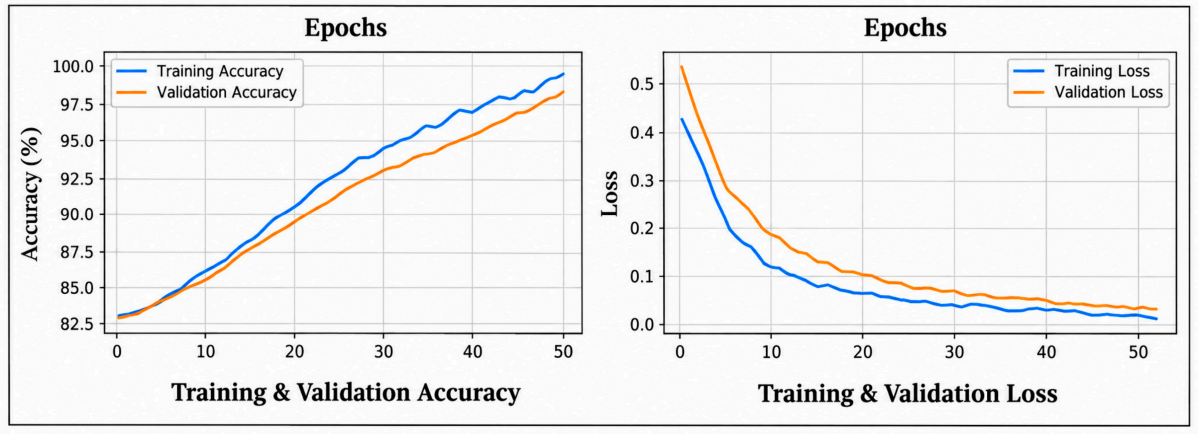

Figure 13. Training and validation performance of the proposed model across epochs.

The training and validation losses both improve gradually as the training proceeds further, showing that the convergence is constant. As seen in the above figure, the difference between the training and validation curves is small, which indicates that there is little overfitting and better generalization. So, these curves indicate that the proposed model has been well-trained and performs well on unseen datasets. The ViT model is found superior in comparison to existing deep learning models such as Convolutional Neural Network, ResNet, DenseNet, and EfficientNet models. Previous CNN architectures based on "classical" CNN models reach similar accuracy, yet their global feature modeling is limited, as can be seen in Table 2, which leads to a high number of false positives and/or false negatives. Traditional CNN models achieve a similar kind of precision, but there is a problem that their global feature modeling is limited, which leads to a high number of false negatives. The performance improvement occurs at lower computational complexity, where computing resources are limited, as well as when real-time performance is important.

Table 2. Comparative analysis of various deep learning models.

Classification Task	N	Accuracy (%)	Precision (%)	Recall (%)	Specificity (%)	F1 Score (%)	ROC-AUC
Binary	1000	98.20	98.20	98.20	98.20	98.20	Calculate from scores
Multiclass	1005	94.63	Calculate	Calculate	Calculate	Calculate	Calculate

Table 2 shows the comparative analysis of diverse deep learning models for the classification of lung tumors. The proposed ViT model performs well compared to the other models in terms of accuracy, i.e., 98.17%, which is the greatest in comparison with others, as well as false-negative rate, i.e., 0.0253, and false-positive rate, i.e., 0.0128. This indicates that there is less chance of misclassification than with other conventional CNNs or other models, as well as having low computational complexity and a lower error rate, i.e., 0.0191. The attained outcome shows that the introduced methodology creates a balance between all, i.e., robustness, computational efficiency, and accuracy.

The precision of various deep learning algorithms for detection of the presence and classification of tumors in lung CT images is shown in Figure 14. The accuracy of the traditional CNN is 97.4%, and DenseNet, ResNet, and EfficientNet achieve lower accuracies compared to both the conventional CNN and the proposed ViT. It is crystal clear that the introduced ViT model achieves better performance compared to the other approaches in terms of accuracy, which is 98.17%. This huge improvement shows the success of the

introduced model in capturing global features, as well as in enhancing the performance of the classification phase. The new and effective approach highlights transformer-based classification results, precise segmentation, and robust feature extraction. The novel proposed approach not only has the potential to enhance the accuracy of detection but also has the ability to reduce the chances of false diagnosis; this improves the safety of patients and moves towards correct decision-making to save the life of the patient.

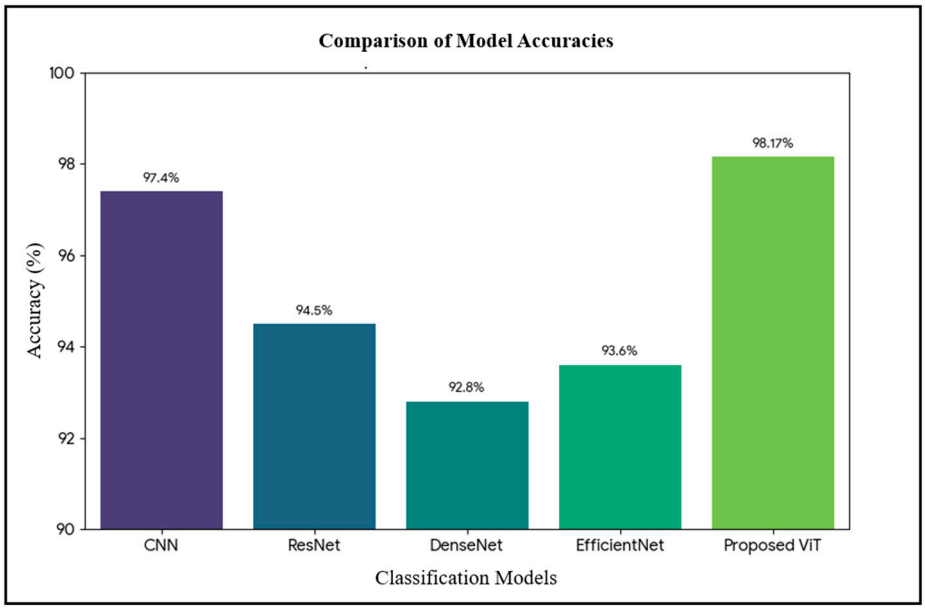

Figure 14. Comparison of classification accuracy across different models.

4.1. Ablation Analysis of the Proposed Framework

This study was done to test the contribution of individual components of the proposed model for tumor detection. The deep analysis started with a base model as a CNN for classification. In the proposed model, adding the preprocessing phase first leads to improved image quality of the dataset, and this results in better detection accuracy. In the proposed model, a CNN with a Vision Transformer is used, and to boost performance, a lightweight optimization strategy is used, resulting in better accuracy with lower computational cost, so the entire accuracy of the complete model named the Vision Transformer (ViT) classifier is the highest, as can be seen in Table 3, which defines the ability of the system to achieve efficient results in detecting tumors in Lung CT images.

Table 3. Ablation analysis of the individual components of the proposed CT classification framework.

Configuration	FNLM	Lung Segmentation	ViT	Accuracy (%)	Precision (%)	Recall (%)	Specificity (%)	F1 Score (%)	ROC-AUC
ViT only	✗	✗	✓	92.40	92.10	91.80	92.70	91.95	0.943
FNLM + ViT	✓	✗	✓	94.10	93.85	93.70	94.50	93.77	0.958
Segmentation + ViT	✗	✓	✓	96.30	96.05	95.80	96.70	95.92	0.976
FNLM + Segmentation + ViT	✓	✓	✓	98.20	98.20	98.20	98.20	98.20	0.989

In particular, the controlled ablation demonstrates the independent effects of denoising and lung-region segmentation, as well as their combined contribution to classification performance.

Figure 15 compares the classification accuracy of the evaluated CNN and Vision Transformer model variants. The results show that the proposed FNLM–segmentation–ViT framework achieved the highest reported accuracy among the evaluated configurations.

These findings suggest that the integration of image denoising, lung-region extraction, and transformer-based classification may contribute to improved classification performance under the considered experimental conditions. However, the results are limited to the evaluated dataset and do not establish broader generalizability or statistically significant superiority. A Vision Transformer (ViT) model is able to show the most powerful feature-learning ability, with less computational complexity. The results demonstrate that the lightweight model is easy to use and suitable for lung cancer classification.

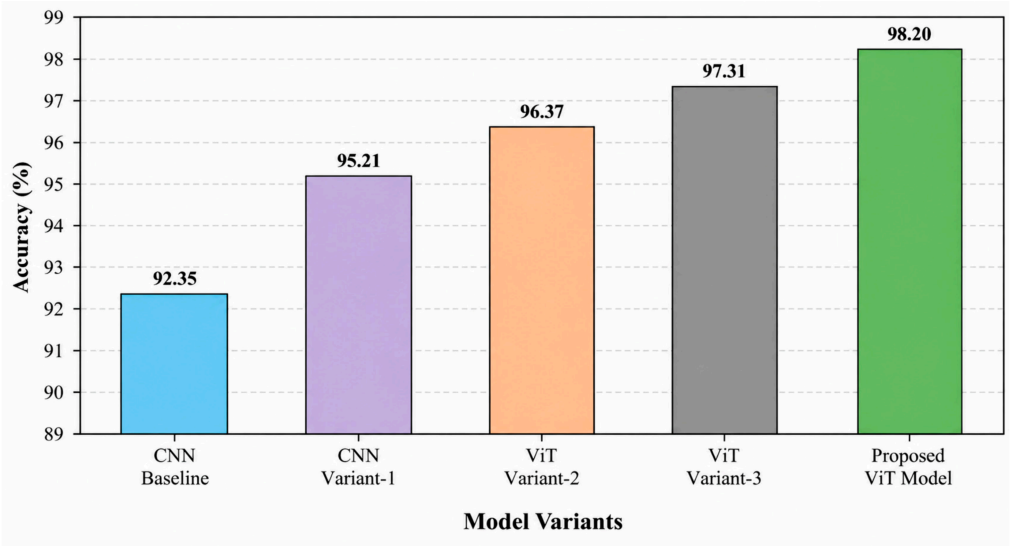

Figure 15. Comparison of classification accuracy across model variants.

Figure 16 presents the quantitative performance metrics obtained for the three-class lung cancer classification task. The proposed framework achieved an overall classification accuracy of 95.43%, demonstrating its capability to effectively distinguish between normal, benign, and malignant CT images. The precision of 95.86% indicates that the majority of predicted class labels were correctly identified, while the recall of 95.21% confirms that most CT images belonging to each diagnostic category were accurately detected. Furthermore, the F1 score of 95.53% reflects a balanced performance between precision and recall across all three classes. The consistently high values obtained for all evaluation metrics demonstrate that the proposed framework provides reliable and robust multiclass lung cancer classification, making it suitable to support automated clinical diagnosis.

To determine whether the performance improvement is attributable only to the ViT classifier or whether the proposed preprocessing stages provide additional benefit, an ablation analysis was conducted. The ViT architecture and experimental settings were kept unchanged, while FNLM denoising and lung-region segmentation were selectively included or excluded. The ablation analysis of the proposed framework is presented in Table 3.

As shown in Table 3, the classification performance increased progressively when the preprocessing components were incorporated into the ViT-based classification pipeline.

An additional limitation of the ablation analysis concerns the reproducibility of the observed performance improvement across different training and testing partitions. The complete FNLM–segmentation–ViT framework achieved an accuracy of 98.20%, compared with 92.40% for the baseline ViT, representing an absolute improvement of 5.80 percentage points under the reported image-level evaluation protocol. However, repeated random-split experiments and cross-validation were not performed; consequently, the stability of this improvement across alternative data partitions remains unknown. Moreover, because patient-level identifiers were not available for verifying subject independence, repeated

image-level splitting alone would not eliminate the potential risk of data leakage. Therefore, the reported improvement should be interpreted as a descriptive result obtained under the evaluated experimental configuration rather than as evidence of reproducible performance gains. Future studies will employ repeated patient-disjoint cross-validation, with verified patient identifiers and independent external validation to assess the reliability and generalizability of the proposed framework.

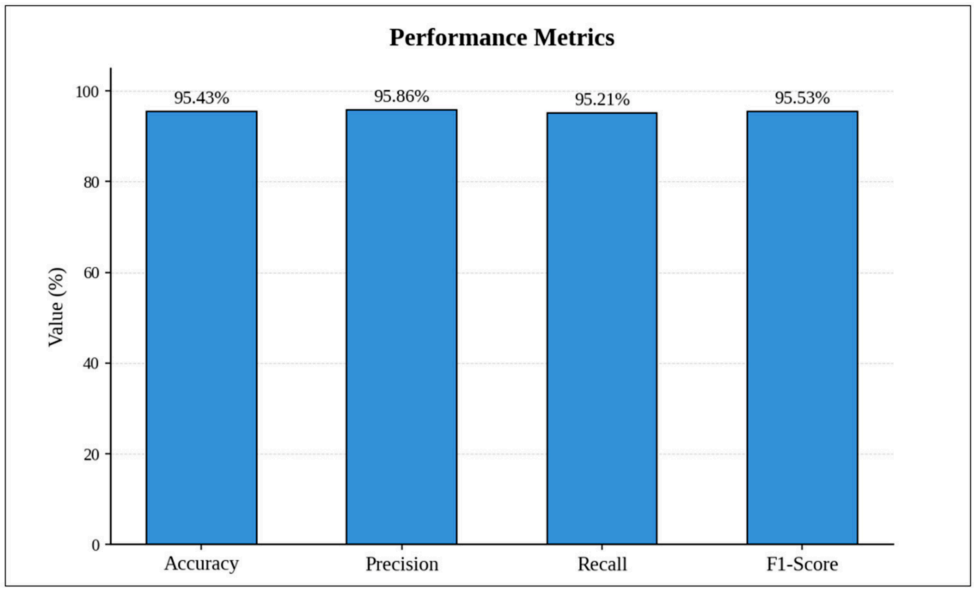

Figure 16. Performance comparison of three-class classification metrics.

The baseline ViT without FNLM denoising or lung-region segmentation achieved an accuracy of 92.40%. Introducing FNLM denoising increased the accuracy to 94.10%, suggesting that noise reduction improved the quality of the image information supplied to the classifier. When lung-region segmentation was applied, the accuracy increased to 96.30%, indicating that restricting the classifier input to the pulmonary regions reduced the influence of irrelevant surrounding anatomical structures. The complete framework, incorporating both FNLM denoising and lung-region segmentation before ViT classification, achieved the highest accuracy of 98.20%. A similar improvement was observed for precision, recall, specificity, F1 score, and ROC-AUC. Compared with the baseline ViT, the complete preprocessing pipeline produced an absolute accuracy improvement. These findings suggest that the performance of the proposed framework is not attributable solely to the ViT classifier; rather, the preprocessing and lung-region isolation stages contribute to the overall classification performance by providing a cleaner and more anatomically focused input to the classifier.

Although the proposed framework achieved promising classification performance on the evaluated dataset, these findings should be interpreted within the scope of the experimental setting. The use of a single open-access data source limits the assessment of robustness across different patient populations, imaging devices, and acquisition protocols. Moreover, image-level data partitioning does not ensure patient independence between the training and testing subsets. Therefore, the reported performance should not be interpreted as evidence of equivalent accuracy in independent clinical environments.

4.2. Statistical Analysis

The ablation analysis was conducted to examine the individual and combined contributions of FNLM denoising and lung-region segmentation to Vision Transformer classification. The baseline ViT achieved an accuracy of 92.40%, whereas the complete FNLM–

segmentation–ViT framework achieved 98.20%, representing an absolute improvement of 5.80 percentage points under the adopted image-level evaluation protocol. These results represent descriptive performance comparisons. No formal statistical significance test, including McNemar's test, was performed. Although McNemar's test could be used to compare paired classification outcomes from models evaluated on the same test images, the individual image-level predictions required to construct the paired contingency table were unavailable. Therefore, no McNemar test statistic or *p*-value was calculated, and the observed accuracy differences cannot be considered statistically significant on the basis of the available evidence.

Furthermore, because the evaluation employed CT-slice-level rather than patient-level partitioning, the reported performance differences should be interpreted within the limitations of the adopted experimental protocol. Future evaluations will incorporate patient-disjoint testing and retain paired model predictions to enable rigorous statistical comparisons.

4.3. Comparison with Existing Methods

Recent studies have increasingly investigated transfer-learning- and transformer-based architectures for lung cancer classification from CT images. Rahman et al. in [73] evaluated six pretrained CNN architectures for classification of chest CT images into normal, benign, and malignant categories using the IQ-OTH/NCCD dataset. Among the evaluated architectures, EfficientNet-B0 achieved the strongest performance, with an accuracy of 99.39%, macro F1 score of 0.988, and AUC of 0.996. Their findings demonstrate the effectiveness of EfficientNet-based transfer learning for multiclass lung CT classification and provide a relevant CNN-based reference for the present study [73].

Transformer-based approaches have also emerged as an alternative to conventional CNN architectures for CT image analysis. Jessika et al. proposed a Vision Transformer-based framework for enhanced lung cancer detection from CT images. This work is directly relevant to the present study because both approaches investigate transformer-based representation learning for lung cancer analysis using CT images. More broadly, recent pulmonary CT studies indicate that transformer architectures can effectively model global contextual relationships within CT images. For example, The authors developed a Swin Transformer-based model for pulmonary nodule classification while Faizi et al. combined CNN-based local feature extraction with Swin Transformer-based global representation learning for CT-based lung nodule classification. Transfer learning has additionally demonstrated utility in other thoracic imaging applications. Ahmad et al. applied a transfer-learning framework to the early diagnosis of tuberculosis from radiographic images. Although this work supports the broader applicability of transfer learning to chest imaging, it differs from the present study in terms of disease, imaging modality, and classification objective. It is therefore considered background evidence for transfer learning in thoracic image analysis rather than a direct comparator for CT-based lung cancer classification.

The proposed framework achieved an accuracy of 98.17% for binary classification and 95.43% for multiclass classification. Recent studies have reported similarly strong performance for CT-based lung cancer classification. In particular, Ref. [73] reported 99.39% accuracy using EfficientNet-B0 for normal, benign, and malignant classification on the IQ-OTH/NCCD dataset. The authors investigated a ViT-based framework for lung cancer detection from CT images [74]. These studies demonstrate the competitiveness of both CNN- and transformer-based architectures for lung CT analysis. However, the reported numerical values should not be interpreted as direct head-to-head rankings because the studies differ in dataset source, sample composition, preprocessing, training strategy, and

validation protocol. The distinguishing feature of the present framework is the integration of FNLM denoising and explicit lung-region extraction prior to ViT-based classification. Table 4 shows the state-of-the-art comparison.

Table 4. State-of-the-art comparison.

Study	Modality	Classification/Task	Model	Reported Result	Comparison with Proposed Work
[73]	Chest CT	Normal/benign/malignant	EfficientNet-B0	Accuracy 99.39%; macro-F1 0.988; AUC 0.996	Highly relevant multiclass comparator
[74]	CT	Lung cancer detection	Vision Transformer	—	Directly relevant ViT study
[12]	CT	Pulmonary nodule classification	Swin Transformer	Accuracy 97.4%	Relevant transformer study; different experimental setting
[14]	CT	Lung nodule classification	CNN + Swin Transformer	Accuracy 90.96%	Relevant transformer study; different experimental setting
[75]	Radiograph	Tuberculosis diagnosis	Transfer learning	—	Background only; not a direct comparator
Proposed framework	CT	Binary + normal/benign/malignant	FNLM + lung-region segmentation + ViT	98.17% binary; 95.43% multiclass	—

4.4. Qualitative Analysis of Preprocessing Effects and Failure Cases

To complement the quantitative ablation analysis, representative CT images were examined to investigate how the preprocessing stages affected classification. For each selected case, the original CT image, FNLM-denoised image, lung-region mask, segmented lung image, and resulting classification were compared. Cases were selected to represent both successful and unsuccessful outcomes rather than only correctly classified examples. In successful cases, FNLM reduced visible image noise while retaining the principal pulmonary structures, whereas lung-region masking suppressed surrounding non-pulmonary image content. For some images, this resulted in a change from an incorrect baseline ViT prediction to the correct classification when the complete preprocessing pipeline was applied. However, the benefit was not uniform across all images. Failure cases were observed when diagnostic appearances were visually similar across classes or when threshold-based lung-region extraction incompletely preserved relevant pulmonary structures. These observations indicate that the preprocessing stages can improve the classifier input in selected cases but cannot be assumed to provide uniform benefit across all CT images.

4.5. Methodological Contribution and Distinction from Existing Architectures

The proposed framework should not be interpreted as a new transformer architecture. ViT and its self-attention mechanism are established methods, and the present study does not modify the internal transformer encoder, introduce a new attention operation, or propose a new optimization objective. Similarly, FNLM denoising, intensity thresholding, morphological processing, and contour extraction are established image-processing techniques. Consequently, the methodological contribution of this study does not arise from the novelty of any individual algorithm.

Instead, the study investigates a specific processing hypothesis: whether explicitly reducing CT image noise and restricting the classifier input to anatomically relevant lung regions provides measurable benefit to ViT-based classification. This distinction is important because an end-to-end classifier supplied with the complete CT slice may learn repre-

sentations from both pulmonary and surrounding anatomical structures. The proposed preprocessing strategy explicitly constrains the image content before patch generation and transformer encoding, thereby providing the ViT with a lung-focused representation rather than modifying its internal attention mechanism.

This positioning differs from recent studies that introduce architectural modifications to transformer-based lung analysis. For example, recent approaches have combined CNN and transformer branches or introduced specialized attention/fusion mechanisms to improve local and global feature modeling. Other work has investigated compact ViT variants specifically for CT nodule classification. The present framework does not compete with these studies on architectural novelty; rather, it evaluates whether preprocessing and anatomical restriction can improve the input representation supplied to an established transformer classifier.

The scientific value of such preprocessing cannot be inferred merely from the inclusion of additional processing stages. It must be demonstrated experimentally. For this reason, the revised study uses component-wise ablation in which the classifier is held constant while FNLM denoising and lung-region extraction are selectively enabled or disabled. This design separates improvements associated with preprocessing from improvements that could otherwise result from changing the classifier architecture. Controlled preprocessing ablation has also been used in recent CT research to quantify the contribution of individual image-processing stages.

Accordingly, the contribution of the proposed framework is primarily empirical and integrative rather than theoretical. The study provides a controlled assessment of whether FNLM denoising and explicit lung-region isolation contribute to binary and multiclass CT classification when the downstream ViT is unchanged. This interpretation also establishes an important limitation: the framework should not be described as a fundamentally new neural architecture. Future work could extend the present approach through task-specific transformer modifications, learned ROI extraction, three-dimensional contextual modeling, or joint optimization of segmentation and classification.

5. Limitations and Discussion

5.1. Potential Reasons for Improved Classification Performance Following Lung-Region Segmentation

The improvement in classification performance observed following lung-region segmentation may be attributed to several complementary factors. First, CT images contain substantial anatomical information outside the pulmonary regions, including bones, surrounding soft tissues, and other structures that may not directly contribute to lung cancer classification. When complete CT images are supplied to the classifier, these regions may introduce irrelevant visual patterns and increase the complexity of feature learning. By applying intensity thresholding, morphological refinement, and contour-based lung extraction, the proposed framework suppresses non-pulmonary structures and provides the Vision Transformer with more anatomically focused input images. This may reduce background interference and facilitate the learning of discriminative pulmonary features.

Second, lung-region segmentation may influence the representation-learning process of the Vision Transformer. Since ViT divides an input image into patches and models relationships among them using self-attention, irrelevant background regions may contribute additional patch representations that are not directly related to the classification objective. Restricting the input to the lung regions may reduce the influence of such patches and encourage the model to capture contextual relationships among pulmonary structures, including variations in tissue appearance, lesion morphology, and local intensity patterns.

However, this interpretation represents a plausible mechanism rather than direct evidence of changes in the model's attention distribution.

Third, segmentation may reduce variability associated with surrounding anatomical structures and differences in image positioning, thereby improving the consistency of the visual information presented to the classifier. This anatomical focusing may be particularly useful when distinguishing normal, benign, and malignant CT images, where relevant abnormalities can exhibit subtle morphological and textural differences. Similar benefits of reducing irrelevant background information through segmentation have been discussed in previous pulmonary CT research [76].

The ablation results support the potential importance of lung-region extraction. The baseline ViT achieved an accuracy of 92.40%, whereas the inclusion of lung segmentation without FNLM denoising increased the accuracy to 96.30%, corresponding to an absolute improvement of 3.90 percentage points. In comparison, FNLM denoising alone improved the accuracy to 94.10%, representing an improvement of 1.70 percentage points. The complete framework combining FNLM denoising and lung-region segmentation achieved the highest accuracy of 98.20%. These observations suggest that anatomical region isolation may contribute more strongly than denoising alone under the evaluated conditions, while the combined preprocessing stages provide complementary benefits. Nevertheless, we observed that improvements should be interpreted with caution. The segmentation procedure may occasionally exclude peripheral or pleural-adjacent lesions, particularly when abnormalities extend beyond the extracted lung boundaries. Furthermore, the ablation results do not independently establish the precise mechanisms responsible for the observed performance differences. Since the dataset was partitioned at the CT-slice level rather than the patient level, the findings also require confirmation through patient-disjoint and independent external validation. Future investigations should include quantitative segmentation assessment, lesion-preservation analysis, and attention visualization to better understand the relationship between anatomical region extraction and transformer-based classification performance.

5.2. Limitations and Future Research

The limitations of this study include the unavailability of verified patient-level identifiers for the 1000 CT images used in binary classification. Consequently, patient-independent data partitioning could not be confirmed, which may have resulted in optimistic performance estimates. Future work will prioritize patient-disjoint data partitioning and independent external validation to establish the generalizability of the proposed framework. The present study has several limitations that should be considered when interpreting its findings. First, the experiments were conducted using a single open-access CT image dataset, which may not adequately represent variations in patient populations, scanner manufacturers, acquisition protocols, and clinical environments. Second, the dataset was partitioned at the image level rather than the patient level, and the exact number of distinct patients could not be verified. Consequently, potential patient overlap between the training and testing subsets cannot be excluded, which may lead to optimistic performance estimates. Third, the proposed framework has not been evaluated using an independent external dataset or in a prospective clinical setting. Therefore, the reported classification performance should be regarded as preliminary evidence of feasibility rather than confirmation of clinical reliability or generalizability. Future investigations should incorporate patient-disjoint evaluation, multi-institutional CT datasets, and prospective clinical validation.

The reported binary classification accuracy of 98.20% and multiclass classification accuracy of 94.63% represent image-level classification performance under the adopted

experimental protocol. These results should not be interpreted as independently validated patient-level diagnostic accuracy. The possibility of optimistic performance estimates due to correlations among CT slices remains an important limitation. To address this limitation, future investigations should employ patient-disjoint partitioning based on verified patient or study identifiers, ensuring that all images belonging to the same patient are assigned exclusively to one subset. In addition, independent external validation using CT datasets from different institutions, scanners, and acquisition protocols is required to assess the robustness and generalizability of the proposed framework. These additional validation procedures are essential before drawing conclusions regarding the clinical applicability of the proposed classification system. Importantly, the dataset was partitioned at the image level rather than the patient level, without explicitly ensuring that CT slices originating from the same patient were assigned exclusively to a single subset. Therefore, patient-level independence across the training, validation, and testing subsets cannot be confirmed. This introduces the possibility of information leakage because CT slices belonging to the same patient may share anatomical and acquisition-related characteristics, potentially leading to optimistic estimates of classification performance. Accordingly, the reported results should be interpreted as image-level classification performance under the adopted experimental protocol rather than as independently validated patient-level diagnostic accuracy. Future investigations should employ patient-disjoint partitioning based on verified patient identifiers to prevent overlap among training, validation, and testing subsets. Furthermore, independent external validation using multicenter CT datasets will be necessary to assess the robustness and generalizability of the proposed framework across different patient populations, imaging scanners, and acquisition protocols.

The ablation results indicate numerical improvements in classification accuracy; formal statistical significance testing was not performed because the paired image-level predictions for the evaluated model configurations were unavailable in the retained experimental records. McNemar's test would be appropriate for comparing classification outcomes from models evaluated on the same test samples; however, the required paired prediction data were not available to construct the corresponding contingency tables. Consequently, the statistical significance of the observed differences could not be established. Therefore, the reported accuracy improvements should be interpreted as descriptive findings rather than statistically confirmed evidence of model superiority. Future studies will incorporate patient-disjoint evaluation, repeated experimental runs, confidence-interval estimation, and appropriate paired statistical tests to assess the reliability of these improvements.

6. Conclusions

This study presented a CT-based lung cancer classification framework integrating image preprocessing, lung-region segmentation, and Vision Transformer (ViT)-based classification. Fast Non-Local Means (FNLM) denoising was employed to reduce image noise, followed by intensity-based thresholding, morphological operations, and contour-based refinement to isolate the lung regions and suppress irrelevant surrounding structures. The resulting lung-region images were subsequently analyzed using a Vision Transformer to learn contextual representations for classification. The proposed framework was evaluated under binary and multiclass classification settings using a single open-access CT image dataset. The experimental results demonstrated promising classification performance under the evaluated conditions, suggesting that conventional lung-region extraction combined with transformer-based representation learning may provide a useful approach for CT image classification. However, these findings should be interpreted cautiously because they do not establish the framework's performance across independent patient populations or clinical environments. Several limitations should be acknowledged. First, reliance on

a single open-access dataset restricts the assessment of generalizability across different institutions, CT scanners, imaging protocols, and patient populations. Second, the available dataset information did not establish the exact number of distinct patients, and image-level partitioning was used rather than patient-disjoint partitioning. Consequently, potential overlap of patient-specific information between training and testing subsets cannot be excluded, which may have resulted in optimistic performance estimates. Third, the segmentation stage has not been independently evaluated using quantitative segmentation metrics, and the framework's computational efficiency and clinical utility require further assessment. Future work will prioritize patient-disjoint evaluation, independent external validation using multi-institutional CT datasets, quantitative assessment of lung-region segmentation, model interpretability, and systematic benchmarking of computational requirements. Prospective clinical evaluation will also be necessary before the framework can be considered for routine clinical application. Future work will also investigate extended training schedules, validation-loss-based checkpoint selection, and convergence behavior to determine whether additional optimization can improve model performance.

Author Contributions: Conceptualization, Writing—review and editing, Methodology, Writing—original draft and Software: I.B. Writing—original draft, Writing—review and editing, Supervision, Formal analysis and Supervision: F.A. Formal analysis, Investigation, Resources and Writing—review and editing: I.d.l.T. Validation, Investigation and Writing—review and editing: M.S.G.d.M. Validation, Resources, Investigation, Writing—review and editing and Data curation; E.S.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Fundación Universidad Europea del Atlántico, Calle Isabel Torres 21, 39011 Santander, Spain (VAT: G39764972).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The CT data analyzed in this study were obtained from the publicly available Low Dose CT Image and Projection Data (LDCT-and-Projection-data) collection hosted by The Cancer Imaging Archive (TCIA), DOI: 10.7937/9NPB-2637. The dataset is publicly accessible through TCIA.

Acknowledgments: During the preparation of this manuscript, the authors used Grammarly (version 14.13) and ChatGPT-6 for the purposes of language refinement. The authors have reviewed and edited the output and take full responsibility for the content of this publication.

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

References

1. Bhatia, I.; Aarti. Lung carcinoma detection at premature stage using deep learning techniques. *AIP Conf. Proc.* **2022**, *2576*, 050009. [CrossRef]
2. Ali, H.; Mohsen, F.; Shah, Z. Improving diagnosis and prognosis of lung cancer using vision transformers: A scoping review. *BMC Med. Imaging* **2023**, *23*, 129. [CrossRef] [PubMed]
3. Bhatia, I.; Aarti. Deep Learning-Based Proposed Approach for Low-Dose CT Scan-Based Early Detection of Lung Cancer. In *Proceedings of the International Conference on Soft Computing and Signal Processing*; Springer Nature: Singapore, 2023; pp. 191–201.
4. Bhatia, I.; Aarti; Ansarullah, S.I.; Amin, F.; Alabrah, A. An advanced lung carcinoma prediction and risk screening model using transfer learning. *Diagnostics* **2024**, *14*, 1378. [CrossRef] [PubMed]
5. Zahid, A.B.; Nisa, F.U.; Malik, A.K.; Qamar, N. Lung Cancer Prediction with Machine Learning, Deep Learning and Hybrid Techniques: A Survey. *LabMed* **2026**, *3*, 7. [CrossRef]
6. Prasad, U.; Chakravarty, S.; Mahto, G. Lung cancer detection and classification using deep neural network based on hybrid metaheuristic algorithm. *Soft Comput.-A Fusion Found. Methodol. Appl.* **2024**, *28*, 8579–8602. [CrossRef]
7. Javed, R.; Abbas, T.; Khan, A.H.; Daud, A.; Bukhari, A.; Alharbey, R. Deep learning for lungs cancer detection: A review. *Artif. Intell. Rev.* **2024**, *57*, 197. [CrossRef]

8. Wani, N.A.; Kumar, R.; Bedi, J. DeepXplainer: An interpretable deep learning based approach for lung cancer detection using explainable artificial intelligence. *Comput. Methods Programs Biomed.* **2024**, *243*, 107879. [CrossRef] [PubMed]
9. Sivasankaran, P.; Dhanaraj, K.R. Lung Cancer Detection Using Image Processing Technique Through Deep Learning Algorithm. *Rev. d'Intelligence Artif.* **2024**, *38*, 297–302. [CrossRef]
10. Crasta, L.J.; Neema, R.; Pais, A.R. A novel Deep Learning architecture for lung cancer detection and diagnosis from Computed Tomography image analysis. *Healthc. Anal.* **2024**, *5*, 100316. [CrossRef]
11. Sumon, R.I.; Mazumdar, M.A.I.; Uddin, S.M.I.; Kim, H.C. Exploring Deep Learning and Machine Learning Techniques for Histopathological Image Classification in Lung Cancer Diagnosis. In *Proceedings of the 2024 International Conference on Electrical, Computer and Energy Technologies (ICECET)*; IEEE: New York, NY, USA, 2024; pp. 1–6.
12. Suzon, M.A.; Shohan, M.S.; Islam, M.S.; Abdullah, M.I.; Aktaruzzaman, M.; Rahman, M.H. Explainable hybrid ensemble framework for lung cancer classification using deep CNNs and random forest. *Sci. Rep.* **2026**. [CrossRef]
13. Sangeetha, S.K.B.; Mathivanan, S.K.; Karthikeyan, P.; Rajadurai, H.; Shivahare, B.D.; Mallik, S.; Qin, H. An enhanced multimodal fusion deep learning neural network for lung cancer classification. *Syst. Soft Comput.* **2024**, *6*, 200068. [CrossRef]
14. Diao, S.; Wan, Y.; Huang, D.; Huang, S.; Sadiq, T.; Khan, M.S.; Hussain, L.; Alkahtani, B.S.; Mazhar, T. Optimizing Bi-LSTM networks for improved lung cancer detection accuracy. *PLoS ONE* **2025**, *20*, e0316136. [CrossRef] [PubMed]
15. Bartusik-Aebisher, D.; Justin Raj, D.R.; Aebisher, D. Artificial Intelligence in Medical Diagnostics: Foundations, Clinical Applications, and Future Directions. *Appl. Sci.* **2026**, *16*, 728. [CrossRef]
16. Yousafzai, S.N.; Nasir, I.M.; Mansour, S.; Negm, N.; Alhashmi, A.A.; Alharbi, M.A.; Kim, E. A hybrid deep learning approach integrating CNN and transformer for lung cancer classification using CT scans. *Sci. Rep.* **2026**, *16*, 15420. [CrossRef] [PubMed]
17. Gautam, N.; Basu, A.; Sarkar, R. Lung cancer detection from thoracic CT scans using an ensemble of deep learning models. *Neural Comput. Appl.* **2024**, *36*, 2459–2477. [CrossRef]
18. Mehta, S.; Bordoloi, D.; Yadav, S.P.; Singh, P.K.; Upadhyay, P. Detection of lung cancer using image processing methods. In *A Practitioner's Approach to Problem-Solving Using AI*; Bentham Science Publishers: Singapore, 2024; pp. 88–103.
19. Bharathi, P.S.; Shalini, C. Advanced hybrid attention-based deep learning network with heuristic algorithm for adaptive CT and PET image fusion in lung cancer detection. *Med. Eng. Phys.* **2024**, *126*, 104138. [CrossRef] [PubMed]
20. Shatnawi, M.Q.; Abuein, Q.; Al-Quraan, R. Deep learning-based approach to diagnose lung cancer using CT-scan images. *Intell.-Based Med.* **2025**, *11*, 100188. [CrossRef]
21. Lee, M.-R.; Kao, M.-H.; Hsieh, Y.-C.; Sun, M.; Tang, K.-T.; Wang, J.-Y.; Ho, C.-C.; Shih, J.-Y.; Yu, C.-J. Cross-site validation of lung cancer diagnosis by electronic nose with deep learning: A multicenter prospective study. *Respir. Res.* **2024**, *25*, 203. [CrossRef] [PubMed]
22. Eldho, K.J.; Nithyanandh, S. Lung cancer detection and severity analysis with a 3D deep learning CNN Model using CT-DICOM clinical dataset. *Indian J. Sci. Technol.* **2024**, *17*, 899–910. [CrossRef]
23. Binson, V.A.; Subramoniam, M.; Mathew, L. Prediction of lung cancer with a sensor array based e-nose system using machine learning methods. *Microsyst. Technol.* **2024**, *30*, 1421–1434. [CrossRef]
24. Sultana, Z.; Foysal, M.; Islam, S.; Al Foysal, A. Lung cancer detection and classification from chest CT images using an ensemble deep learning approach. In *Proceedings of the2024 6th International Conference on Electrical Engineering and Information & Communication Technology (ICEEICT)*; IEEE: New York, NY, USA, 2024; pp. 364–369.
25. Saravanaprasad, P.; Karuppusamy, S.A. Advanced lung tumor diagnosis using a 3D deep neural network based CAD system. *Biomed. Signal Process. Control* **2024**, *89*, 105650. [CrossRef]
26. Bhatia, I.; Aarti; Ansarullah, S.I.; Amin, F.; Alabrah, A. Lightweight Advanced Deep Neural Network (DNN) Model for Early-Stage Lung Cancer Detection. *Diagnostics* **2024**, *14*, 2356. [CrossRef] [PubMed]
27. Meeradevi, T.; Sasikala, S.; Murali, L.; Manikandan, N.; Ramaswamy, K. Lung cancer detection with machine learning classifiers with multi-attribute decision-making system and deep learning model. *Sci. Rep.* **2025**, *15*, 8565. [CrossRef] [PubMed]
28. Wang, X.; Sharpnack, J.; Lee, T.C. Improving lung cancer diagnosis and survival prediction with deep learning and CT imaging. *PLoS ONE* **2025**, *20*, e0323174. [CrossRef] [PubMed]
29. Ansari, M.M.; Kumar, S.; Bin Heyat, B.; Ullah, H.; Bin Hayat, M.A.; Parveen, S.; Ali, A.; Zhang, T. SVMVGGNet-16: A Novel Machine and Deep Learning Based Approaches for Lung Cancer Detection using Combined SVM and VGGNet-16. *Curr. Med. Imaging* **2025**, *21*, e15734056348824. [CrossRef] [PubMed]
30. Sankar, B.S.; Natarajasivan, D.; Reddy, M.P. Enhanced lung cancer detection using BiLSTM attention and transformer models on clinical text data. *Sci. Rep.* **2026**, *16*, 27546. [CrossRef] [PubMed]
31. Debnath, R.; Mondal, R.; Chakraborty, A.; Chatterjee, S. Advances in Artificial Intelligence for Lung Cancer Detection and Diagnostic Accuracy: A Comprehensive Review. *Int. J. Innov. Sci. Res. Technol.* **2025**, *10*, 1579–1586. [CrossRef]
32. Ozdemir, B.; Aslan, E.; Pacal, I. Attention enhanced inceptionnext based hybrid deep learning model for lung cancer detection. *IEEE Access* **2025**, *13*, 27050–27069. [CrossRef]

33. Ukwuoma, C.C.; Cai, D.; Eziefuna, E.O.; Oluwasanmi, A.; Abdi, S.F.; Muoka, G.W.; Thomas, D.; Sarpong, K. Enhancing histopathological medical image classification for Early cancer diagnosis using deep learning and explainable AI–LIME & SHAP. *Biomed. Signal Process. Control* **2025**, *100*, 107014. [CrossRef]
34. Bherje, A.; Jidge, A.; Roy, C.; Hulke, A.; Aswathy, M.A.; Yadav, V.; Veenamol, K.V. Design of Deep Learning-based Approach to Predict Lung Cancer on CT Scan Images. In Proceedings of the 2024 5th International Conference on Innovative Trends in Information Technology (ICITIIT), Kottayam, India, 15–16 March 2024.
35. Mehrzadi, H.; Maghanaki, M.; Shahin, M.; Chen, F.F.; Hosseinzadeh, A.; Tarhani, S. Enhanced lung cancer detection in CT scans using deep learning architectures. In *Proceedings of the Third International Conference on Advances in Computing Research (ACR'25)*; International Conference on Advances in Computing Research; Springer Nature: Cham, Switzerland, 2025; pp. 127–135.
36. Cay, T. Lung cancer diagnosis with GAN supported deep learning models. *Bio-Med Mater. Eng.* **2025**, *36*, 209–211. [CrossRef] [PubMed]
37. Dawood, H.; Nawaz, M.; Ilyas, M.U.; Nazir, T.; Javed, A. Attention-guided CenterNet deep learning approach for lung cancer detection. *Comput. Biol. Med.* **2025**, *186*, 109613. [CrossRef] [PubMed]
38. Golkarieha, A.; Kiashemshakib, K.; Boroujenic, S.R.; Isakand, N.A. Advanced U-net architectures with CNN backbones for automated lung cancer detection and segmentation in chest CT images. *arXiv* **2025**, arXiv:2507.09898.
39. Shariff, V.; Paritala, C.; Ankala, K.M. Optimizing non small cell lung cancer detection with convolutional neural networks and differential augmentation. *Sci. Rep.* **2025**, *15*, 15640. [CrossRef] [PubMed]
40. Priya, A.; Bharathi, P.S. SE-ResNeXt-50-CNN: A deep learning model for lung cancer classification. *Appl. Soft Comput.* **2025**, *171*, 112696. [CrossRef]
41. Tawfeek, M.A.; Alrashdi, I.; Alruwaili, M.; Shaban, W.M.; Talaat, F.M. Enhancing the efficiency of lung cancer screening: Predictive models utilizing deep learning from CT scans. *Neural Comput. Appl.* **2025**, *37*, 11459–11477. [CrossRef]
42. Ejiyi, C.J.; Qin, Z.; Agbesi, V.K.; Yi, D.; Atwereboannah, A.A.; Chikwendu, I.A.; Bamisile, O.F.; Kissanga, G.-M.B.; Bamisile, O.O. Advancing cancer diagnosis and prognostication through deep learning mastery in breast, colon, and lung histopathology with ResoMergeNet. *Comput. Biol. Med.* **2025**, *185*, 109494. [CrossRef] [PubMed]
43. Ali, F.; Almuhaimeed, A.; Alghamdi, W.; Aldossary, H.; Asiry, O.; Masmoudi, A. Leveraging deep learning for epigenetic protein prediction: A novel approach for early lung cancer diagnosis and drug discovery. *Health Inf. Sci. Syst.* **2025**, *13*, 28. [CrossRef] [PubMed]
44. Nguyen, M.N. A scoping review of deep learning approaches for lung cancer detection using chest radiographs and computed tomography scans. *Biomed. Eng. Adv.* **2025**, *9*, 100138. [CrossRef]
45. Liz-Lopez, H.; de Sojo-Hernández, Á.A.; D'Antonio-Maceiras, S.; Diaz-Martinez, M.A.; Camacho, D. Deep learning innovations in the detection of lung cancer: Advances, trends, and open challenges. *Cogn. Comput.* **2025**, *17*, 67. [CrossRef]
46. Pathipati, H.; Ramisetti, L.N.B.; Reddy, D.N.; Pesaru, S.; Balakrishna, M.; Anitha, T. Optimizing cancer detection: Swarm algorithms combined with deep learning in colon and lung cancer using biomedical images. *Diyala J. Eng. Sci.* **2025**, *18*, 91–102. [CrossRef]
47. Li, X.; Li, X.; Qin, J.; Lei, L.; Guo, H.; Zheng, X.; Zeng, X. Machine learning-derived peripheral blood transcriptomic biomarkers for early lung cancer diagnosis: Unveiling tumor-immune interaction mechanisms. *BioFactors* **2025**, *51*, e2129. [CrossRef] [PubMed]
48. Aishwarya, K.V.; Asuntha, A. A survey on comparative study of lung nodules applying machine learning and deep learning techniques. *Multimed. Tools Appl.* **2025**, *84*, 2127–2181. [CrossRef]
49. Mahmud, M.R.; Fardin, H.; Siddiqui, M.I.H.; Sakib, A.H.; Sakib, A.A. Hybrid deep learning for interpretable lung cancer recognition across computed tomography and histopathological imaging modalities. *Int. J. Sci. Res. Arch.* **2025**, *15*, 1798–1810. [CrossRef]
50. Gokavarapu, S.; Rao, K.V.; Srinivas, G.; Krishna, D.S. Enhancing lung cancer detection: Optimizing deep learning with convolutional block attention module. *J. Artif. Intell. Technol.* **2025**, *5*, 143–149. [CrossRef]
51. Ahmed, K.; Ahmed, S.S.; Talukdar, A.; Chakrabarty, D. An empirical study on lung cancer detection and classification using machine learning and image processing techniques. In *Revolutionizing Healthcare 5.0: The Power of Generative AI: Advancements in Patient Care Through Generative AI Algorithms*; Springer Nature: Cham, Switzerland, 2025; pp. 165–176.
52. Usharani, C.; Selvapandian, A. FedLRes: Enhancing lung cancer detection using federated learning with convolution neural network (ResNet50). *Neural Comput. Appl.* **2025**, *37*, 8273–8284. [CrossRef]
53. Mannepalli, D.; Tak, T.K.; Krishnan, S.B.; Sreenivas, V. GSC-DVIT: A vision transformer based deep learning model for lung cancer classification in CT images. *Biomed. Signal Process. Control* **2025**, *103*, 107371. [CrossRef]
54. Kashyap, M.; Wang, X.; Panjwani, N.; Hasan, M.; Zhang, Q.; Huang, C.; Bush, K.; Chin, A.; Vitzthum, L.K.; Dong, P.; et al. Automated deep learning–based detection and segmentation of lung tumors at ct imaging. *Radiology* **2025**, *314*, e233029. [CrossRef] [PubMed]
55. Neoaz, N.; Amin, M.H. Leveraging Artificial Intelligence for Early Lung Cancer Detection through Advanced Imaging Analysis. *Glob. J. Comput. Sci. Artif. Intell.* **2025**, *1*, 55–65. [CrossRef]

56. Shahriyar, O.; Moghaddam, B.N.; Yousefi, D.; Mirzaei, A.; Hoseini, F. An analysis of the combination of feature selection and machine learning methods for an accurate and timely detection of lung cancer. *arXiv* **2025**, arXiv:2501.10980.
57. Hammad, M.; ElAffendi, M.; El-Latif, A.A.A.; Ateya, A.A.; Ali, G.; Plawiak, P. Explainable AI for lung cancer detection via a custom CNN on CT images. *Sci. Rep.* **2025**, *15*, 12707. [CrossRef] [PubMed]
58. Vyas, A.; Kumar, K.; Sharma, A.; Verma, D.; Bhatia, D.; Wahi, N.; Yadav, A.K. Advancing the frontier of artificial intelligence on emerging technologies to redefine cancer diagnosis and care. *Comput. Biol. Med.* **2025**, *191*, 110178. [CrossRef] [PubMed]
59. Tidke, D.; Banait, S.S. Harnessing Deep Learning for Lung Cancer Detection and Prevention: A Comprehensive Survey. *EPJ Web Conf.* **2025**, *328*, 01007. [CrossRef]
60. Saxena, S.; Prasad, S.N.; Polnaya, A.M.; Agarwala, S. Hybrid deep convolution model for lung cancer detection with transfer learning. *arXiv* **2025**, arXiv:2501.02785.
61. Saha, N.; Mondal, R.; Banerjee, A.; Debnath, R.; Chatterjee, S. Advanced Deep Lung Care Net: A Next Generation Framework for Lung Cancer Prediction. *Int. J. Innov. Sci. Res. Technol.* **2025**, *10*, 2312–2320. [CrossRef]
62. Kavitha, S.; Patnala, E.; Sangaraju, H.R.; Bingu, R.; Adinarayana, S.; Dhatterwal, J.S. An optimized multi-head attention based fused depthwise convolutional model for lung cancer detection. *Expert Syst. Appl.* **2025**, *271*, 126596. [CrossRef]
63. Bhukya, S.; Ganagoni, V.; Nangunoori, S.S.; Enapothula, S.T. A Deep Learning Framework Using Enhanced Convolutional Neural Network for Detection of Lung Cancer from CT Images. *Int. Res. J. Multidiscip. Technovation* **2025**, *7*, 138–150. [CrossRef]
64. Tu, H.; Zhao, Y.; Cui, J.; Lu, W.; Sun, G.; Xu, X.; Hu, Q.; Hu, K.; Wu, M.; Wu, X. Improving Lung Cancer Risk Prediction Using Machine Learning: A Comparative Analysis of Stacking Models and Traditional Approaches. *Cancers* **2025**, *17*, 1651. [CrossRef] [PubMed]
65. Mehrnia, S.S.; Safahi, Z.; Mousavi, A.; Panahandeh, F.; Farmani, A.; Yuan, R.; Rahmim, A.; Salmanpour, M.R. Landscape of 2D Deep Learning Segmentation Networks Applied to CT Scan from Lung Cancer Patients: A Systematic Review. *J. Imaging Inform. Med.* **2025**, *38*, 3711–3740. [CrossRef] [PubMed]
66. Alavinejad, M.; Shirzad, M.; Javid-Naderi, M.J.; Rahdar, A.; Fathi-Karkan, S.; Pandey, S. Smart nanomedicines powered by artificial intelligence: A breakthrough in lung cancer diagnosis and treatment. *Med. Oncol.* **2025**, *42*, 134. [CrossRef] [PubMed]
67. Tusher, M.I.; Hossain, R.; Akter, A.; Mahin, R.H.; Akhi, S.S.; Chy, S.K.; Haider, M.; Akter, S.; Hasan, M.; Shaima, M. Deep learning meets early diagnosis: A hybrid CNN-DNN framework for lung cancer prediction and clinical translation. *Int. J. Med. Sci. Public Health Res.* **2025**, *6*, 63–72. [CrossRef]
68. Gupta, A.; Kumar, A.; Rautela, K. UDCT: Lung Cancer detection and classification using U-net and DARTS for medical CT images. *Multimed. Tools Appl.* **2025**, *84*, 19065–19085. [CrossRef]
69. Henriksen, M.B.; Van Daalen, F.; Wee, L.; Hansen, T.F.; Jensen, L.H.; Brasen, C.L.; Hilberg, O.; Bermejo, I. Lung Cancer Detection Using Bayesian Networks: A Retrospective Development and Validation Study on a Danish Population of High-Risk Individuals. *Cancer Med.* **2025**, *14*, e70458. [CrossRef] [PubMed]
70. Ochoa-Ornelas, R.; Gudiño-Ochoa, A.; García-Rodríguez, J.A.; Uribe-Toscano, S. A robust transfer learning approach with histopathological images for lung and colon cancer detection using EfficientNetB3. *Healthc. Anal.* **2025**, *7*, 100391. [CrossRef]
71. Chen, H.; Liu, H.; Xing, L.; Fan, D.; Chen, N.; Ma, P.; Zhang, X. Deep Learning-driven Microfluidic-SERS to Characterize the Heterogeneity in Exosomes for Classifying Non-Small Cell Lung Cancer Subtypes. *ACS Sens.* **2025**, *10*, 2872–2882. [CrossRef] [PubMed]
72. Tan, S.L.; Selvachandran, G.; Paramesran, R.; Ding, W. Lung cancer detection systems applied to medical images: A state-of-the-art survey. *Arch. Comput. Methods Eng.* **2025**, *32*, 343–380. [CrossRef]
73. Rahman, R.; Sultana, S.; Das, M.; Aziz Susom, A.; Sabbirul Mohosin Naim, S.A.; Bin Sharif, N. Lung Cancer Classification from Chest CT Scans using EfficientNet. *Int. J. Comput. Appl.* **2026**, *187*, 35–41. [CrossRef]
74. Jessika, S.; Panneerselvam, R.; M, T.; C, G.; B, N.; S, S. Vision Transformer-Based Framework for Enhanced Lung Cancer Detection in CT Images. In Proceedings of the 2025 2nd Global AI Summit—International Conference on Artificial Intelligence and Emerging Technology (AI Summit), Noida, India, 19–21 November 2025; pp. 555–560. [CrossRef]
75. Ahmad, K.F.; Vashisth, A.; Kaur, G.; Bakhshizada, R. A Transfer Learning-based Framework for Early Diagnosis of Tuberculosis using Radiographic Images. In Proceedings of the International Conference on Cognitive Computing and Networking Systems (ICC-CNS), Guntur, India, 11–13 June 2026; pp. 472–477.
76. Pawar, S.P.; Talbar, S.N. Multi-level deep learning based lung cancer classifier for classification based on tumour-node-metastasis approach. *Int. J. Imaging Syst. Technol.* **2023**, *33*, 881–894. [CrossRef]

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