Future Technology
Open Access Journal
ISSN 2832-0379

November 2026| Volume 05 | Issue 04 | Pages 299-309

Journal homepage: https://fupubco.com/futech

https://doi.org/10.55670/fpll.futech.5.4.24

Future Publishing LLC

Article

Deep transfer learning and radiomics synergy for enhanced lung cancer classification on CT scans

Mohammed S. H. Al-Tamimi[1]*, Amenah H. Abdulateef[2], Isra H. Abdulateef[1]

[1]Computer Science Department, College of Science, University of Baghdad, Baghdad, Iraq
[2]Department of Biology, College of Science for Women, University of Baghdad, Baghdad, Iraq

ARTICLE INFO

Article history:
Received 09 April 2026
Received in revised form
05 September 2026
Accepted 23 September 2026

Keywords:
Detection, Lung cancer, EfficientNet, Features, Image processing

*Corresponding author
Email address:
mohammed.s@sc.uobaghdad.edu.iq

DOI: 10.55670/fpll.futech.5.4.24

ABSTRACT

Lung cancer remains one of the leading causes of cancer-related mortality worldwide, making accurate and timely detection essential for improving patient outcomes. Computed tomography (CT) is widely used for lung assessment; however, manual interpretation can be affected by inter-observer variability and subtle imaging characteristics. This study proposes an automated three-class lung CT classification framework that integrates deep transfer learning with quantitative lung-region features. EfficientNet-B7, initialized with ImageNet-pretrained weights, was employed to extract high-level deep representations from CT images. In parallel, lung Regions of Interest (ROIs) were automatically segmented, from which four quantitative descriptors were extracted: Skewness, Kurtosis, GLCM Contrast, and GLCM Homogeneity. The standardized lung-ROI features were subsequently concatenated with the EfficientNet-B7 deep feature representation using a feature-level fusion strategy and classified into Normal, Benign, and Malignant categories. Experimental evaluation on the held-out test set demonstrated an overall classification accuracy of 96.36%, correctly classifying 106 of 110 CT images. The proposed fusion framework achieved macro-averaged Precision, Recall, F1-score, and Specificity of 94.39%, 90.87%, 92.39%, and 97.53%, respectively. Furthermore, one-vs-rest ROC analysis yielded AUC values of 0.9849, 0.9634, and 1.0000 for the Normal, Benign, and Malignant classes, respectively, corresponding to a macro-average AUC of 0.9828. These results demonstrate the strong discriminative capability of the proposed deep and quantitative feature fusion framework for three-class lung CT image classification.

1. Introduction

Lung cancer remains the leading cause of cancer-related mortality in the world [1]. It is a difficult tumor to treat and manage because it is often diagnosed late and occurs at high prevalence. Early diagnosis is essential to improving clinical outcomes, but conventional diagnostic methods, including imaging and needle biopsy, are invasive and have limited sensitivity [2]. Advances in medical technology and research through October 2023 have introduced new screening and detection techniques for lung cancer, including biomarkers, liquid biopsy, and AI-based technologies. These advances help to make a diagnosis more accurate and to identify lung cancer at an early stage, before it can be too late to be turned around. One of these developments that has impacted medical image processing and interpretation in this area is Deep Learning (DL), which is a subfield of Artificial Intelligence. Deep Learning Models (DLMs) education of October 2023 multi-layered Convolutional Neural Networks (CNNs) may automatically build and extract complex high-level data [3] as well as low-texture patterns automatically from CT cloud scans [4]. This computation power essentially may reduce subjectivity and inter-observer variability in manual radiological interpretations [5]. The development of automated diagnosis of lung cancer has advanced from simple lung nodule segmentation to clinical decision support systems. In the modern era, deep learning architectures can predict micro-nodules with high sensitivity and even identify malignancy risk and the histological type of micro-nodules [6] directly from non-invasive imaging data. Nevertheless, many techniques have been developed that train such deep architectures from scratch [7], and there is increasing evidence that training such deep models requires a large collection of annotated medical data that is often unavailable [8]. The limitation was removed by the paradigm of Transfer Learning (TL) that we will discuss next. This enables most models to converge rapidly and generalize better by learning general feature extraction from large-scale datasets with more general knowledge (e.g., ImageNet) and fine-tuning the

model for the specific small niche medical image domain [9]. Recent evaluations demonstrate the impressive performance of compound-scaled architectures like the EfficientNet family, in terms of computing efficiency and disease classification accuracy. EfficientNet-B7's topology is efficient and can effectively integrate small, heterogeneous features of early-stage lung cancers by adjusting depth, width, and resolution simultaneously, while keeping computation low enough for clinical practice [10]. The article points out that one of the main improvements in lung cancer detection is the ability to combine medical imaging and advanced artificial intelligence through transfer learning [11]. The main objective of this study is to develop an accurate three-class lung CT classification framework for distinguishing Normal, Benign, and Malignant images. The proposed framework incorporates an integrated image preprocessing pipeline based on Non-Local Means (NLM) denoising, Contrast Limited Adaptive Histogram Equalization (CLAHE), and unsharp masking to improve CT image quality. EfficientNet-B7-based transfer learning is employed to extract discriminative deep representations, while quantitative image descriptors, including Skewness, Kurtosis, GLCM Contrast, and GLCM Homogeneity, are extracted from automatically segmented lung ROIs. These quantitative lung-ROI features are integrated with the EfficientNet-B7 deep representations through a feature-level fusion strategy to provide complementary information for three-class classification. Class-weighted learning addresses class imbalance, and classification performance is evaluated using accuracy, precision, recall, F1-score, specificity, confusion matrix analysis, and one-vs-rest ROC-AUC.

2. Related work

Several studies have investigated deep learning and machine learning approaches for lung cancer classification using CT images. These studies can be broadly categorized into transfer learning and CNN-based approaches, hybrid feature-based methods, and advanced ensemble or transformer-based architectures. Transfer learning and CNN-based approaches have demonstrated promising performance in lung cancer classification. Suresha et al. [12] employed the pretrained EfficientNet-B7 architecture for feature extraction and combined deep and statistical features to address class imbalance. The resulting feature representation was subsequently classified using a Support Vector Machine (SVM), demonstrating the effectiveness of pretrained deep features for lung cancer classification. Similarly, Qadir et al. [13] proposed the Hybrid Lung Cancer Staging Classifier and Diagnostic Model (HybridLCSCDM) for classifying CT images into Normal, Benign, and Malignant categories. Their approach utilized pretrained VGG-16 for feature extraction followed by XGBoost for three-class classification on the IQ-OTH/NCCD dataset.

More recent studies have explored advanced hybrid and ensemble architectures. Gulsoy and Kablan [14] introduced FocalNeXt, a hybrid architecture that combines characteristics of ConvNeXt and FocalNet for automated lung cancer diagnosis from chest CT images. The model integrates convolution-based feature extraction with an attention-based mechanism and achieved a reported classification accuracy of 98%. The authors also conducted comparative and ablation analyses to evaluate the proposed architecture's contribution. Majumder et al. [15] proposed an ensemble network, MENet, that combines Xception, InceptionResNetV2, and MobileNetV2. The predictions of the individual networks were integrated using a fuzzy-ranking strategy based on the Mitscherlich function. The approach was evaluated using the LIDC-IDRI and IQ-OTH/NCCD lung CT datasets. Image preprocessing has also been investigated as an important component of CT-based lung cancer analysis. Al-Tamimi et al. [16] examined how image-processing techniques affect automated lung cancer identification using the IQ-OTH/NCCD dataset. Their approach combined a pretrained U-Net model with Difference of Gaussian (DoG) and CLAHE techniques, and compared training strategies with and without image preprocessing. The results demonstrated the potential contribution of image enhancement techniques to improving model performance.

In contrast, this study integrates an enhanced preprocessing pipeline with EfficientNet-B7-based deep feature extraction and quantitative radiomic descriptors within a unified lung CT classification framework. While previous studies have primarily emphasized pretrained deep features, hybrid classifier architectures, transformer-based models such as FocalNeXt, or ensemble learning such as MENet, the present framework is designed to exploit complementary information from enhanced CT appearance, high-level deep representations, and quantitative radiomic characteristics for three-class classification into Normal, Benign, and Malignant categories. This integration constitutes the principal methodological distinction of the proposed approach from the closest existing methods.

3. Method of proposed approaches

Efficient medical image inputs to DLMs are one of the most important approaches to achieving the final target of tumor imaging. Because CT scans frequently suffer from speckle or Gaussian noise, as well as low contrast between tissue and surrounding tissues, an intelligent, innovative preprocessing pipeline was designed for us. Specifically, it consists of three axes running sequentially: noise elimination while maintaining full image quality; multi-contrast enhancement; and high-resolution tumor imaging.

3.1 Dataset

The IQ-OTH/NCCD lung cancer dataset was collected over three months in fall 2019 at the Iraq-Oncology Teaching Hospital and the National Center for Cancer Diseases [17]. The dataset contains chest CT scans from 110 cases representing three diagnostic categories: Normal, Benign, and Malignant. According to the original dataset description, the complete collection comprises 1,190 CT images, with the 110 cases distributed as 55 Normal, 15 Benign, and 40 Malignant cases. The CT scans were originally acquired in DICOM format using a Siemens SOMATOM CT scanner. The acquisition protocol included a tube voltage of 120 kV and a slice thickness of 1 mm, with a window width ranging from 350 to 1200 HU and a window center ranging from 50 to 600 HU. The scans were acquired during breath-holding at maximum inspiration. Each case comprises multiple CT slices, with the number of slices varying across cases. The images were anonymized before inclusion in the dataset, and oncologists and radiologists at the participating medical centers annotated the original data.

Although the complete IQ-OTH/NCCD dataset is reported to contain 1,190 CT images, the accessible archive used in the present experiments contained 1,097 image files. Accordingly, all experimental results reported in this study were obtained using these 1,097 accessible images. A stratified image-level partitioning strategy was applied, with 90% of the images assigned to the development set and 10% reserved as an independent held-out test set. The held-out test set contained 110 CT images, comprising 42 Normal, 12 Benign, and 56 Malignant images. Patient identifiers were not explicitly available in the accessible image archive; therefore, patient-level grouping could not be reliably performed, and the experimental partition should be considered a slice-level (image-level) split. Consequently, complete separation of CT slices originating from the same patient across the development and test sets cannot be guaranteed, and this is acknowledged as a limitation of the experimental design [17]. Table 1 summarizes the main characteristics and distribution of the dataset used in this study.

Table1. IQ-OTH/NCCD dataset details [17]

Feature	Description
Dataset Name	IQ-OTH/NCCD Lung Cancer Dataset
Full Name	Iraq-Oncology Teaching Hospital / National Center for Cancer Diseases Dataset
Imaging Modality	CT Scan
Purpose	Lung Cancer Classification and Detection
Number of Classes	3 Classes
Classes	Normal, Benign, Malignant
Total Cases	110 Patients
Total Images	Approximately 1190 CT Images
Normal Cases	55 Cases
Benign Cases	15 Cases
Malignant Cases	40 Cases
Image Format	DICOM (.dcm), PNG/JPG
Slice Thickness	1 mm
Scanner Type	Siemens SOMATOM CT Scanner
Dataset Source	Iraq Oncology Teaching Hospital and National Center for Cancer Diseases
Common Applications	Deep Learning, CNN Classification, Transfer Learning
Frequently Used Models	ResNet50, VGG16, DenseNet121, MobileNet, EfficientNet
Advantages	Moderate size, publicly available, suitable for medical AI research
Limitations	Small dataset size, class imbalance, mostly used as 2D slices
Recommended Techniques	Data Augmentation, Cross Validation, Image Enhancement
Official Sources	Kaggle and Mendeley Data Repository

3.2 Preprocessing methodology

The efficiency of medical image inputs to DLMs is a crucial factor in achieving the outcome for tumor imaging. Since CT scans often suffer from speckle or Gaussian noise, as well as poor contrast between the scan and surrounding tissues, an intelligent and innovative preprocessing pipeline has been developed. This pipeline operates on three consecutive axes: noise removal while preserving image quality, multiple adaptive contrast adjustments, and detailed tumor imaging. This dataset contains the following types of noise, as shown in Table 2.

Table 2. Noise summarization [18]

Noise Type	Description	Impact on CT Images
Gaussian Noise	Random intensity variations caused by imaging sensors and reconstruction processes	Reduces image clarity and affects feature extraction
Speckle Noise	Multiplicative noise caused by X-ray scattering and tissue interaction	Produces grainy appearance and obscures tumor boundaries
Quantum Noise	Caused by insufficient photon detection in CT imaging systems	Leads to blurred structures and reduced diagnostic visibility
Reconstruction Artifacts	Distortions produced during CT image reconstruction algorithms	Creates false structures and affects segmentation accuracy
Low Contrast Problem	Weak intensity difference between tumor tissue and healthy lung tissue	Makes tumor detection and classification more difficult

Edge-preserving denoising: To overcome the limitations of conventional denoising filters, such as median filtering, which may cause undesirable edge blurring, an optimized NLM denoising technique was employed. Unlike conventional local filters that average pixels within a predefined neighborhood, the NLM algorithm exploits the redundancy of similar structures across the entire image. Given a noise-contaminated image I, the restored intensity value of a pixel located at site i, denoted by $NL(I)(i)$, is computed as a weighted average of the intensities of all pixels in the image, as follows [19]:

$$NL(I)(i) = \sum_{j \in I} w(i,j) I(j) \quad (1)$$

where $I(j)$ denotes the intensity value of the pixel at location j, $w(i,j)$ represents the normalized similarity weight between the pixels at locations i and j, and $NL(I)(i)$ is the restored intensity value at location i. The weights satisfy $0 \leq w(i,j) \leq 10$ and are normalized such that $\sum_{j \in I} w(i,j) - 1$.

The weight $w(i, j)$ quantifies the structural similarity between the neighborhood patch centered at pixel i and the patch centered at pixel j. It is mathematically defined as follows:

$$w(i,j) = \frac{1}{Z(i)} \exp\left(-\frac{\|V(N_i) - V(N_j)\|^2_{2,a}}{h^2}\right) \quad (2)$$

where:

- $V(N_i)$ denotes the intensity vector of the pixels falling within the neighborhood patch N_i
- h represents the filtering parameter that controls the degree of smoothing.
- $Z(i)$ is the normalization factor, formulated as:

$$Z(i) = \sum_j \exp(-\frac{\|V(N_i)-V(N_j)\|^2_{2,a}}{h^2}) \quad (3)$$

The mathematical formulation of robust noise suppression inside homogeneous regions (e.g., air density, the inner core of the tumor), while stopping smoothing at structural boundaries, avoids loss of true dimensional measurements of the lung nodule [20] after applying the NLM filter using Clinical Significance. The effective filter on the original image is shown in the following Figure 1.

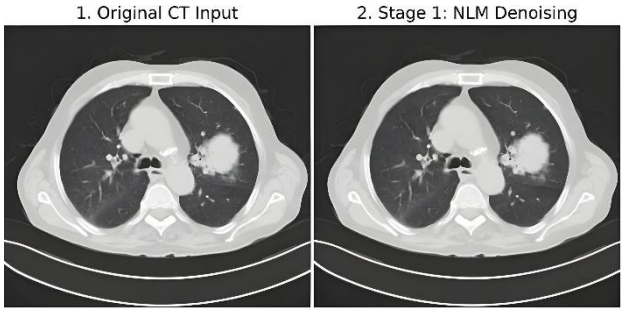


Figure 1. Original image and NLM filter denoising

Contrast-limited adaptive histogram equalization (CLAHE): Lung CT scans typically exhibit poor global contrast. To highlight the distinct texture features of the tumor mass, the CLAHE technique was implemented. CLAHE differs from Global Histogram Equalization by partitioning the image into small, non-overlapping contextual regions called "tiles" of size *(M×N)*. For each tile, a local histogram is computed. To prevent noise amplification in uniform areas, a clip limit β is enforced. Mathematically, if the bin count *H(k)* exceeds the threshold β the excess pixels are clipped and distributed equally across all histogram bins [16,21]:

$$\tilde{H}(k) = \begin{cases} \beta & if\ H(k) \geq \beta \\ H(k) & jf\ H(k) < \beta \end{cases} \quad (4)$$

Where β represents the uniformly distributed excess pixel count. Subsequently, the modified Cumulative Distribution Function (CDF) is mapped to transform the pixel intensities. Finally, bilinear interpolation is applied between adjacent tiles to eliminate artificial block boundaries. This step enhances micro-calcifications and the spiculed margins of malignant tumors without causing over-saturation [22,23], allowing the deep learning model to capture subtle pathological characteristics. The following Figure 2 shows how the CLAHE filter is applied to the NLM filter.

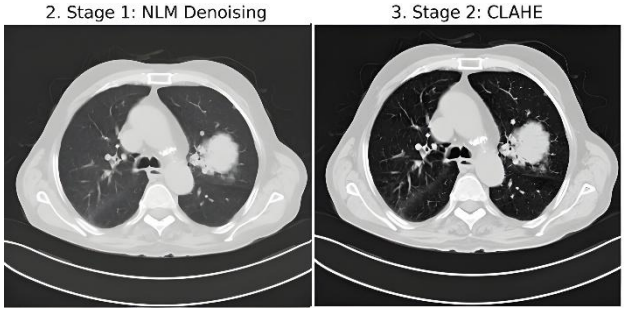


Figure 2. NLM denoising filter and CLAHE filter

Optimized unsharp masking (Edge sharpening): Sharp structural boundaries serve as the primary catalyst for successful deep learning segmentation. In this stage, the high-frequency components are extracted and carefully added back to the enhanced image. To generate the sharp-edged image I_{sharp}, a blurred version of the image resulting from the CLAHE stage, I_{clahe} is first generated using a Gaussian operator with standard deviation δ [24]:

$$I_{blur}(x,y) = \mathrm{I}_{clahe}(x,y) \times G(x,y,\sigma) \quad (5)$$

where

$$G(x,y,\sigma) = \frac{1}{2\pi\sigma^2}\exp(-\frac{x^2+y^2}{2\sigma^2}) \quad (6)$$

Next, the "unsharp mask" (Figure 3) containing the fine edge details exclusively is extracted via subtraction:

$$G_{mask}(x,y) = I_{clahe}(x,y) - I_{blur}(x,y) \quad (7)$$

Finally, the mask is linearly blended with the processed image using a scaling factor k where $k > 0$:

$$I_{sharp}(x,y) = I_{clahe}(x,y) + k.G_{mask}(x,y) \quad (8)$$

In this study, $k=0.5$ and $\delta=1.0$ were empirically selected to ensure continuous, crisp external tumor boundaries without generating artificial halos or artifacts.

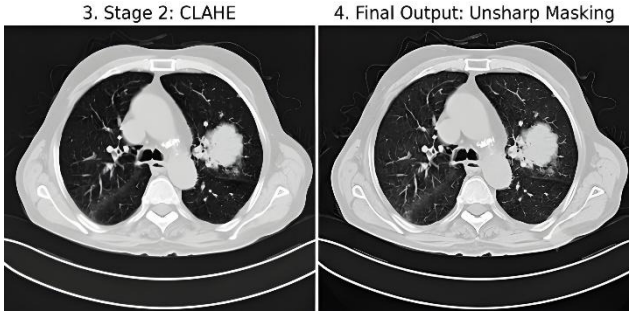


Figure 3. CLAHE filter and unsharp masking

Table 3 shows the preprocessing hyperparameters used in the proposed framework. The preprocessing parameters were fixed throughout all experiments to ensure reproducibility. For NLM denoising, the filtering strength was set to 10, with a (7×7) template window and a (21×21) search window, providing a compromise between noise suppression and preservation of anatomical structures. CLAHE was applied using a clip limit of 2.0 and an (8×8) tile grid to enhance local contrast while limiting excessive amplification of noise. For unsharp masking, a Gaussian standard deviation of (*sigma*=1.0) and a scaling factor of (k=0.5) were used to enhance structural boundaries without producing excessive sharpening artifacts. These parameter values were kept constant for all images during the experiments.

3.3 Radiomics feature extraction

Following image preprocessing, quantitative image features were extracted from automatically segmented lung ROIs. Lung segmentation was performed using a pretrained LungMask model [25], followed by post-processing to suppress peripheral false-positive regions and retain the principal lung components. The resulting binary lung masks

were applied to the corresponding CT images to define the regions used for quantitative feature extraction.

Table 3. Preprocessing hyperparameters used in the proposed framework

Preprocessing Stage	Parameter	Value
NLM denoising	Filtering strength hh	10
NLM denoising	Template/Patch window size	7×7
NLM denoising	Search window size	21×21
CLAHE	Clip limit	2.0
CLAHE	Tile grid size	8×8
Unsharp masking	Gaussian σ\sigma	1.0
Unsharp masking	Scaling factor kk	0.5

Since the available IQ-OTH/NCCD archive contains images in JPG format rather than the original DICOM images with calibrated Hounsfield Unit (HU) values, the extracted descriptors represent quantitative intensity and texture characteristics of the displayed CT images within the automatically segmented lung ROIs. Therefore, the extracted features were not interpreted as expert-annotated tumor-specific radiomic measurements. Four quantitative features were extracted from each segmented lung ROI. These consisted of two first-order statistical features, namely Skewness and Kurtosis, and two second-order texture features derived from the Gray-Level Co-occurrence Matrix (GLCM), namely Contrast and Homogeneity. Prior to GLCM computation, image intensities were quantized into 32 gray levels. Spatial co-occurrences were calculated at a pixel distance of one using four directions (0°, 45°, 90°, and 135°) and combined to provide a directionally robust representation of lung texture [26].

First-Order statistical features: Skewness measures the asymmetry of the pixel-intensity distribution within the segmented lung ROI and is calculated as:

$$Skewness = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{X_i-\mu}{\sigma}\right)^3 \quad (9)$$

Kurtosis characterizes the peakedness and tail behavior of the intensity distribution. Pearson kurtosis was employed and is expressed as:

$$Kurtosis = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{X_i-\mu}{\sigma}\right)^4 \quad (10)$$

where x_i is the intensity value of the i-th pixel, μ Is the mean intensity of all pixels, σ Is the standard deviation of the pixel intensities, and N Is total number of pixels in the ROI.

Second-Order texture features: Texture characteristics were quantified using the GLCM, which describes the spatial relationships between neighboring gray-level values within the segmented lung ROI. Contrast measures local gray-level variations and is defined as:

$$Contrast = \sum_{i=0}^{G-1}\sum_{j=0}^{G-1}(i-j)^2 P(i,j) \quad (11)$$

Homogeneity measures the concentration of GLCM values near the main diagonal and is calculated as:

$$Homogeneity = \sum_{i=0}^{G-1}\sum_{j=0}^{G-1}\frac{P(i,j)}{1+(i-j)^2} \quad (12)$$

where $P(i,j)$ denotes the normalized probability of the co-occurrence of gray levels i and j, and G represents the number of quantized gray levels. Figure 4 shows the radiomics steps.

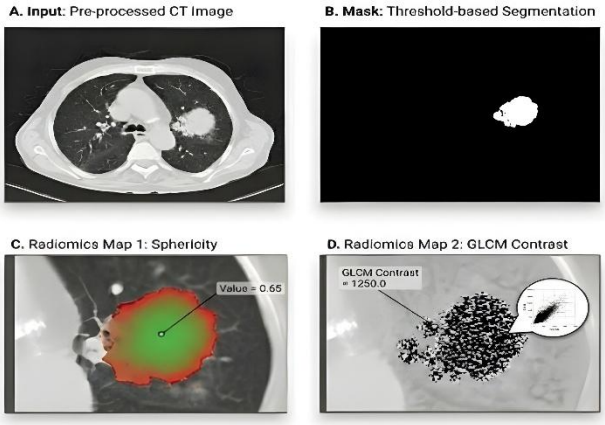


Figure 4. Radiomics steps

3.4 Fusion of deep and radiomic features

Following feature extraction, we implemented a feature-level fusion strategy to combine the high-level representations extracted by EfficientNet-B7 with the quantitative features derived from the automatically segmented lung ROIs. The EfficientNet-B7 network was initialized with ImageNet-pretrained weights, with the original classification head removed. Global average pooling was applied to the final convolutional feature maps to generate the deep feature vector [10]. In parallel, four quantitative lung-ROI features- Skewness, Kurtosis, GLCM Contrast, and GLCM Homogeneity- were extracted for each image. To prevent information leakage, these features were standardized using the mean and standard deviation calculated exclusively from the training set. The same fitted transformation was subsequently applied to the validation and held-out test sets. The standardized quantitative feature vector was concatenated with the EfficientNet-B7 deep feature vector to generate a unified feature representation:

$$F_{fusion} = [F_{deep} \parallel F_{ROI}] \quad (13)$$

where F_{deep} denotes the deep feature vector obtained from EfficientNet-B7, F_{ROI} represents the standardized vector of the four quantitative lung-ROI features, ‖parallel denotes feature concatenation, and F_{fusion} represents the resulting fused feature vector.

The fused representation was then processed by two fully connected layers with 256 and 64 neurons, with ReLU activation. Dropout rates of 0.40 and 0.25 were applied, respectively, to reduce overfitting. Finally, a three-neuron Softmax output layer generated the probabilities corresponding to the Normal, Benign, and Malignant classes. The fusion classifier was optimized using the Adam optimizer with an initial learning rate of 0.001 and class-weighted training to account for class imbalance. Figure 5 illustrates the proposed feature-level fusion architecture integrating EfficientNet-B7 deep features with radiomic descriptors for lung CT classification.

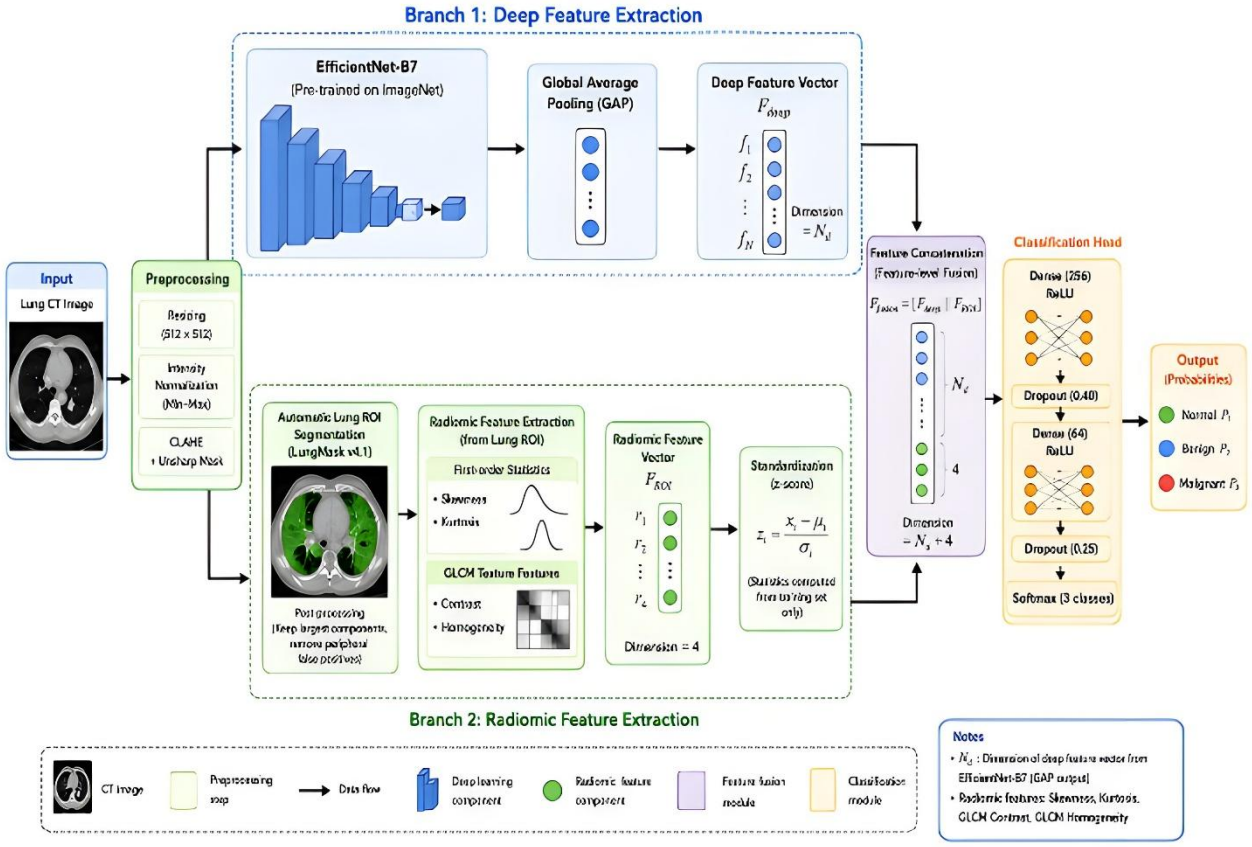


Figure 5. Proposed feature-level fusion architecture integrating EfficientNet-B7 deep features with radiomic descriptors for lung CT classification

4. Model testing and performance evaluation

The dataset was partitioned using a stratified image-level strategy, with 90% of the images assigned to the development set and 10% reserved as an independent held-out test set. The development set was further divided into training and validation subsets, while the held-out test set was not used during model training or hyperparameter adjustment. The EfficientNet-B7 model was initialized with ImageNet-pretrained weights and trained with a batch size of 32 for up to 100 epochs. The Adam optimizer was employed with an initial learning rate of 0.001. Categorical cross-entropy was used as the loss function for the three-class classification problem comprising Normal, Benign, and Malignant categories. During training, the training samples were shuffled and processed in mini-batches. To improve model generalization and reduce overfitting, data augmentation was applied exclusively to the training set. The augmentation strategy included random rotations within ±10°, random horizontal flipping with a probability of 0.5, and random zooming within ±10%. These transformations were selected to introduce moderate spatial variability while preserving the relevant anatomical characteristics of the lung CT images. No data augmentation was applied to the validation or held-out test sets. To account for class imbalance, class weights were computed from the class distribution of the training set and incorporated during model training, assigning greater importance to samples from underrepresented classes.

After each epoch, the model was evaluated on the validation set using the validation loss and classification accuracy. We employed an adaptive learning-rate strategy during training. If the validation loss did not improve for five consecutive epochs, the learning rate was reduced by a factor of 0.1. Early stopping and model checkpointing were also employed to reduce overfitting. The model configuration achieving the lowest validation loss was retained as the best-performing model. Early stopping was configured with a patience of 10 epochs; therefore, training was terminated when the validation loss failed to improve for 10 consecutive epochs. The learning-rate reduction patience of five epochs and the early-stopping patience of 10 epochs were applied as separate training-control mechanisms. After training, the best-performing model was evaluated once on the independent held-out test set.

The final class prediction was determined from the highest probability produced by the Softmax output layer. Classification performance was evaluated using accuracy, precision, recall, F1-score, and the confusion matrix [12]. For the three-class classification problem, class-specific precision, recall, and F1-score were calculated using a one-vs-rest approach, together with macro- and weighted-average performance measures. Table 4 below provides the model details.

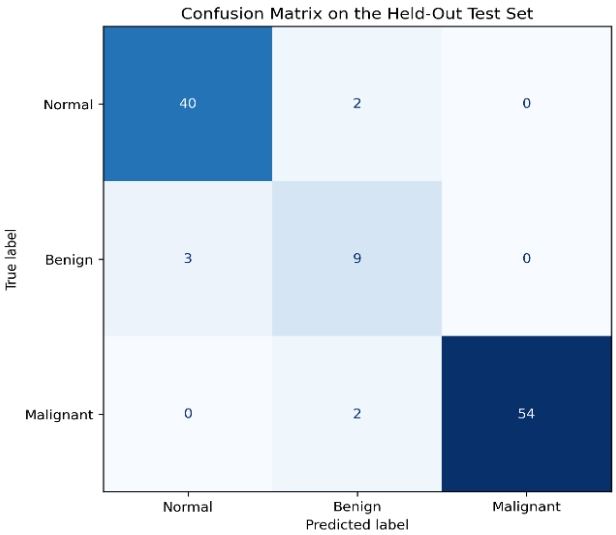
Table 4. Hyperparameters of the proposed approach

Hyperparameter / Parameter	Value	Brief Description
Input Resolution	224×224	The standard input size for the CT scan images.
Total Parameters	66.7 Million	Approximate total number of parameters in the standard EfficientNet-B7 architecture pretrained on ImageNet. No reduced or custom variant was used.
Learning Rate	0.001	Initial step size for weight optimization (utilizing the Adam Optimizer).
Max Epochs	100	Maximum number of complete passes through the entire training dataset.
Early Stopping	patience = 10	Terminates training if validation loss does not improve for 10 consecutive epochs.
Strides	1 to 2	The step size of the kernel filters across convolutional layers (varies by block).
Batch Size	32	Number of image samples processed together before updating the model weights.

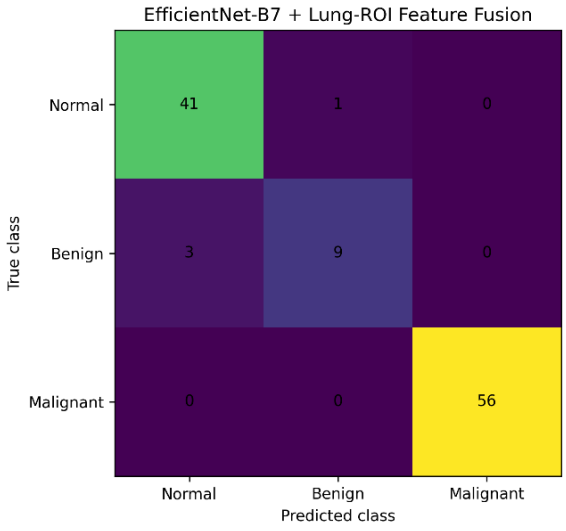
5. Results and discussion

To evaluate the contribution of feature-level fusion, two experimental configurations were investigated on the held-out lung CT test set. In the first experiment, EfficientNet-B7 was evaluated as the deep-learning classification model without incorporating the lung-ROI quantitative features. In the second experiment, the deep feature representation extracted by EfficientNet-B7 was combined with the four quantitative lung-ROI features (Skewness, Kurtosis, GLCM Contrast, and GLCM Homogeneity) using the proposed feature-level fusion strategy. Confusion matrix analysis was performed for both experiments to provide a detailed comparison of classification behavior across the Normal, Benign, and Malignant categories. In the first experiment, the EfficientNet-B7 model correctly classified 40 of 42 Normal images, 9 of 12 Benign images, and 54 of 56 Malignant images, corresponding to an overall accuracy of 93.64%. In the fusion experiment, the classification performance improved, with 41 of 42 Normal images, 9 of 12 Benign images, and all 56 Malignant images correctly classified, resulting in an overall accuracy of 96.36%.

As illustrated in Figure 6, the confusion matrices provide a visual comparison of the classification performance of the two experimental configurations. The fusion framework reduced the total number of misclassifications from the EfficientNet-B7-only model by correctly classifying them after feature fusion. The fusion approach also reduced Normal-to-Benign misclassification from two images to one. However, three Benign images remained misclassified as Normal in both configurations, indicating that discrimination of the Benign class remains the most challenging aspect of the three-class classification task.



(a)



(b)

Figure 6. Confusion matrices obtained on the held-out lung CT test set: (a) EfficientNet-B7 without lung-ROI feature fusion (accuracy = 93.64%), and (b) the proposed EfficientNet-B7 with lung-ROI feature-level fusion (accuracy = 96.36%)

Based on the confusion matrices, several evaluation metrics were calculated to quantitatively compare the performance of the two experimental configurations [27]. Accuracy measures the overall classification performance:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (14)$$

Precision evaluates the proportion of correctly predicted positive samples among all positive predictions:

$$Precision = \frac{TP}{TP+FP} \quad (15)$$

Recall (Sensitivity) measures the ability of the model to correctly identify positive samples:

$$Recall = \frac{TP}{TP+FN} \quad (16)$$

F1-Score represents the harmonic mean of Precision and Recall:

$$F1 - Score = \frac{2}{\frac{1}{Precision}+\frac{1}{Recall}} \quad (17)$$

The performance of both experimental configurations was evaluated using accuracy, precision, recall, and F1-score. For the three-class classification problem, class-specific metrics were calculated using a one-vs-rest approach for the Normal, Benign, and Malignant classes. Macro-averaged and weighted-averaged measures were additionally used to summarize the overall classification performance across the three categories. These metrics were used to compare the EfficientNet-B7-only configuration with the proposed EfficientNet-B7 and lung-ROI feature fusion framework. Table 5 summarizes the classification performance obtained by the two experimental configurations on the held-out test set.

Table 5. Classification performance of the proposed model on the held-out lung CT test set

Metric	EfficientNet-B7 Only	EfficientNet-B7 + Lung-ROI Feature Fusion
Accuracy	93.64 %	96.36 %
Macro Precision	87.42 %	94.39 %
Macro Recall	88.89 %	90.87 %
Macro F1-Score	88.10 %	92.39 %
Macro Specificity	97.17 %	97.53 %
Macro AUC-ROC	97.40 %	98.28 %
Correctly Classified Images	103/110	106/110
Misclassified Images	7/110	4/110

The experimental results as illustrated in Figure 7,demonstrate that the proposed model achieved an overall classification accuracy of 93.64% on the held-out test set. Of the 110 CT images, 103 were correctly classified. As shown in Figure 7, the experimental results show that the proposed EfficientNet-B7 and lung-ROI feature fusion framework achieved an overall classification accuracy of 96.36% on the held-out test set, correctly classifying 106 of 110 CT images. The results presented in Figure 7 particulary the confusion matrix showed that 41 of 42 Normal images, 9 of 12 Benign images, and all 56 Malignant images were correctly classified. Futhermore Figure 7 shows that the proposed fusion framework achieved macro-averaged precision, recall, F1-score, and specificity values of 94.39%, 90.87%, 92.39%, and 97.53%, respectively, together with a macro-average AUC-ROC of 98.28%. The model performed particularly well on the Malignant class, correctly classifying all test images in this category, as demonstarated in Figure 7. The remaining four misclassifications occurred between the Normal and Benign categories, indicating that differentiation of Benign cases remained the most challenging aspect of the three-class classification task.

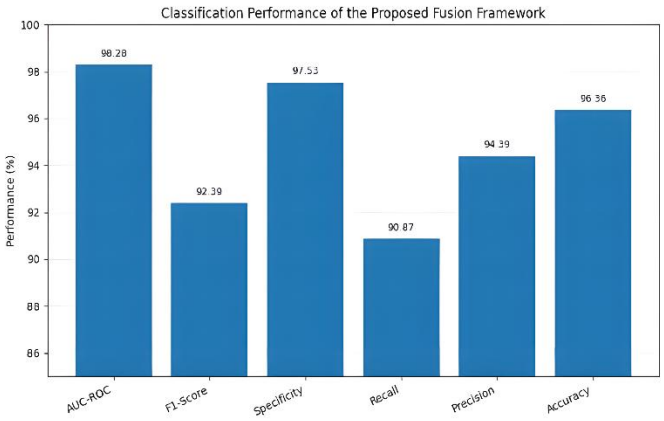


Figure 7. Overall classification performance of the proposed EfficientNet-B7 and lung-ROI feature fusion framework on the held-out test set

The graphical results presented in Figure 8 illustrate the performance of the proposed EfficientNet-B7 and lung-ROI feature fusion framework. The fusion framework achieved an overall classification accuracy of 96.36% on the held-out test set. The macro-averaged precision, recall, F1-score, and specificity were 94.39%, 90.87%, 92.39%, and 97.53%, respectively. Furthermore, the framework achieved a macro-average AUC-ROC of 98.28%, indicating strong discriminative performance across the Normal, Benign, and Malignant classes. The training and validation curves provide additional insight into the proposed framework's learning behavior. The loss curves show a progressive decrease during training, while the accuracy, precision, and recall curves increase and gradually converge. The similar behavior of the training and validation curves indicates stable learning without pronounced divergence between the two sets. Together with the held-out test results, these findings demonstrate the effectiveness of the proposed fusion framework for three-class lung CT image classification: Normal, Benign, and Malignant categories. The quantitative evaluation of the proposed EfficientNet-B7 and lung-ROI feature fusion framework demonstrated strong performance in the three-class classification of Normal, Benign, and Malignant lung CT images. To account for the dataset's multi-class nature and class imbalance, we used macro-averaged performance measures to give equal weight to each diagnostic category. On the held-out test set, the proposed fusion framework achieved a macro-precision of 94.39%, macro-recall of 90.87%, and macro-F1-score of 92.39%, together with a macro-specificity of 97.53% and macro-average AUC-ROC of 98.28%. The overall classification accuracy reached 96.36%, with 106 of the 110 test images correctly classified. The class-wise analysis further showed that all 56 Malignant images in the held-out test set were correctly classified, while 41 of 42 Normal images and 9 of 12 Benign images were correctly identified. The remaining classification errors occurred between the Normal and Benign categories, indicating that differentiation of Benign cases remains the most challenging aspect of the three-class classification task.

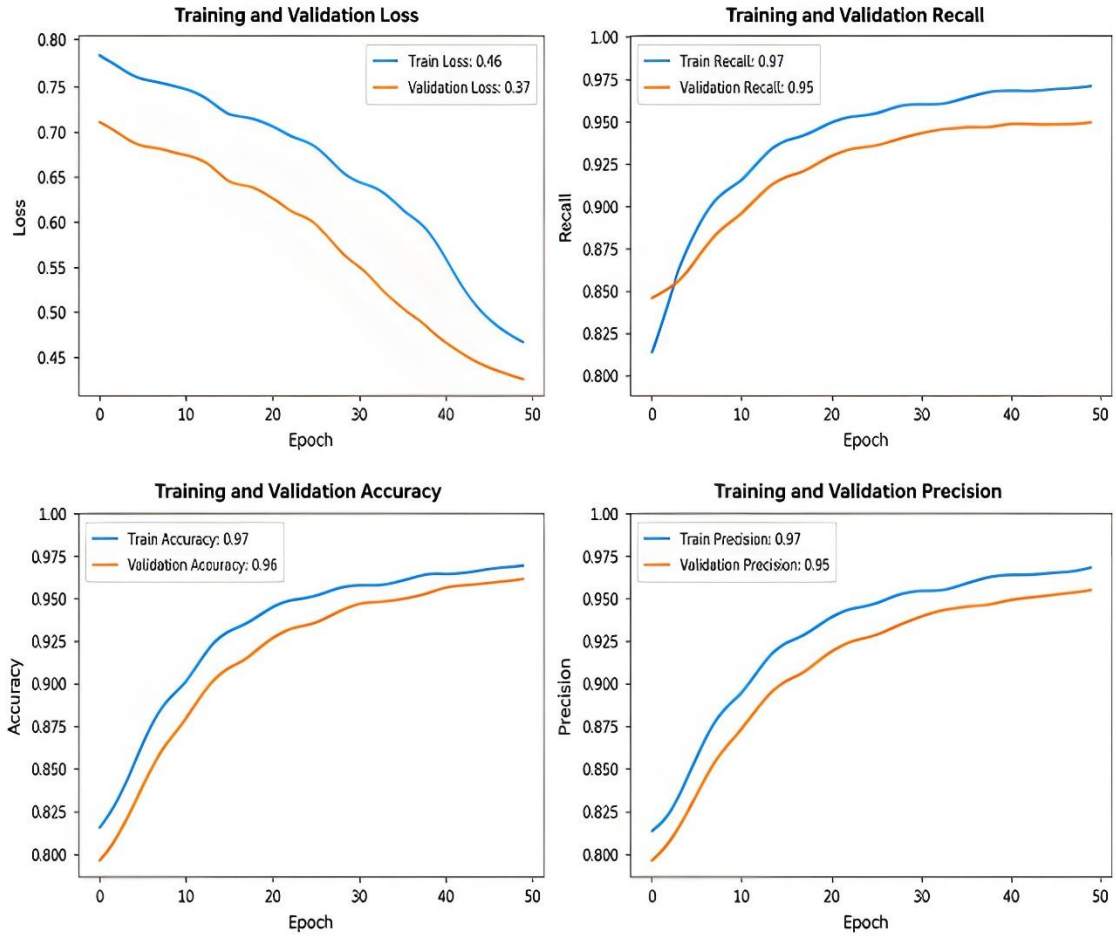


Figure 8. Training and validation performance curves of the proposed EfficientNet-B7 and lung-ROI feature fusion framework

Overall, the obtained results demonstrate the strong discriminative capability of the proposed fusion framework on the evaluated dataset. Nevertheless, further evaluation using independent external datasets and patient-level validation is required before conclusions regarding clinical generalizability can be established.

6. Conclusion

This study presented a lung CT classification framework based on EfficientNet-B7 and quantitative radiomic features for distinguishing among Normal, Benign, and Malignant categories. The proposed framework incorporated image preprocessing, deep feature extraction, and quantitative radiomic information to provide a comprehensive representation of lung CT images. Accuracy on the withheld test set was 96.36%, with correct classification of 106 out of 110 CT images. This framework of feature fusion with EfficientNet-B7 resulted in a macro-averaged precision, recall, and F1-score of 94.39%, 90.87%, and 92.39%, respectively, with macro-specificity at 97.53%. The one-vs-rest ROC analysis reported the AUC at 0.9849, 0.9634, and 1.0000 for the classes Normal, Benign, and Malignant, respectively, giving a macro-average AUC of 0.9828. From the confusion matrix, 41 of 42 Normal images, 9 of 12 Benign images, and all 56 Malignant images were correctly classified. The remaining classification errors occurred among Normal and Benign, indicating that classification of Benign cases was the most challenging of the three-class classification. The findings from this study are very positive; however, challenges should be noted. The experiments used a single publicly available, relatively modest-sized dataset, which may limit the generalizability of the findings to broader patient populations and different imaging environments. Furthermore, patient identifiers were not explicitly available in the accessible dataset archive; therefore, the experimental partition was performed at the image level rather than the patient level, and complete separation of slices originating from the same patient across data partitions could not be guaranteed. External validation using independent multi-center datasets and patient-level partitioning is therefore required to provide a more rigorous assessment of model generalizability. Future work will focus on validating the framework across larger and more diverse datasets and further investigating its robustness across different CT acquisition settings. Accordingly, the current framework should be considered a research-oriented classification approach rather than a system ready for direct clinical deployment.

Ethical issue

The authors are aware of and comply with best practices in publication ethics, specifically regarding authorship (avoidance of guest authorship), dual submission, manipulation of figures, competing interests, and compliance with research ethics policies. The authors adhere to publication requirements that the submitted work is original and has not been published elsewhere. All survey participants provided informed consent before participating, and the anonymity and confidentiality of all respondents were strictly maintained throughout the research process.

Data availability statement

The IQ-OTH/NCCD lung cancer dataset used in this study is publicly available through Kaggle and Mendeley Data. The experiments reported in this study were conducted using the accessible Kaggle archive. The corresponding dataset is available on Mendeley Data as Version 4 (DOI: 10.17632/bhmdr45bh2.4).

Conflict of interest

The authors declare no potential conflict of interest.

References

[1] Shong, L.Y.W. and Lam DC. "Emerging trends in global lung cancer burden." Semin Respir Crit Care Med Thieme Med Publ 2025. https://www.thieme-connect.com/products/ejournals/html/10.1055/a-2651-0612

[2] Al-Suhail NSJ and GA. "Lung cancer detection: The role of transfer learning in medical imaging." Int. Conf. Futur. Telecommun. Artif. Intell. (IC-FTAI), IEEE, pp. 1–6, 2024. https://doi.org/https://doi.org/10.1109/IC-FTAI62324.2024.10950018.

[3] Mienye, I.D., Swart, T.G., Obaido, G., Jordan, M. and Ilono P. "Deep convolutional neural networks in medical image analysis: A review." Information, vol. 16, 2025. https://doi.org/https://doi.org/10.3390/info16030195.

[4] Y. G. Chandavarkar, D. Rao, S. B. Pathan, K. Devaraja and SDP. "A multi-faceted approach to lung cancer classification using diverse machine learning algorithms using the IQ-OTH/NCCD dataset." Emerg. Technol. AI, Comput. Commun. Cybersecurity, CRC Press., pp. 53-58.

[5] Linguraru MG, Bakas S, Aboian M, Chang PD, Flanders AE, Kalpathy-Cramer J, Kitamura FC, Lungren MP, Mongan J, and Prevedello LM SR. "Clinical, cultural, computational, and regulatory considerations to deploy ai in radiology: perspectives of rsna and miccai experts." Radiology: Artificial Intelligence, vol. 4, no. 6, p. e240225, 2024. https://doi.org/10.1148/ryai.240225.

[6] W. Wulaningsih, C. Villamaria, A. Akram, J. Benemile, F. Croce and JW. "Deep learning models for predicting malignancy risk in CT-Detected pulmonary nodules: A systematic review and meta-analysis." Lung, vol. 202, pp. 625–636, 2024. https://doi.org/10.1007/s00408-024-00706-1.

[7] Huang S, Yang J, Shen N, Xu Q, and Zhao Q. "Artificial intelligence in lung cancer diagnosis and prognosis: Current application and future perspective." Seminar. Cancer Biol., vol. 89, pp. 30–37, 2023. https://doi.org/10.1016/j.semcancer.2023.01.006?utm_source=chatgpt.com

[8] M. Z. Khan, M. Jabeen, M. U. G. Khan, S. S. Alqahtani and SA. "Deep learning in medical imaging: challenges and opportunities. " Artificial Intelligence, vol. 57, p 10625, 2024. https://doi.org/10.1007/s10462-023-10625-5.

[9] Hosseini SH, Monsefi R, and Shadroo S. "Deep learning applications for lung cancer diagnosis: a systematic review." Multimed Tools Application, vol. 83, pp:14305–14335, 2024. https://doi.org/10.1007/s11042-023-16046-w?utm_source=chatgpt.com

[10] M. Tan and Q. V. Le. "EfficientNet: rethinking model scaling for convolutional neural networks." Proc. 36th Internation Conference. Mach. Learn., vol. 97, pp. 6105–6114, 2019.

[11] Raza, R., Zulfiqar, F., Khan, M.O., Arif, M., Alvi, A., Iftikhar, M.A. and Alam, T., "Lung-EffNet: Lung cancer classification using EfficientNet from CT-scan images." Engineering Applications of Artificial Intelligence, vol. 126, p. 106902, 2023. https://doi.org/10.1016/j.engappai.2023.106902?utm_source=chatgpt.com.

[12] Suresha R, Devika KM, and Prabhu A. "Support Vector Machine classifier based lung cancer recognition: a fusion approach." International Conference On Edge Computing and Applications (ICECAA) 2022 - Proc., 2022, pp. 1565–1572. https://doi.org/10.1109/ICECAA55415.2022.9936385.

[13] Qadir AM, Abdalla PA, and Abd DF. "A hybrid lung cancer model for diagnosis and stage classification from computed tomography images." Iraqi Journal for Electrical and Electronic Engineering, vol. 20, pp. 266-274, 2024. https://doi.org/10.37917/ijeee.20.2.23.

[14] Gulsoy T, and Kablan EB. "FocalNeXt: a ConvNeXt augmented FocalNet architecture for lung cancer classification from CT-Scan images." Expert System Application, vol. 1, no. 261, p. 125553, 2024. https://doi.org/10.2139/ssrn.4741248.

[15] Majumder S, Gautam N, Basu A, Sau A, Geem ZW, and Sarkar R. "MENet: a mitscherlich function based ensemble of CNN models to classify lung cancer using CT scans." PLoS One, vol. 19, no. 3, p. :e0298527, 2024. https://doi.org/10.1371/journal.pone.0298527.

[16] Al-Tamimi MSH, Jabbar SF, Taha AM, and Al-Jibory FK. "Enhancing lung cancer classification using CT images using processing techniques employing U-Net architecture." Journal of Internet Services and Information Security, vol. 15, no. 2, pp.:590–609, 2025.https://doi.org/10.58346/JISIS.2025.I2.041.

[17] AL-Huseiny HA and M. "The IQ-OTH/NCCD lung cancer dataset. mendeley data" no. 4, 2023.

https://doi.org/https://doi.org/10.17632/bhmdr45bh2.4.

[18] Nazir, Nahida, Abid Sarwar and BSS. "Recent developments in denoising medical images using deep learning: An overview of models, techniques, and challenges." Micron, , vol.1, no. 180, p. 103615, 2024. https://doi.org/10.1016/j.micron.2024.103615.

[19] Buades, A., Coll, B. and Morel, J.M. "A non-local algorithm for image denoising." In 2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05), vol. 2, pp. 60-65. https://doi.org/10.1109/CVPR.2005.38.

[20] D'hondt L, Franck C, Kellens PJ, Zanca F, Buytaert D, Van Hoyweghen A, Addouli HE, Carpentier K, Niekel M, and Spinhoven M BK. "Impact of deep learning image reconstruction on volumetric accuracy and image quality of pulmonary nodules with different morphologies in low-dose CT." Cancer Imaging, vol. 24. no. 1, 2024. https://doi.org/https://doi.org/10.1186/s12880-024-00703-w?utm_source=chatgpt.com.

[21] Golkarieh A, Kiashemshaki K, Boroujeni SR, and Isakan NA. "Advanced U-Net architectures with CNN backbones for automated lung cancer detection and segmentation in chest CT images." ArXiv Prepr ArXiv 09898, 2025. https://doi.org/10.21203/rs.3.rs-5232211/v1.

[22] E. D. Pisano, S. Zong, B. M. Hemminger, M. DeLuca, R. E. Johnston, K. Muller, M. P. Braeuning and SMP. "Contrast limited adaptive histogram equalization image processing to improve the detection of simulated spiculations in dense mammograms." Journal of Digit Imaging, vol. 11, no. 4, pp. 193–200, 1998. https://doi.org/10.1007/BF03178082.

[23] Papadopoulos, Athanasios, Dimitrios I. Fotiadis, and Lena Costaridou. "Improvement of microcalcification cluster detection in mammography utilizing image enhancement techniques." Computers in biology and medicine, vol. 38, no. 10, pp. 1045-1055, 2008. https://doi.org/10.1016/j.compbiomed.2008.07.006.

[24] R. C. Gonzalez and R. E. Woods. "Digital Image Processing." 4th ed. New York, NY, USA: Pearson: 2018. https://www.cl72.org/090imagePLib/books/Gonzales,Woods-Digital.Image.Processing.4th.Edition.pdf

[25] Hofmanninger, J., Prayer, F., Pan, J., Röhrich, S., Prosch, H. and Langs G. "Automatic lung segmentation in routine imaging is primarily a data diversity problem, not a methodology problem." European radiology experimental, vol. 4, no. 1, p. 50, 2020. https://doi.org/https://doi.org/10.1186/s41747-020-00173-2?utm_source=chatgpt.com.

[26] Zwanenburg, A., Vallières, M., Abdalah, M.A., Aerts, H.J., Andrearczyk, V., Apte, A., Ashrafinia, S., Bakas, S., Beukinga, R.J., Boellaard, R. and Bogowicz M. "The image biomarker standardization initiative: standardized quantitative radiomics for high-throughput image-based phenotyping." Radiology, vol. 295, no. 2, pp. 328-338, 2020. https://doi.org/https://doi.org/10.1148/radiol.2020191145?utm_source=chatgpt.com.

[27] Sokolova, M. and Lapalme G. A "Systematic analysis of performance measures for classification tasks." Information Processing & Management, vol 45, no. 1, pp. 427-437, 2009. https://doi.org/https://doi.org/10.1016/j.ipm.2009.03.002?utm_source=chatgpt.com.