

PulmoDiagAI: An Intelligent Medical Diagnosis System for Automated Pulmonary Pathology Classification using Deep Learning

Abdellatif BOUZID-DAHO

`abdellatif.bouزيد-daho@ummto.dz`

Mouloud Mammeri University of Tizi-Ouzou <https://orcid.org/0009-0006-6531-8935>

Merzouk YOUNSI

Mouloud Mammeri University of Tizi-Ouzou <https://orcid.org/0000-0002-5542-2900>

Djedjiga DERBANE

Mouloud Mammeri University of Tizi-Ouzou

Kamilia FODIL

Mouloud Mammeri University of Tizi-Ouzou

Research Article

Keywords: Lung cancer, COVID'19, Deep learning, Convolutional neural networks, Transfer learning, Medical imaging, Computer-aided diagnosis

Posted Date: October 28th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11510345/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: The authors declare no competing interests.

PulmoDiagAI: An Intelligent Medical Diagnosis System for Automated Pulmonary Pathology Classification using Deep Learning

¹Abdellatif BOUZID-DAHO (corresponding author)

Laboratoire Vision Artificielle et Automatique des Systèmes (LVAAS)
Departement of Biomedical Engineering, Faculty of Electrical and Computer Engineering,
Mouloud Mammeri University, Tizi-Ouzou, Algeria
<https://orcid.org/0009-0006-6531-8935>
abdellatif.bouزيد-daho@ummto.dz

²Merzouk YOUNSI

Laboratoire Vision Artificielle et Automatique des Systèmes (LVAAS)
Mouloud Mammeri University (UMMTO)
Tizi-Ouzou, Algeria
<https://orcid.org/0000-0002-5542-2900>

³Djedjiga DERBANE

Departement of Biomedical Engineering, Faculty of Electrical Engineering and Computer Science,
Mouloud Mammeri University, Tizi-Ouzou, Algeria

⁴Kamilia FODIL

Departement of Biomedical Engineering, Faculty of Electrical Engineering and Computer Science,
Mouloud Mammeri University, Tizi-Ouzou, Algeria

Abstract: Lung cancer and Coronavirus Disease 2019 (COVID'19) remain among the leading causes of respiratory-related morbidity and mortality worldwide, highlighting the critical need for rapid, accurate, and automated diagnostic solutions. This study presents PulmoDiagAI, an intelligent computer-aided diagnosis (CAD) system that integrates state of the art deep learning techniques for the automated detection and classification of pulmonary diseases from chest X-ray and computed tomography (CT) images. Five pre-trained convolutional neural network (CNN) architectures, namely VGG19, ResNet50, MobileNetV3, EfficientNet-B3, and DenseNet121 were fine-tuned using a transfer learning strategy to identify the most effective model for each diagnostic task. A comprehensive preprocessing pipeline, including image standardization, median filtering, resizing, intensity normalization, and Contrast Limited Adaptive Histogram Equalization (CLAHE), was employed to enhance image quality and improve feature extraction. The proposed framework was evaluated on two publicly available benchmark datasets: IQ-OTH/NCCD for lung cancer detection and COVID-QU-Ex for COVID'19 diagnosis. Experimental results demonstrate that EfficientNet-B3 achieved superior performance for COVID'19 classifications, reaching an accuracy of 97.0% and an AUC-ROC of 99.43%. For lung cancer detection, MobileNetV3 and ResNet50 delivered outstanding results, achieving accuracies of 99.0% and 98.0%, respectively, with a perfect AUC-ROC of 100%. To facilitate clinical adoption, the proposed model was deployed as an intuitive graphical user interface (GUI) that provides real time diagnostic predictions along with confidence scores. By combining high diagnostic accuracy, computational efficiency, and practical usability, PulmoDiagAI offers a reliable decision-support system capable of assisting radiologists and clinicians in the early detection of pulmonary diseases, ultimately contributing to improved diagnostic workflows and patient outcomes.

Keywords: Lung cancer, COVID'19, Deep learning, Convolutional neural networks, Transfer learning, Medical imaging, Computer-aided diagnosis.

1. Introduction

Pulmonary diseases constitute a major global health burden, accounting for a significant proportion of morbidity and mortality worldwide. Among these, lung cancer remains the leading cause of cancer-related deaths globally, with approximately 1.8 million deaths reported in 2020 [1]. The prognosis

of lung cancer is critically dependent on early detection, making the development of effective screening methodologies an urgent priority. Similarly, the COVID-19 pandemic, caused by the SARS-CoV-2 virus, has demonstrated the devastating impact of respiratory infections, necessitating rapid and accurate diagnostic tools to manage outbreaks effectively [2].

Medical imaging plays a fundamental role in the diagnosis of pulmonary pathologies. Chest radiography (X-ray) and computed tomography (CT) are the primary imaging modalities employed for detecting lung abnormalities, including nodules, masses, ground-glass opacities, and consolidations [3]. However, the interpretation of these images remains largely dependent on the expertise of radiologists and pulmonologists, a process that is inherently time-consuming, subject to inter-observer variability, and vulnerable to human error. In contexts of high caseloads, such as during the COVID-19 pandemic, this manual analysis can lead to diagnostic delays and increased clinician workload.

Recent advances in artificial intelligence (AI), particularly deep learning, have opened new perspectives for automated medical image analysis. Convolutional neural networks (CNNs) have demonstrated remarkable efficacy in image classification, detection, and segmentation tasks, owing to their capacity to automatically learn hierarchical feature representations from visual data [4] [5]. The application of deep learning to chest imaging has shown promising results for detecting various pulmonary conditions, including lung nodules, pneumonia, and COVID-19 [6] [7]. Transfer learning, which leverages knowledge from models pre-trained on large datasets like ImageNet, has proven particularly valuable for medical imaging applications where annotated data is scarce [8] [9].

Despite these advances, several challenges remain in developing clinically deployable AI systems for pulmonary diagnosis. These include the need for robust preprocessing to handle variations in image quality [10] [11], the selection of appropriate model architectures that balance accuracy and computational efficiency [12], and the development of interpretable interfaces for clinical use. Furthermore, there is ongoing debate regarding which CNN architectures are most effective for specific pulmonary pathologies [13] [14].

This study addresses these challenges by presenting PulmoDiagAI, a comprehensive deep learning-based system for automated classification of pulmonary pathologies. Specifically, we aim to: (1) develop and evaluate five state of the art CNN architectures for lung cancer and COVID-19 classifications; (2) implement a robust preprocessing pipeline optimized for medical imaging; (3) deploy the system as an intuitive graphical user interface for clinical use; and (4) provide a comparative analysis to identify the most effective architectures for each pathology.

2. Related works

The application of deep learning to pulmonary disease diagnosis has attracted considerable research attention, with numerous studies exploring various architectures and approaches. Recent reviews have comprehensively surveyed the state-of-the-art in AI-based lung disease detection, highlighting the predominant role of CNNs and transfer learning in achieving high diagnostic accuracy [1] [3].

For COVID-19 detection, multiple studies have demonstrated the potential of deep learning on chest X-ray and CT images. A comparative study on the COVID-QU-Ex dataset found that EfficientNetV2-small achieved superior performance with 96.46% accuracy, outperforming other architectures including MobileNetV3, which showed lower performance at 91.35% accuracy [7]. ResNet-based architectures have also demonstrated strong performance, with one study reporting 95.31% accuracy using ResNet152V2 for multi-disease classification including COVID-19, pneumonia, and lung cancer. More recently, vision transformer architectures have emerged as promising alternatives, with the WSDC-ViT model achieving 95.13% accuracy on chest X-ray datasets and 99.36% on CT datasets for pneumonia classification [9].

Lung cancer classification has also benefited from deep learning advances. Hybrid approaches combining CNNs with tree-based classifiers have shown promising results, with ResNet-18 feature extraction coupled with XGBoost achieving F1-scores between 0.977 and 0.982 for five-class lung disease classification [2]. A study comparing multiple architectures reported that VGG19+CNN achieved 98.05% accuracy for distinguishing COVID-19, pneumonia, and lung cancer, outperforming ResNet based models. Novel architectures such as LIC-UMNET, combining Dual Attention U-Net with MobileNet and whale optimization, have reported 98.86% classification accuracy for lung diseases including lung cancer and COVID-19 [6].

The importance of preprocessing for medical image classification has been consistently emphasized. Contrast Limited Adaptive Histogram Equalization (CLAHE) has been widely adopted for enhancing chest X-ray images, improving visibility of pathological features [4] [6]. Studies have shown that CLAHE is particularly effective for distinguishing between normal and pathological lung regions [10]. The choice of preprocessing techniques can significantly impact model performance, with some studies reporting that appropriate enhancement can improve accuracy by several percentage points [4] [11].

Transfer learning has become a standard approach in medical image classification, enabling models to leverage features learned from large natural image datasets [3] [5]. Studies have demonstrated that pre-trained architectures such as VGG19, ResNet50, and DenseNet121 can achieve high accuracy with limited medical training data [7]. However, the optimal architecture choice may vary depending on the specific pathology and dataset characteristics, with shallower models sometimes outperforming deeper ones for certain tasks [7].

Quantitative chest imaging analysis using AI has also progressed significantly, particularly for occupational lung diseases. Recent applications include CNN-based detection of pneumoconiosis with 83.8% accuracy [11], transformer-based factorized encoders achieving 97.06% accuracy on 3D CT data [11], and the novel Kolmogorov-Arnold Networks (KAN) achieving 98.03% accuracy for binary classification [11]. These advances highlight the rapid evolution of AI technologies for pulmonary diagnosis.

Despite these substantial contributions, there remains a need for comprehensive comparative studies evaluating multiple architectures on consistent datasets with standardized preprocessing. Additionally, clinically deployable systems require not only high accuracy but also user-friendly interfaces for practical implementation. The present study addresses these gaps by providing a systematic comparison of five CNN architectures across two key pulmonary pathologies, implemented with a complete preprocessing pipeline and deployed via an interactive graphical interface.

3. Materials and methods

3.1 Datasets

The following two public datasets were utilized in this study:

IQ-OTH/NCCD Lung Cancer Dataset: This dataset was collected from the Iraqi Oncology Teaching Hospital and National Center for Cancer Diseases. It comprises 1,190 thoracic CT images from 110 cases, categorized into three classes: 416 normal cases, 120 benign tumors, and 561 malignant tumors. The CT scans were acquired using a Siemens SOMATOM scanner and evaluated by oncologists and radiologists [16].

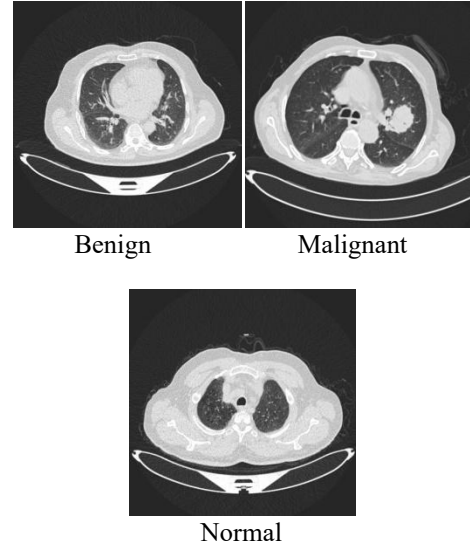


Figure 1: Sample CT images illustrating the three classes of the IQ-OTH/NCCD lung cancer dataset (Normal, Benign, and Malignant).

COVID-QU-Ex Dataset: This public dataset, developed by Qatar University, contains 33,920 chest X-ray images distributed across three classes: 11,956 confirmed COVID-19 cases, 11,263 non-COVID pneumonia cases, and 10,701 normal cases. The dataset includes ground-truth lung masks and careful label generation protocols based on RT-PCR test windows [17].

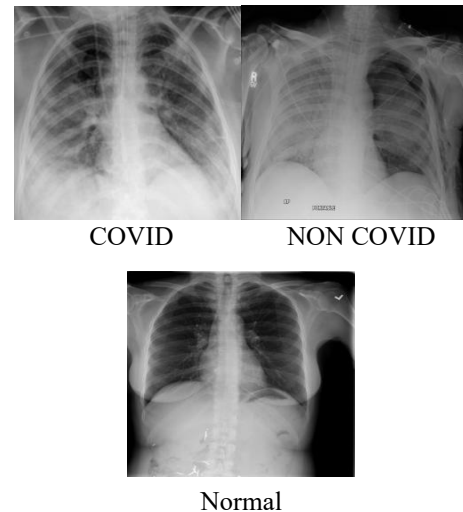


Figure 2: Representative chest X-ray images from the COVID-QU-Ex dataset used in this study.

3.2 Preprocessing pipeline

A comprehensive preprocessing pipeline was implemented to standardize image quality and enhance pathological features:

Standardization: All RGB images were converted to grayscale using the standard luminance formula: $\text{Gray} = 0.299 \times R + 0.587 \times G + 0.114 \times B$. This

conversion preserves clinically relevant intensity and texture information while reducing computational complexity.

Median filtering: A median filter was applied to reduce impulse noise while preserving edges. Comparative analysis of 3×3 and 5×5 kernel sizes indicated that the 3×3 kernel provided the optimal trade-off between noise reduction and detail preservation, with higher PSNR values for both datasets.

Resizing: All images were resized to 224×224 pixels to match the input requirements of the pre-trained CNN architectures. This standardization ensures compatibility across models while maintaining sufficient spatial resolution for pathological feature recognition.

Contrast enhancement: The Contrast Limited Adaptive Histogram Equalization (CLAHE) method was applied to improve local contrast and enhance visibility of pulmonary structures. CLAHE divides the image into small tiles and applies histogram equalization locally with a clip limit to prevent excessive noise amplification [4] [6]. This technique was particularly effective for highlighting ground-glass opacities and nodules.

Normalization: Pixel intensities were normalized to the [0,1] range by dividing by 255. This scaling facilitates stable and rapid model convergence during training.

3.3 Model architectures

Five state-of-the-art CNN architectures were implemented and compared:

VGG19: A deep architecture with 19 layers (16 convolutional, 3 fully connected). Known for its uniform 3×3 convolution filters and simplicity, though computationally intensive with 138 million parameters.

ResNet50: A 50-layer architecture incorporating residual blocks with skip connections. These connections mitigate the vanishing gradient problem, enabling effective training of deeper networks [2].

MobileNetV3: A lightweight architecture optimized for mobile and embedded devices. Employs depthwise separable convolutions and automated architecture search for efficient feature extraction.

EfficientNet-B3: Part of the EfficientNet family, using a compound scaling method to balance network depth, width, and resolution. Achieves state-of-the-art performance with optimized computational efficiency [7].

DenseNet121: Features densely connected layers where each layer receives inputs from all previous layers. This design encourages feature reuse and reduces parameter count while maintaining high performance [4].

3.4 Transfer learning approach

All models were initialized with ImageNet pre-trained weights through transfer learning. The base convolutional layers were frozen initially, while newly added fully connected layers were trained on the medical datasets. The top layers included Global Average Pooling, a Dense layer (128 neurons, ReLU activation), a Dropout layer (0.2), and a Softmax output layer with class-specific neurons. Models were compiled with the Adam optimizer (learning rate = 0.0001) using categorical cross-entropy loss. Training was conducted for 25 epochs with a batch size of 32.

3.5 Evaluation metrics

To comprehensively assess the performance of the proposed deep learning models, several widely used evaluation metrics were employed. These metrics provide complementary information regarding the classification capability of each model and are computed based on the values of the confusion matrix, namely True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN).

Accuracy: Accuracy measures the overall proportion of correctly classified samples among all tested instances. It provides a global indication of the classifier's performance.

$$\text{Accuracy (A)} = (\text{TP} + \text{TN}) / (\text{TP} + \text{TN} + \text{FP} + \text{FN}) \quad (1)$$

where:

TP: correctly classified positive samples.

TN: correctly classified negative samples.

FP: negative samples incorrectly classified as positive.

FN: positive samples incorrectly classified as negative.

Although accuracy is intuitive and widely used, it may provide misleading results when dealing with imbalanced datasets.

Precision: also referred to as the Positive Predictive Value (PPV), measures the proportion of correctly predicted positive samples among all samples classified as positive. High precision indicates a low false-positive rate.

$$\text{Precision (P)} = \text{TP} / (\text{TP} + \text{FP}) \quad (2)$$

A high precision value is particularly important in medical diagnosis, where reducing false alarms can prevent unnecessary clinical examinations and treatments.

Recall (Sensitivity): also known as Sensitivity or True Positive Rate (TPR), evaluates the model's ability to correctly identify positive cases.

$$\text{Recall (R)} = \text{TP} / (\text{TP} + \text{FN}) \quad (\%) \dots \dots \dots (3)$$

A high recall is crucial in healthcare applications because failing to detect a diseased patient (false negative) may lead to delayed diagnosis and severe clinical consequences.

F1-score: is the harmonic mean of Precision and Recall, providing a balanced measure when both false positives and false negatives are important.

$$\text{F1-score} = 2 * (\text{Precision} * \text{Recall}) / (\text{Precision} + \text{Recall}) \quad (\%) \dots \dots \dots (4)$$

The F1-score ranges from 0 to 1, where a value closer to 1 indicates superior classification performance.

Confusion Matrix: provides a detailed summary of the classification results by comparing predicted labels with the corresponding ground truth labels.

For binary classification, the confusion matrix is represented as:

		Predicted	
		Positive	Negative
Actual	Positive	TP	FN
	Negative	FP	TN

For multiclass classification, the confusion matrix extends to an N×N matrix, where N denotes the number of classes. It allows the identification of specific misclassification patterns and provides valuable insights into model behavior.

Receiver Operating Characteristic (ROC): curve illustrates the trade-off between the True Positive Rate (Sensitivity) and the False Positive Rate (FPR) across different classification thresholds.

The False Positive Rate is defined as:

$$\text{FPR} = \text{FP} / (\text{FP} + \text{TN}) \quad (\%) \dots \dots \dots (5)$$

whereas the True Positive Rate corresponds to Recall:

$$\text{TPR} = \text{TP} / (\text{TP} + \text{FN}) \quad (\%) \dots \dots \dots (6)$$

An ideal classifier produces a ROC curve approaching the upper-left corner of the graph,

indicating both high sensitivity and low false-positive rate.

Area Under the ROC Curve (AUC): summarizes the overall discrimination capability of the classifier independently of the decision threshold.

Mathematically :

$$0 \leq \text{AUC} \leq 1 \dots \dots \dots (7)$$

where:

AUC = 1 indicates perfect classification.

AUC = 0.5 corresponds to random guessing.

AUC < 0.5 indicates performance worse than random classification.

A higher AUC value reflects a stronger ability of the model to distinguish between different classes.

Overall performance assessment

The combination of these evaluation metrics provides a comprehensive assessment of the proposed deep learning models. While Accuracy measures the overall classification performance, Precision quantifies the reliability of positive predictions, Recall evaluates the ability to detect diseased patients, and the F1-score balances both measures. Furthermore, the Confusion Matrix offers a detailed analysis of classification errors, whereas the ROC curve and AUC assess the robustness and discriminative power of each model across different decision thresholds. Together, these metrics enable a rigorous comparison of the five CNN architectures and facilitate the identification of the most suitable model for lung cancer and COVID-19 diagnosis.

3.6 Graphical User Interface

The system was deployed as an interactive python graphical user interface (GUI). The GUI enables users to select the diagnostic type (COVID'19 or lung cancer), upload chest images, and receive instant diagnosis with confidence scores. In this interface, MobileNetV3 is integrated as the baseline model to process input images, chosen for its lightweight design and suitability for real-time applications.

3.7 Proposed framework and flowchart

The proposed PulmoDiagAI framework consists of a systematic pipeline encompassing data preprocessing, model training, and clinical deployment (Figure 1).

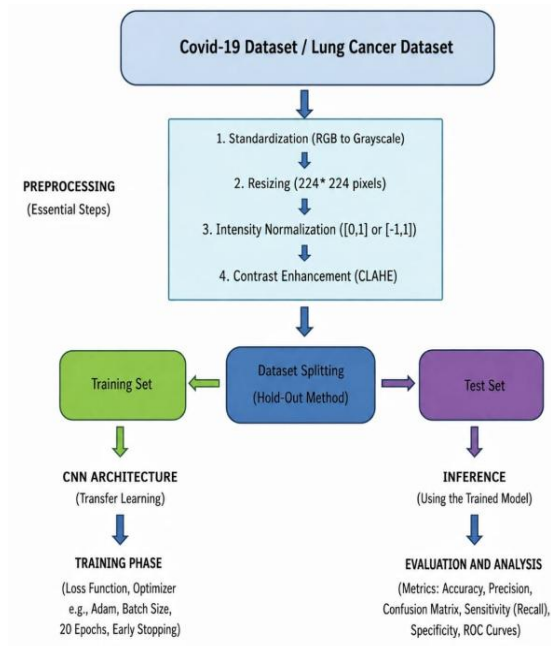


Figure 3: Overall methodology flowchart

Phase 1: Data preparation

- Collection and consolidation of medical datasets.
- Rigorous preprocessing (grayscale conversion, median filtering, resizing, CLAHE enhancement, and normalization).
- Data partitioning using hold-out validation (80/20 for COVID-QU-Ex, 70/30 for IQ-OTH/NCCD).

Phase 2: Model development

- Implementation of five pre-trained CNN architectures.
- Transfer learning with ImageNet weights.
- Custom top layers adapted for multi-class classification.
- Training with Adam optimizer and categorical cross-entropy loss.
- Learning rate optimization and dropout regularization.

Phase 3: Evaluation and analysis

- Performance assessment using accuracy, precision, recall, F1-score.
- Confusion matrix analysis for class-wise performance.
- ROC curve and AUC evaluation.
- Comparative ranking of architectures for each pathology.

Phase 4: Clinical deployment

- Selection of optimal model (MobileNet-V3) for GUI integration.

- Development of interactive interface using python.
- Real-time diagnosis with confidence scoring.
- Application for both COVID-19 and lung cancer detection.

4. Experimental results and discussions

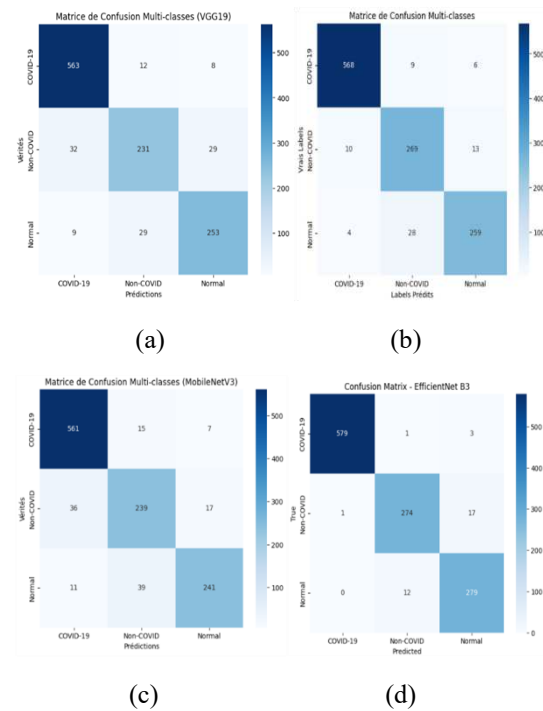
4.1 Preprocessing evaluation

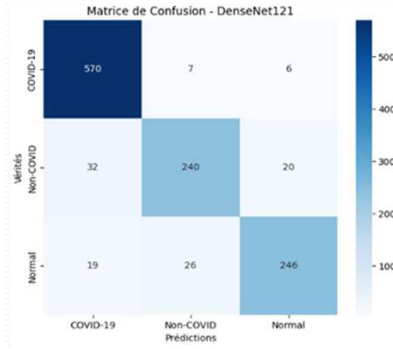
The effectiveness of preprocessing techniques was quantitatively assessed using PSNR, SNR, and MSE metrics. For COVID-QU-Ex images, the median filter with 3×3 kernel achieved superior results (PSNR: 36.94 dB, MSE: 13.12) compared to the 5×5 kernel (PSNR: 32.73 dB, MSE: 24.79). Similarly, for IQ-OTH/NCCD images, the 3×3 filter maintained higher PSNR (16.74 dB) and lower MSE (1375.03) than the 5×5 filter. These results confirm the selection of the 3×3 median filter for optimal noise reduction with minimal detail loss.

CLAHE enhancement visibly improved contrast and visual clarity of pathological features, including ground-glass opacities (COVID-19) and nodules (lung cancer). The method enhanced tissue differentiation while preserving diagnostically relevant details.

4.2 COVID-19 Classification results

Figure 4 presents confusion matrices for all five models on the COVID-QU-Ex dataset.





(c)

Figure 4: Confusion matrices for COVID'19: (a) VGG19, (b) ResNet-50, (c) MobileNet-V3, (d) EfficientNet-B3, (e) DenseNet-121

EfficientNet-B3 demonstrated superior performance, achieving the highest true positive counts across all classes: 579 for COVID-19 (4 false negatives), 274 for non-COVID (18 false negatives), and 279 for normal (12 false negatives). ResNet50 followed closely with 568, 269, and 259 true positives respectively. DenseNet121 showed moderate performance with 570 COVID-19 true positives but higher errors on non-COVID (52 false negatives) and normal (45 false negatives). MobileNetV3 and VGG19 exhibited the weakest performance, with VGG19 having 61 false negatives on non-COVID.

The confusion matrices obtained for the five CNN architectures are comparatively analyzed in Table 1. These matrices provide a comprehensive overview of the classification performance of each model by illustrating the distribution of correctly and incorrectly classified samples across all classes. They enable a detailed assessment of each architecture's discriminative capability and facilitate the identification of potential misclassification patterns. Consequently, the confusion matrix serves as an essential tool for evaluating the robustness and reliability of the proposed deep learning models.

Table 1. Comparative analysis of the classification performance of the five CNN architectures based on their confusion matrices.

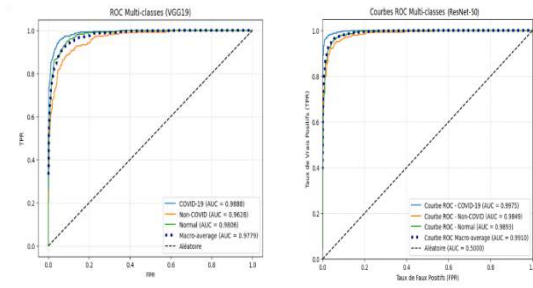
Model	COVID'19 (TP/FN)	Non-COVID (TP/FN)	Normal (TP/FN)
VGG19	563 / 20	231 / 61	253 / 38
ResNet50	568 / 15	269 / 23	259 / 32
EfficientNet B3	579 / 4	274 / 18	279 / 12
DenseNet121	570 / 13	240 / 52	246 / 45
MobileNetV3	561 / 22	239 / 53	241 / 50

TP = (True Positives): Number of correctly classified images belonging to the corresponding class.

FN (False Negatives): Number of images belonging to the corresponding class that were incorrectly classified as another class)

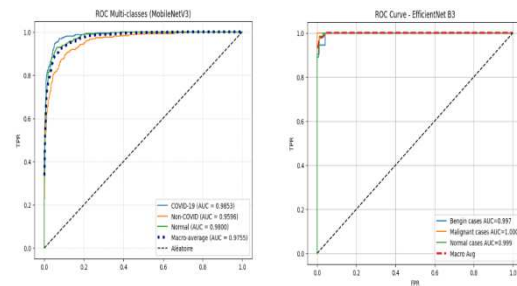
Table 1 summarizes, in overall, the confusion matrix rates. The analysis of this table clearly demonstrates that EfficientNet-B3 provides the most reliable multiclass classification performance, achieving the highest true-positive rates and the lowest false-negative rates across all diagnostic categories, making it the most suitable architecture for automated pulmonary disease diagnosis.

Analysis of the ROC curves, shown in Figure 5, confirmed EfficientNet-B3's superiority with macro-AUC of 0.9940, followed by ResNet50 (0.9910), DenseNet121 (0.9809), VGG19 (0.9779), and MobileNetV3 (0.9755).



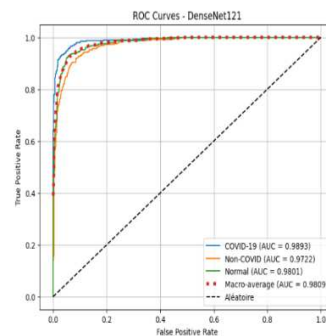
(a)

(b)



(c)

(d)



(e)

Figure 5: ROC curves for COVID'19 classifications: (a) VGG19, (b) ResNet-50, (c) MobileNet-V3, (d) EfficientNet-B3, (e) DenseNet-121

The Receiver Operating Characteristic (ROC) curves obtained for the five evaluated CNN architectures demonstrate consistently high classification performance, although noticeable differences can be observed among the models in terms of their discriminative capability across the three diagnostic classes.

EfficientNet-B3 achieved the best overall performance, with AUC values exceeding 0.99 for all classes and an exceptional AUC of 0.999 for the COVID'19 class. Its ROC curves rapidly approach the upper-left corner of the ROC space, indicating an almost perfect trade-off between sensitivity and specificity. This outstanding behavior confirms the model's excellent ability to distinguish COVID'19, Non-COVID Pneumonia, and Normal cases with minimal false-positive rates.

ResNet50 ranked second, exhibiting highly consistent and well-balanced performance with a macro-average AUC of 0.9910. The three ROC curves remain closely grouped, reflecting robust generalization across all diagnostic categories. Notably, the model achieved an AUC of 0.9975 for COVID'19 detections, representing the second-highest value among all evaluated architectures. These results demonstrate the effectiveness of residual learning in extracting highly discriminative radiological features while maintaining stable performance across multiple classes.

DenseNet121 achieved a macro-average AUC of 0.9809, with its highest discrimination capability observed for the COVID'19 class (AUC = 0.9893). However, its performance slightly decreased for the Non-COVID Pneumonia class (AUC = 0.9722), suggesting increased difficulty in distinguishing non-COVID pulmonary infections from other thoracic abnormalities due to overlapping radiographic characteristics.

VGG19 obtained a macro-average AUC of 0.9779. Although the model demonstrated satisfactory overall performance, its ROC curve for the Non-COVID Pneumonia class (AUC = 0.9628) exhibited a slight decline within the false-positive rate (FPR) interval of 0.05–0.20. This behavior indicates the presence of moderate confusion between Non-COVID Pneumonia and the remaining classes, leading to a slight reduction in classification accuracy within this region.

Among the evaluated models, MobileNetV3 produced the lowest overall discrimination

performance, with a macro-average AUC of 0.9755. Similar to VGG19 and DenseNet121, the Non-COVID Pneumonia class remained the most challenging category, yielding an AUC of 0.9596. This outcome is consistent with the lightweight design philosophy of MobileNetV3, which prioritizes computational efficiency and reduced model complexity over achieving maximum classification accuracy.

Table 2. Comparison in terms of AUC of the five CNN architectures based on COVID-QU-Ex.

Model	COVID'19	Non-COVID	Normal	Macro-avg
EfficientNet B3	0.999	0.991	0.993	0.994
ResNet50	0.9975	0.9849	0.9893	0.9910
DenseNet121	0.9893	0.9722	0.9801	0.9809
VGG19	0.9888	0.9628	0.9806	0.9779
MobileNetV3	0.9853	0.9596	0.9800	0.9755

Table 2. presents the Area Under the Receiver Operating Characteristic Curve (AUC-ROC) obtained by the five evaluated CNN architectures for the three diagnostic classes (COVID'19, Non-COVID, and Normal), together with their corresponding macro-average AUC values. The AUC metric provides a threshold-independent assessment of a model's discriminative capability, where values approaching 1.0 indicate excellent classification performance.

Among the evaluated models, EfficientNet-B3 achieved the highest overall performance, with a macro-average AUC of 0.994. It consistently outperformed the other architectures across all diagnostic categories, reaching an exceptional AUC of 0.999 for the COVID'19 class, 0.991 for Non-COVID Pneumonia, and 0.993 for the Normal class. These near-perfect AUC values demonstrate the remarkable ability of EfficientNet-B3 to accurately distinguish pulmonary pathologies while maintaining an excellent balance between sensitivity and specificity. The superior performance can be attributed to its compound scaling strategy, which efficiently optimizes network depth, width, and input resolution, enabling the extraction of highly discriminative image features.

ResNet50 ranked second with a macro-average AUC of 0.9910, demonstrating highly stable and balanced performance across all three classes. In particular, it achieved an AUC of 0.9975 for

COVID'19 detections, representing the second-best result among all evaluated architectures. The relatively small variation between its class-specific AUC values confirms the robustness of residual learning in capturing complex radiographic patterns while maintaining excellent generalization capability.

DenseNet121 obtained a macro-average AUC of 0.9809, indicating strong overall classification performance. The model achieved its highest discrimination capability for the COVID'19 class (AUC = 0.9893), whereas the Non-COVID Pneumonia class yielded the lowest score (AUC = 0.9722). This reduction suggests that DenseNet121 encounters greater difficulty in differentiating non-COVID pulmonary infections from other thoracic abnormalities, likely due to the substantial overlap in radiological manifestations among these diseases.

Similarly, VGG19 achieved a macro-average AUC of 0.9779, reflecting satisfactory but comparatively lower performance. Although the model maintained excellent discrimination for the COVID'19 (0.9888) and Normal (0.9806) classes, its performance decreased for the Non-COVID Pneumonia class (0.9628). This finding indicates a higher level of class overlap and misclassification for this category, highlighting the limitations of conventional deep convolutional architectures in learning highly discriminative features from complex medical images.

Among all evaluated architectures, MobileNetV3 produced the lowest macro-average AUC (0.9755). While its performance remained excellent for the COVID'19 (0.9853) and Normal (0.9800) classes, the Non-COVID Pneumonia category again represented the most challenging classification task, with an AUC of 0.9596. This outcome is consistent with the lightweight architecture of MobileNetV3, which was primarily designed to minimize computational complexity and inference time, resulting in a slight reduction in discriminative performance compared with larger and deeper CNN models.

Overall, the results clearly indicate that COVID-19 is the most easily distinguishable class for all evaluated architectures, with AUC values consistently exceeding 0.98. Conversely, the Non-COVID Pneumonia class remains the most difficult category to classify accurately due to the strong visual similarity between various pulmonary infections and inflammatory conditions. Nevertheless, all models achieved macro-average AUC values above 0.97, demonstrating excellent overall diagnostic capability. Among them, EfficientNet-B3 consistently delivered the highest discrimination performance across all classes,

confirming its superiority as the most reliable deep learning architecture for automated pulmonary disease classification in the proposed PulmoDiagAI framework.

Table 3. Comparison of the five CNN architecture models in terms of Precision, Recall, F1 score, and Accuracy.

Metrics	Class		
	COVID'19	Non-COVID	Normal
DenseNet121			
Precision	0.92	0.88	0.90
Recall	0.98	0.82	0.85
F1-score	0.95	0.85	0.87
Accuracy	0.90		
EfficientNet B3			
Precision	1.00	0.95	0.93
Recall	0.99	0.94	0.96
F1-score	1.00	0.95	0.95
Accuracy	0.97		
MobileNetV3			
Precision	0.92	0.82	0.91
Recall	0.96	0.82	0.83
F1-score	0.94	0.82	0.87
Accuracy	0.89		
ResNet50			
Precision	0.98	0.88	0.93
Recall	0.97	0.92	0.89
F1-score	0.97	0.90	0.91
Accuracy	0.94		
VGG19			
Precision	0.93	0.85	0.87
Recall	0.97	0.79	0.87
F1-score	0.95	0.82	0.87
Accuracy	0.90		

Table 3 summarizes the classification performance of the five evaluated CNN architectures in terms of Precision, Recall, F1-score, and Overall Accuracy for the three diagnostic classes (COVID'19, Non-COVID, and Normal). These metrics provide a comprehensive assessment of the models' ability to correctly identify pulmonary diseases while minimizing both false-positive and false-negative predictions.

Among all evaluated architectures, EfficientNet-B3 achieved the best overall performance, reaching an overall accuracy of 97%, the highest obtained in this study. It also exhibited the most balanced classification performance across all classes, with a precision of 1.00, recall of 0.99, and F1-score of 1.00 for the COVID'19 class. These outstanding results indicate that EfficientNet-B3 is capable of detecting COVID'19 cases with almost perfect reliability while producing virtually no false-positive predictions. Furthermore, the model maintained excellent performance for the Non-COVID Pneumonia class, achieving a precision of 0.95, recall of 0.94, and F1-score of 0.95, as well as strong performance for the Normal class (precision

= 0.93, recall = 0.96, F1-score = 0.95). This balanced behavior demonstrates the superior feature extraction capability and robustness of EfficientNet-B3 in multiclass pulmonary disease classification.

ResNet50 ranked second with an overall accuracy of 94%, confirming the effectiveness of residual learning for medical image analysis. The model achieved excellent performance for COVID'19, with a precision of 0.98, recall of 0.97, and F1-score of 0.97. For the Non-COVID Pneumonia class, ResNet50 obtained a recall of 0.92, which is the second-highest among all evaluated architectures, indicating its strong capability to correctly identify pneumonia cases while maintaining a precision of 0.88. Performance for the Normal class also remained consistently high, with an F1-score of 0.91, highlighting the model's balanced generalization across all diagnostic categories.

DenseNet121 achieved an overall accuracy of 90%, with particularly strong performance for the COVID'19 class, where it reached a recall of 0.98, demonstrating its high sensitivity in identifying infected patients. However, its performance decreased for the Non-COVID Pneumonia class, yielding a precision of 0.88, recall of 0.82, and F1-score of 0.85. This reduction suggests that DenseNet121 encounters greater difficulty in distinguishing non-COVID pulmonary infections from other thoracic abnormalities that exhibit similar radiographic features.

Similarly, VGG19 obtained an overall accuracy of 90%. Although it achieved satisfactory performance for the COVID'19 class (precision = 0.93, recall = 0.97, F1-score = 0.95), its classification capability decreased considerably for Non-COVID Pneumonia, where the recall dropped to 0.79, representing the lowest value among all evaluated models. This finding indicates that VGG19 tends to miss a larger proportion of pneumonia cases, thereby increasing the number of false-negative predictions. Performance for the Normal class remained moderate, with both recall and F1-score equal to 0.87.

Among the evaluated architectures, MobileNetV3 achieved the lowest overall accuracy (89%). Despite its competitive performance for the COVID'19 class (precision = 0.92, recall = 0.96, F1-score = 0.94), the model showed limited discriminative capability for the Non-COVID Pneumonia class, with both precision and recall equal to 0.82. Its performance also decreased for the Normal class, where the recall reached 0.83 and the F1-score 0.87. These results are consistent with the lightweight design philosophy of MobileNetV3, which prioritizes computational efficiency and

reduced model complexity over maximum classification accuracy.

Overall, the quantitative evaluation demonstrates that COVID'19 is the most accurately recognized class across all evaluated CNN architectures, consistently achieving the highest precision, recall, and F1-score values. In contrast, the Non-COVID Pneumonia class remains the most challenging category, owing to the significant radiographic similarities between different pulmonary infections and inflammatory conditions. Among all investigated models, EfficientNet-B3 consistently outperformed the competing architectures across every evaluation metric, confirming its superior robustness, discriminative capability, and clinical applicability for automated pulmonary disease diagnosis. The superior classification metrics obtained by EfficientNet-B3 are fully consistent with the previously reported confusion matrices and ROC-AUC analysis, further validating its effectiveness as the core deep learning model integrated into the proposed PulmoDiagAI computer-aided diagnosis system.

4.2.1 Comparative study with state-of-the-art literature on COVID-QU-Ex Dataset

In Table 04, we present a comparative performance evaluation between the five evaluated CNN approaches (DenseNet121, EfficientNet B3, MobileNetV3, ResNet50, and VGG19) and existing state-of-the-art methods from the literature, including H-SGE-Seq2D [18], LitCovNet [19], LungMaxViT [20], and U-Net [21], on the COVID-QU-Ex dataset. Among the evaluated models, EfficientNet B3 demonstrates superior performance, achieving the highest classification accuracy of 0.97. This matches the top-performing state-of-the-art approach by Schraut et al. (U-Net) [21] (0.97) and outperforms advanced hybrid models such as LungMaxViT by Fu et al. [20] (0.96), LitCovNet by Hossain et al. [19] (0.95), and H-SGE-Seq2D by Mustapha et al. [18] (0.94). Furthermore, ResNet50 yields a competitive accuracy of 0.94, on par with recent attention-guided architectures, while DenseNet121 and VGG19 both achieve a solid 0.90 accuracy. Although lightweight models like MobileNetV3 show a slight drop in accuracy (0.89), they still offer a highly efficient trade-off for resource-constrained environments. These results confirm the effectiveness of compound-scaled architectures like EfficientNet B3 in extracting fine-grained visual representations from COVID-QU-Ex chest X-ray images.

Table 4. Comparative study on COVID-QU-Ex dataset

Approach	Accuracy
DenseNet121	0.90
EfficientNet B3	0.97
MobileNetV3	0.89
ResNet50	0.94
VGG19	0.90
H-SGE-Seq2D [18]	0.94
LitCovNet [19]	0.95
LungMax ViT [20]	0.96
U-Net [21]	0.97

4.3 Lung Cancer Classification Results

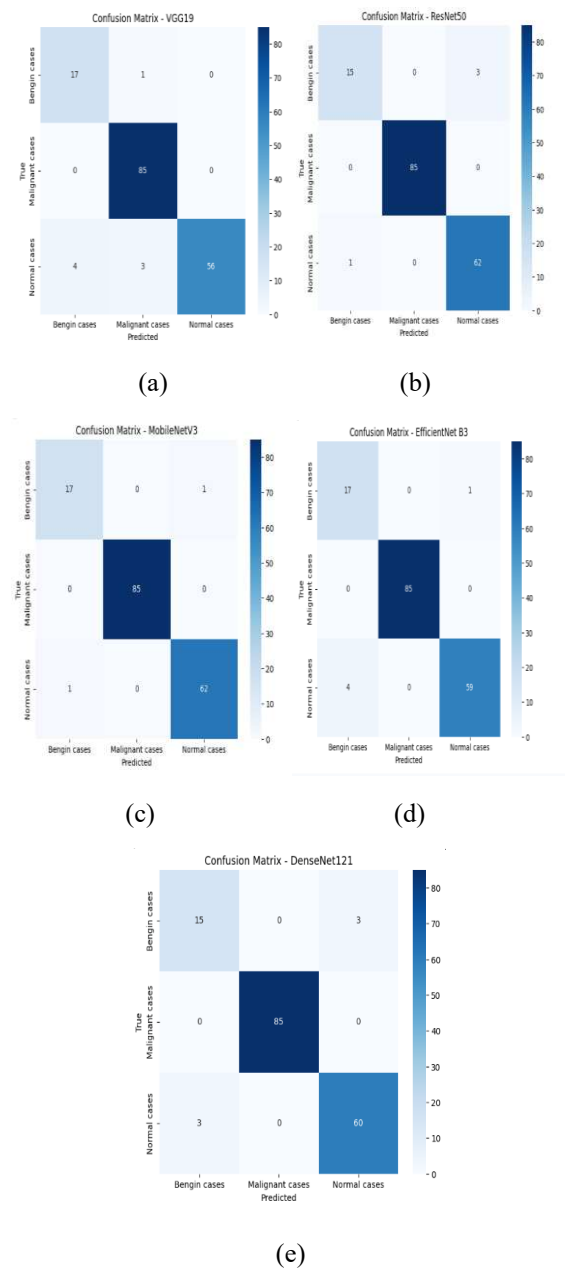


Figure 6: Confusion matrices for lung cancer: (a) VGG19, (b) ResNet-50, (c) MobileNet-V3, (d) EfficientNet-B3, (e) DenseNet-121

Figure 6 presents the confusion matrices obtained by the five evaluated convolutional neural network (CNN) architectures on the IQ-OTH/NCCD lung cancer dataset. The confusion matrices provide detailed insights into the classification behavior of each model by illustrating both correctly classified samples and misclassification patterns, thereby enabling a comprehensive assessment of their diagnostic reliability.

Overall, the five architectures demonstrated excellent performance in identifying the Malignant class, which represents the most clinically significant category due to its direct impact on patient prognosis. Remarkably, ResNet50, EfficientNet-B3, MobileNetV3, DenseNet121, and VGG19 all achieved 100% sensitivity for malignant tumor detection, correctly classifying all 85 malignant cases without producing any false negatives. From a clinical perspective, this finding is of paramount importance, as false-negative diagnoses may delay treatment initiation and significantly compromise patient survival.

Among the evaluated models, ResNet50 exhibited the most balanced classification performance. In addition to perfectly detecting all malignant cases, it correctly classified 62 of the 63 normal images, with only one normal case misclassified as benign. For the Benign category, only three samples were incorrectly assigned to the Normal class, indicating excellent discrimination capability and strong generalization across all categories. These results demonstrate the effectiveness of residual learning in extracting highly discriminative imaging features while maintaining stable performance under multiclass classification.

EfficientNet-B3 achieved performance comparable to ResNet50, confirming the effectiveness of compound model scaling for medical image analysis. Similar to ResNet50, it detected all malignant tumors correctly without generating false negatives. Furthermore, it correctly identified 17 out of 18 benign cases, with only one benign lesion misclassified as Normal, demonstrating excellent sensitivity for early-stage lesions. The only noticeable reduction in performance was observed for the Normal class, where 59 of 63 images were correctly classified, while four normal images were incorrectly predicted as benign. Although these false-positive predictions may lead to additional clinical investigations, they remain preferable to false-negative cancer diagnoses in practical screening applications.

MobileNetV3 produced particularly noteworthy results considering its lightweight architecture. Despite its substantially lower computational complexity, the model achieved perfect malignant tumor detection (85/85) while correctly classifying

62 normal cases, matching the best-performing architecture for this class. Only one benign sample was misclassified as Normal, resulting in 17 correctly identified benign lesions. These findings demonstrate that MobileNetV3 successfully combines high diagnostic reliability with computational efficiency, making it especially suitable for deployment in portable medical devices, edge computing platforms, and real-time clinical decision-support systems where hardware resources are limited.

DenseNet121 also demonstrated excellent performance for malignant tumor detection, achieving 100% recall for this critical category. However, compared with ResNet50 and EfficientNet-B3, its classification accuracy slightly decreased for the remaining categories. The model correctly classified 60 normal cases, while three normal images were misclassified as benign. Likewise, three benign lesions were incorrectly predicted as Normal, suggesting that DenseNet121 experiences greater difficulty in distinguishing subtle radiological characteristics between benign lesions and healthy pulmonary tissues.

Although VGG19 successfully identified all malignant tumors, its overall discriminative performance was comparatively lower than that of the other architectures. The model correctly classified 17 benign cases, but incorrectly assigned one benign lesion to the Malignant class, generating an unnecessary false-positive cancer diagnosis. Moreover, VGG19 correctly identified only 56 normal cases, while seven normal images were misclassified, including four predicted as benign and three as malignant. These findings indicate that the relatively shallow feature extraction capability of VGG19 limits its effectiveness in capturing subtle inter-class variations compared with more recent deep learning architectures.

The comparative analysis of the confusion matrices clearly demonstrates that ResNet50, EfficientNet-B3, and MobileNetV3 constitute the three most reliable architectures for automated lung tumor classification. Beyond achieving 100% recall for malignant tumors, they maintain very low misclassification rates for benign and normal cases, thereby ensuring a high level of diagnostic safety. Among these models, the outstanding performance of MobileNetV3 is particularly noteworthy because it achieves diagnostic accuracy comparable to considerably larger networks while requiring substantially fewer computational resources. This favorable balance between predictive performance and computational efficiency makes MobileNetV3 an attractive candidate for embedded artificial intelligence applications, mobile healthcare systems, and point-of-care computer-aided diagnosis (CAD) solutions.

Overall, the confusion matrix analysis confirms the robustness and clinical applicability of the proposed deep learning framework. The complete elimination of false-negative malignant diagnoses across all evaluated architectures represents a critical achievement from a clinical standpoint, as maximizing cancer detection sensitivity remains the primary objective of intelligent diagnostic systems designed to support radiologists and improve early lung cancer diagnosis.

Table 5. Comparative analysis of the classification performance of the five CNN architectures based on their confusion matrices.

Model	Benign TP/FN	Malignant TP/FN	Normal TP/FN	Total errors
ResNet50	15 / 3	85 / 0	62 / 1	4
EfficientNet B3	17 / 1	85 / 0	59 / 4	5
DenseNet121	15 / 3	85 / 0	60 / 3	6
VGG19	17 / 1	85 / 0	56 / 7	8
MobileNetV3	17 / 1	85 / 0	62 / 1	2
TP = True Positives (Number of images correctly classified in their true class)				
FN = False Negatives (Number of images in this class that were incorrectly classified)				

Table 5 summarizes the classification performance of the five evaluated convolutional neural network (CNN) architectures based on the number of True Positives (TP) and False Negatives (FN) obtained for the three diagnostic categories, namely Benign, Malignant, and Normal, together with the total number of classification errors. This analysis provides a detailed assessment of the diagnostic reliability of each model and highlights their ability to correctly identify clinically relevant pulmonary lesions.

A remarkable observation from Table 4 is that all five CNN architectures achieved perfect detection of malignant tumors, correctly identifying all 85 malignant cases without generating any false negatives. From a clinical perspective, this represents an outstanding result because missing a malignant lesion may delay diagnosis and treatment, potentially leading to severe consequences for patient survival. Consequently, all evaluated models achieved 100% sensitivity for the most critical diagnostic category, demonstrating their strong potential for computer-aided lung cancer screening.

Despite this common strength, noticeable differences emerge in the classification of Benign and Normal cases. Among all evaluated architectures, MobileNetV3 achieved the best overall performance, producing only two

classification errors across the entire test dataset. The model correctly identified 17 of the 18 benign lesions, misclassifying only one benign case as normal, while simultaneously recognizing 62 of the 63 normal images, with a single misclassification. These results demonstrate that MobileNetV3 successfully combines excellent diagnostic accuracy with a lightweight architecture, making it particularly attractive for real-time and resource-constrained clinical applications.

ResNet50 ranked second with four total classification errors. Besides achieving perfect recognition of malignant tumors, it correctly classified 62 normal images while misclassifying only one normal case as benign. The model also correctly identified 15 benign lesions, with three benign samples incorrectly assigned to the Normal category. This balanced performance confirms the effectiveness of residual learning in extracting robust and discriminative image representations, resulting in excellent generalization across different pathological categories.

EfficientNet-B3, although producing five classification errors, demonstrated outstanding discrimination capability for benign lesions by correctly identifying 17 of 18 benign cases, corresponding to only one false negative. Similar to ResNet50 and MobileNetV3, it achieved perfect malignant tumor detection. However, its performance slightly decreased for the Normal category, where 59 of the 63 normal images were correctly classified, while four normal cases were predicted as benign. From a clinical standpoint, these false-positive findings are generally less critical than false-negative cancer diagnoses, since they primarily result in additional medical examinations rather than missed malignant lesions.

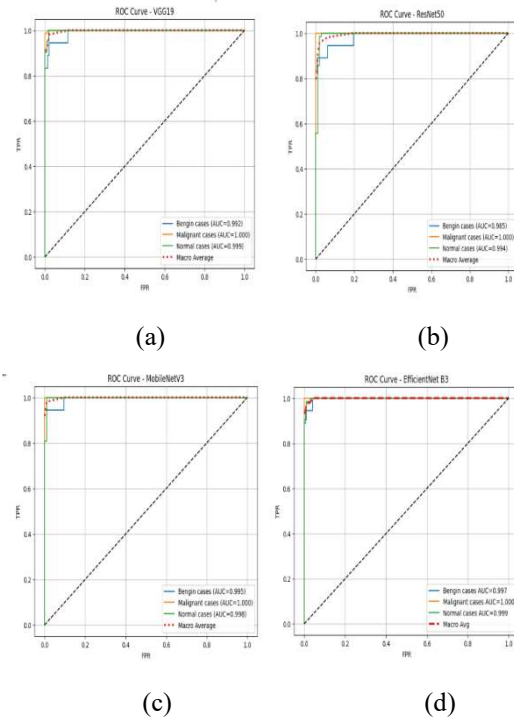
DenseNet121 achieved similarly excellent performance for malignant tumors but exhibited a moderate reduction in classification accuracy for the remaining classes. The model correctly classified 15 benign lesions and 60 normal images, while producing three misclassifications in each category. These results suggest that DenseNet121 encounters greater difficulty in separating benign pulmonary abnormalities from normal lung tissue, likely because of overlapping radiological characteristics that challenge feature discrimination.

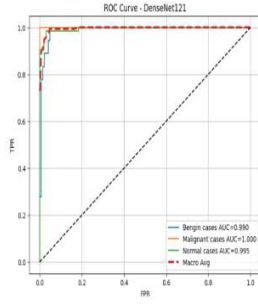
Among the evaluated models, VGG19 produced the largest number of classification errors (eight misclassified images). Although it successfully detected all malignant tumors, its performance for the Normal class was noticeably lower, correctly recognizing only 56 of the 63 normal images, while seven normal cases were misclassified. The model also correctly identified 17 benign lesions, missing only one benign case. The relatively higher error

rate observed for VGG19 can be attributed to its conventional architecture, which lacks the advanced feature reuse and optimization mechanisms introduced in more recent deep learning models.

Overall, the comparative analysis clearly demonstrates that MobileNetV3 provides the most favorable trade-off between diagnostic accuracy and computational efficiency, achieving the lowest overall number of classification errors while preserving perfect malignant tumor detection. ResNet50 and EfficientNet-B3 also delivered outstanding performance, confirming their robustness and excellent discriminative capability for multiclass lung tumor classification. More importantly, the absence of false-negative malignant diagnoses across all five architectures represents a highly significant clinical achievement, as maximizing sensitivity for cancer detection remains the primary objective of computer-aided diagnostic systems. These findings strongly support the integration of the proposed deep learning framework into intelligent clinical decision-support systems aimed at improving the early diagnosis of lung cancer while minimizing the risk of missed malignant cases.

ROC curves, given in Figure 7, showed near-perfect AUC values across models. ResNet50 and MobileNetV3 achieved 1.000 for malignant cases and macro-AUC of 0.997 and 0.993 respectively.





(e)

Figure 7: ROC curves for lung cancer classifications: (a) VGG19, (b) ResNet-50, (c) MobileNet-V3, (d) EfficientNet-B3, (e) DenseNet-121

Figure 7 illustrates the Receiver Operating Characteristic (ROC) curves obtained for the five evaluated convolutional neural network (CNN) architectures. Overall, all models exhibited excellent discriminative performance, with Area Under the Curve (AUC) values approaching the theoretical optimum. Nevertheless, noticeable differences emerged regarding their ability to separate the three diagnostic classes (Benign, Malignant, and Normal), particularly for the benign category, which represents the most challenging classification task.

Among the evaluated architectures, ResNet50 and MobileNetV3 demonstrated the best overall discriminative performance. Both models achieved perfect or near-perfect AUC values ($AUC \approx 1.000$) across all diagnostic categories, producing ROC curves that rapidly rise toward the upper-left corner of the ROC space. This behavior indicates an almost ideal balance between sensitivity and specificity, with extremely low false-positive rates while maintaining maximal true-positive rates. The nearly vertical trajectories observed at the lowest false-positive rate (FPR) values confirm the outstanding capability of these two architectures to establish highly discriminative decision boundaries. While the superior performance of ResNet50 is consistent with its well-established robustness in medical image analysis, the remarkable performance of MobileNetV3 is particularly noteworthy considering its lightweight architecture and significantly lower computational complexity. These findings demonstrate that an optimized compact network can achieve diagnostic performance comparable to considerably deeper and computationally demanding CNN models.

EfficientNet-B3 also achieved outstanding classification performance, producing ROC curves that closely approach the ideal classifier. The model reached a perfect AUC of 1.000 for the Malignant class, while achieving AUC values of 0.999 and

0.997 for the Normal and Benign classes, respectively. A slight deviation from the ideal ROC trajectory can be observed for the benign category at very low false-positive rates, indicating that distinguishing benign pulmonary lesions from normal lung tissues remains the most challenging aspect of the classification task. Nevertheless, the observed reduction is minimal and confirms the remarkable robustness of EfficientNet-B3 in multiclass lung tumor diagnosis.

Despite being one of the earliest deep CNN architectures, VGG19 maintained excellent discriminative capability, particularly for the Malignant class, where it also achieved a perfect AUC of 1.000. The model further obtained AUC values of 0.999 for the Normal class and 0.992 for the Benign class. Although these values remain exceptionally high, the ROC curve associated with benign lesions exhibits a slight deviation from the ideal trajectory, suggesting a modest reduction in sensitivity when differentiating benign abnormalities from healthy pulmonary tissues. From a clinical perspective, however, this decrease remains relatively minor and does not substantially compromise the overall diagnostic performance of the model.

Among the evaluated architectures, DenseNet121 produced the lowest, yet still excellent, discrimination performance. Similar to the other CNN models, it perfectly detected malignant tumors ($AUC = 1.000$), confirming its excellent capability for identifying the clinically most significant category. However, the model achieved slightly lower AUC values for the Normal (0.995) and Benign (0.990) classes. The corresponding ROC curve exhibits a more noticeable deviation within the FPR range of approximately 0.05–0.15, indicating increased overlap between benign lesions and normal pulmonary structures. This behavior suggests that DenseNet121 experiences greater difficulty in learning highly discriminative features for subtle morphological differences between these two categories.

Overall, the ROC analysis strongly corroborates the observations derived from the confusion matrices and quantitative performance metrics. All evaluated architectures demonstrated excellent discrimination capability, particularly for the Malignant class, where perfect AUC values were consistently achieved. However, ResNet50 and MobileNetV3 emerged as the most reliable classifiers by combining near-perfect ROC characteristics, maximum AUC values, and exceptionally low false-positive rates across all diagnostic categories. The outstanding performance achieved by MobileNetV3 is especially significant because it matches the predictive capability of considerably larger architectures while maintaining substantially

lower computational requirements. This favorable balance between diagnostic accuracy and computational efficiency makes MobileNetV3 particularly attractive for deployment in embedded artificial intelligence systems, mobile healthcare platforms, and real-time computer-aided diagnosis (CAD) applications. Collectively, these findings confirm the robustness, reliability, and clinical applicability of the proposed deep learning framework for automated lung tumor classification and further support its potential integration into intelligent clinical decision-support systems for early lung cancer diagnosis.

Table 6: Comparison in terms of AUC of the five CNN architectures based on IQ-OTH/NCCD dataset.

Model	Benign	Malignant	Normal	Macro-avg
EfficientNet B3	0.997	1.000	0.999	0.998
ResNet50	0.985	1.000	0.994	0.993
DenseNet121	0.990	1.000	0.995	0.995
VGG19	0.992	1.000	0.999	0.997
MobileNetV3	0.995	1.000	0.998	0.997

Table 6 presents the comparative Area Under the Receiver Operating Characteristic Curve (AUC-ROC) values obtained by the five evaluated convolutional neural network (CNN) architectures on the IQ-OTH/NCCD lung cancer dataset. The AUC is one of the most informative threshold-independent evaluation metrics, measuring the ability of a classifier to distinguish between different diagnostic categories regardless of the selected decision threshold. An AUC value close to 1.000 indicates an almost perfect discriminative capability.

The results demonstrate that all evaluated architectures achieved outstanding diagnostic performance, with macro-average AUC values ranging from 0.993 to 0.998, confirming the effectiveness of deep transfer learning for automated lung tumor classification from CT images.

Among all models, EfficientNet-B3 achieved the highest overall performance with a macro-average AUC of 0.998. The model obtained AUC values of 0.997, 1.000, and 0.999 for the Benign, Malignant, and Normal classes, respectively. The perfect discrimination achieved for malignant tumors demonstrates the model's exceptional capability to identify the clinically most critical cases without ambiguity. Moreover, its near-perfect performance for benign and normal tissues confirms the effectiveness of EfficientNet's compound scaling

strategy in learning highly discriminative and robust feature representations across all diagnostic categories.

VGG19 and MobileNetV3 ranked second, both achieving a macro-average AUC of 0.997, although their class-wise performances differed slightly. VGG19 reached AUC values of 0.992, 1.000, and 0.999 for the Benign, Malignant, and Normal classes, respectively, demonstrating remarkable stability despite its relatively simple and conventional architecture. In contrast, MobileNetV3 achieved superior discrimination for the Benign class (AUC = 0.995) while maintaining nearly perfect performance for the Normal (0.998) and Malignant (1.000) categories. These findings are particularly significant because MobileNetV3 reaches a level of diagnostic accuracy comparable to considerably deeper architectures while requiring substantially fewer computational resources, making it highly suitable for real-time and edge-based medical applications.

DenseNet121 achieved a macro-average AUC of 0.995, confirming its strong overall discriminative capability. The model perfectly identified malignant tumors (AUC = 1.000) and maintained excellent performance for the Normal (0.995) and Benign (0.990) classes. Although slightly lower than EfficientNet-B3 and MobileNetV3, these results demonstrate the effectiveness of dense feature propagation in preserving diagnostic information throughout the network while maintaining excellent classification reliability.

Among the evaluated architectures, ResNet50 obtained the lowest macro-average AUC (0.993), despite achieving perfect discrimination of malignant tumors (AUC = 1.000). Its performance remained excellent for the Normal class (AUC = 0.994), whereas the Benign category produced the lowest class-specific AUC (0.985) among all models. This slight decrease suggests that residual learning, while extremely effective for detecting malignant lesions, encounters greater difficulty in separating benign abnormalities from normal pulmonary structures due to the subtle morphological similarities between these classes.

A noteworthy observation is that all five CNN architectures achieved an identical AUC of 1.000 for the Malignant class, indicating perfect separability between malignant tumors and the remaining categories. This finding highlights the strong discriminative power of deep convolutional networks for detecting malignant pulmonary lesions and emphasizes the robustness of the proposed preprocessing and transfer learning strategy. Conversely, the Benign class consistently yielded the lowest AUC values across all architectures, suggesting that benign lesions remain the most

challenging category to classify because of their greater radiological similarity to healthy lung tissues.

Overall, the AUC analysis confirms that the proposed deep learning framework provides near-perfect diagnostic discrimination on the IQ-OTH/NCCD dataset. While all evaluated CNN architectures achieved outstanding classification performance, EfficientNet-B3 emerged as the most reliable model by consistently producing the highest macro-average AUC and maintaining excellent discrimination across all diagnostic categories. The competitive performance achieved by MobileNetV3 is equally noteworthy, demonstrating that lightweight CNN architectures can deliver diagnostic accuracy comparable to larger and more computationally demanding models. These findings are fully consistent with the previously reported confusion matrices and classification metrics, further validating the robustness, reliability, and clinical applicability of the proposed PulmoDiagAI computer-aided diagnosis framework for automated lung cancer detection.

Table 7: Comparison, in terms of Precision, Recall, F1 score, and Accuracy, between the five CNN architecture models.

Metrics	Class		
	Benign	Malignant	Normal
DenseNet121			
Precision	0.83	1.00	0.95
Rappel	0.83	1.00	0.95
F1-score	0.83	1.00	0.95
Accuracy	0.96		
EfficientNet B3			
Precision	0.81	1.00	0.98
Rappel	0.94	1.00	0.94
F1-score	0.87	1.00	0.96
Accuracy	0.97		
MobileNetV3			
Precision	0.94	1.00	0.98
Rappel	0.94	1.00	0.98
F1-score	0.94	1.00	0.98
Accuracy	0.99		
ResNet50			
Precision	0.94	1.00	0.95
Rappel	0.83	1.00	0.98
F1-score	0.88	1.00	0.97
Accuracy	0.98		
VGG19			
Precision	0.81	0.96	1.00
Rappel	0.94	1.00	0.89
F1-score	0.87	0.98	0.94
Accuracy	0.95		

Table 7 summarizes the quantitative performance of the five evaluated CNN architectures in terms of

Precision, Recall, F1-score, and Overall Accuracy on the IQ-OTH/NCCD lung cancer dataset. These metrics provide complementary insights into the diagnostic capability of each model by evaluating classification reliability, sensitivity, and the balance between false-positive and false-negative predictions across the three diagnostic categories (Benign, Malignant, and Normal).

Overall, all evaluated architectures achieved excellent performance, with overall accuracies ranging from 95% to 99%, confirming the effectiveness of transfer learning for multiclass lung tumor classification. A particularly noteworthy observation is that all models achieved a recall of 1.00 for the Malignant class, indicating that every malignant tumor was correctly identified. From a clinical perspective, this result is highly significant because it demonstrates the complete elimination of false-negative malignant diagnoses, thereby maximizing patient safety and reducing the risk of delayed therapeutic intervention.

Among the evaluated architectures, MobileNetV3 achieved the best overall performance, reaching an overall accuracy of 99%. The model obtained precision, recall, and F1-score values of 0.94 for the Benign class while simultaneously achieving perfect performance (1.00) for the Malignant class. For the Normal category, MobileNetV3 further demonstrated outstanding discrimination with precision, recall, and F1-score values of 0.98. These highly balanced results indicate that MobileNetV3 successfully minimizes both false-positive and false-negative predictions while maintaining exceptional diagnostic reliability. Considering its lightweight architecture and low computational complexity, these findings highlight its strong potential for deployment in real-time clinical environments and resource-constrained medical devices.

ResNet50 ranked second with an overall accuracy of 98%. Similar to MobileNetV3, it achieved perfect precision and recall (1.00) for malignant tumor detection, confirming its excellent sensitivity for identifying the clinically most critical category. The model also demonstrated high precision (0.94) for benign lesions and excellent recall (0.98) for the Normal class, resulting in an F1-score of 0.97. However, the lower recall (0.83) observed for benign lesions indicates that a small proportion of benign tumors were misclassified, suggesting that distinguishing benign abnormalities from normal pulmonary structures remains the principal source of classification error.

EfficientNet-B3 achieved an overall accuracy of 97%, confirming its excellent classification capability. The model correctly identified all malignant tumors with perfect precision, recall, and

F1-score values of 1.00. For the Normal class, it achieved a precision of 0.98 and an F1-score of 0.96, reflecting strong discrimination between healthy lung tissues and pathological findings. Nevertheless, the comparatively lower precision (0.81) for benign lesions indicates the presence of a limited number of false-positive predictions, although the high recall (0.94) demonstrates that most benign cases were successfully detected.

DenseNet121 obtained an overall accuracy of 96%, exhibiting stable and well-balanced performance across all diagnostic categories. The model achieved perfect classification of malignant tumors while maintaining identical precision, recall, and F1-score values of 0.83 for benign lesions and 0.95 for the Normal class. These consistent values indicate balanced classification behavior, although its performance remains slightly below that of MobileNetV3 and ResNet50, particularly for benign tumor discrimination.

Among the evaluated architectures, VGG19 produced the lowest overall accuracy (95%). Although the model achieved perfect recall (1.00) for malignant tumors and perfect precision (1.00) for the Normal class, its recall for the Normal category decreased to 0.89, indicating that several normal images were incorrectly classified as pathological. Furthermore, the model exhibited the lowest precision (0.81) for benign lesions, resulting in an F1-score of 0.87. These findings suggest that the relatively shallow feature extraction capability of VGG19 limits its effectiveness in distinguishing subtle morphological differences between benign lesions and normal pulmonary tissues.

A general trend observed across all evaluated CNN architectures is that the Malignant class consistently achieved the highest classification performance, with perfect or near-perfect values for all evaluation metrics. This reflects the distinct radiological characteristics of malignant lung tumors, which facilitate reliable feature extraction by deep convolutional networks. In contrast, the Benign class remained the most challenging category, exhibiting comparatively lower precision and recall values across most architectures due to the considerable visual similarity between benign lesions and normal pulmonary tissues.

Overall, the comparative evaluation demonstrates that MobileNetV3 provides the best compromise between diagnostic accuracy, robustness, and computational efficiency, outperforming all competing architectures with an overall accuracy of 99% while maintaining consistently high precision, recall, and F1-score values across all diagnostic classes. The superiority of MobileNetV3, together with its lightweight design, makes it particularly attractive for integration into intelligent computer-

aided diagnosis (CAD) systems intended for real-time lung cancer screening. These findings are fully consistent with the confusion matrix and ROC-AUC analyses, further confirming the robustness and clinical applicability of the proposed PulmoDiagAI framework for automated lung tumor diagnosis.

4.3.1 Comparative study with state-of-the-art literature on IQ-OTH/NCCD dataset

In this section, we present a comparative study of various convolutional neural network architectures and baseline approaches for lung cancer classification on computed tomography (CT) images. Specifically, Table 08 presents a comparative performance evaluation between the five evaluated CNN approaches (DenseNet121, EfficientNet B3, MobileNetV3, ResNet50, and VGG19) and existing state-of-the-art methods from the literature, including AlexNet [22], Gabor-GLCM + SVM [23], GoogLeNet [24], and Self-Attention-based VGG16 [25], on the IQ-OTH/NCCD lung cancer dataset. Among the CNN implementations, ResNet50 achieves the highest overall accuracy of 0.98, outperforming both lightweight architectures like MobileNetV3 (0.97) and modern optimized backbones such as EfficientNet B3 (0.97). Deeper models like DenseNet121 and VGG19 also demonstrate competitive results with classification accuracies of 0.96 and 0.95, respectively. When benchmarking against state-of-the-art literature on the same dataset, all five evaluated CNN architectures consistently exceed traditional feature extraction methods, such as the Gabor-GLCM + SVM approach by Kareem et al. [23] (0.89), as well as earlier deep frameworks like AlexNet [22] (0.93) and GoogLeNet [24] (0.94). Furthermore, our top-performing model (ResNet50) slightly surpasses the recent Self-Attention-based VGG16 network proposed by Ghosh & Chatterjee [25] (0.97). These findings highlight the efficacy of residual learning in capturing complex features within CT slices, offering a robust balance between feature representation and diagnostic accuracy for lung cancer detection.

Table 8: Comparative study on IQ-OTH/NCCD lung cancer dataset

Approach	Accuracy
DenseNet121	0.96
EfficientNet B3	0.97
MobileNetV3	0.97
ResNet50	0.98
VGG19	0.95
AlexNet [22]	0.93
Gabor-GLCM + SVM [23]	0.89
GoogLeNet [24]	0.94
Self-Attention-based VGG16 [25]	0.97

4.4 Deployment of the Proposed Computer-Aided Diagnosis System

Following the comparative evaluation of the five convolutional neural network (CNN) architectures, the best-performing model was deployed within an interactive Graphical User Interface (GUI) developed using Google Colab Graphical User Interface (Colab GUI). The objective of this deployment is to transform the proposed deep learning framework into a practical computer-aided diagnosis (CAD) system capable of assisting clinicians in real-time pulmonary disease diagnosis.

Considering the experimental results presented in the previous sections, MobileNetV3 was selected as the inference engine of the proposed application. Although originally designed as a lightweight CNN architecture for resource-constrained devices, the optimized MobileNetV3 model achieved classification performance comparable to, and in some experiments even superior to, considerably larger networks. In addition to its excellent diagnostic accuracy, MobileNetV3 offers significantly reduced computational complexity, faster inference time, and lower memory consumption, making it particularly suitable for deployment in portable healthcare systems and real-time clinical environments.

The proposed GUI provides a simple and intuitive workflow that enables healthcare professionals to analyze chest medical images without requiring any expertise in artificial intelligence or deep learning. The complete diagnostic process is fully automated, from image loading to the final clinical prediction.

4.4.1 Graphical User Interface Architecture

Figure 8 illustrates the graphical interface of the proposed PulmoDiagAI system. The GUI integrates the trained MobileNetV3 model with the image preprocessing pipeline and the classification engine, providing an end-to-end decision-support tool for pulmonary disease diagnosis.

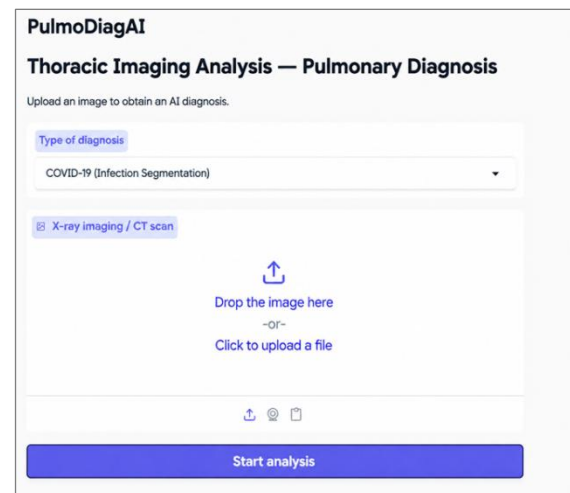


Figure 8: Graphical User Interface (GUI) of the proposed PulmoDiagAI computer-aided diagnosis system

The interface consists of four principal modules:

- Diagnostic mode selection.
- Medical image upload.
- Automatic AI analysis.
- Diagnostic report generation

4.4.1.1 Diagnostic Mode Selection

The first component allows the clinician to select the desired diagnostic protocol according to the examined pathology:

- COVID'19 Detection.
- Lung Cancer Diagnosis.

This selection automatically loads the corresponding optimized MobileNetV3 model and its associated preprocessing pipeline.



Figure 9: Diagnostic mode selection

4.4.1.2 Medical Image Upload

The clinician uploads a chest medical image directly through the graphical interface.

Depending on the selected diagnostic task, the application accepts:

- Chest X-ray images for COVID'19 screening.
- Chest CT scan images for lung cancer diagnosis.

Immediately after loading, the image undergoes the preprocessing operations developed in this work, including resizing, normalization, median filtering, and Contrast Limited Adaptive Histogram Equalization (CLAHE), before being forwarded to the deep learning classifier.

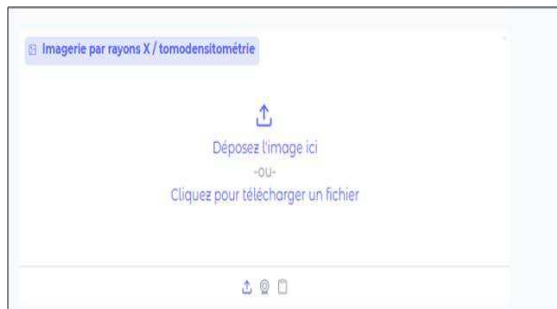


Figure 10: Medical Image Upload

4.4.1.3 Automated AI-Based Diagnosis

After pressing the Analyze button, the complete inference process is executed automatically.

The MobileNetV3 model extracts hierarchical image features through successive convolutional layers and predicts the most probable diagnostic class.

Depending on the selected application, the system produces one of the following outputs.

COVID'19 Screening: the classifier predicts one of three diagnostic categories:

- COVID'19

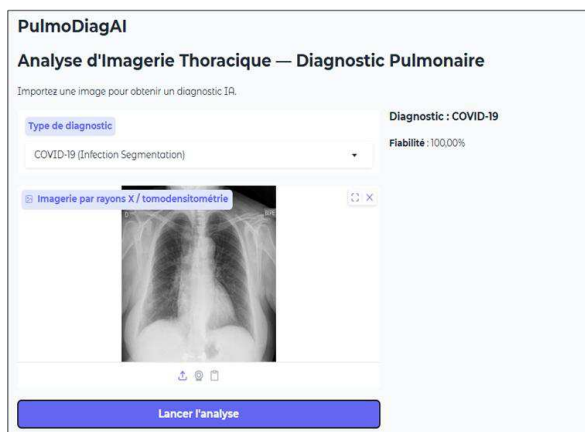


Figure 11: COVID'19 positive prediction

- Non-COVID Pneumonia



Figure 12: Non-COVID pneumonia prediction

- Normal



Figure 13: Normal chest prediction

Lung Cancer Diagnosis: For CT images, the classifier predicts:

- Malignant Tumor



Figure 14: Malignant lung tumor prediction

- Benign Tumor

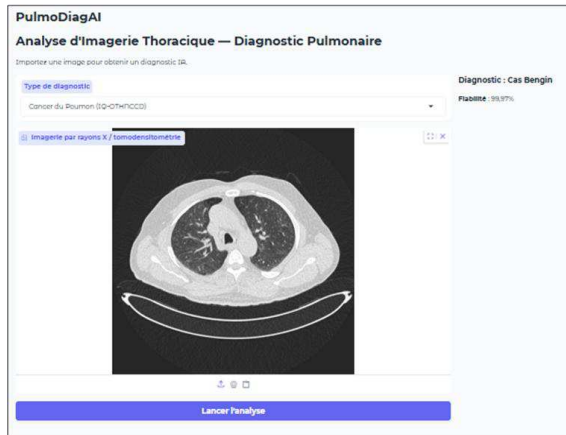


Figure 15: Benign lung tumor prediction

- Normal Lung

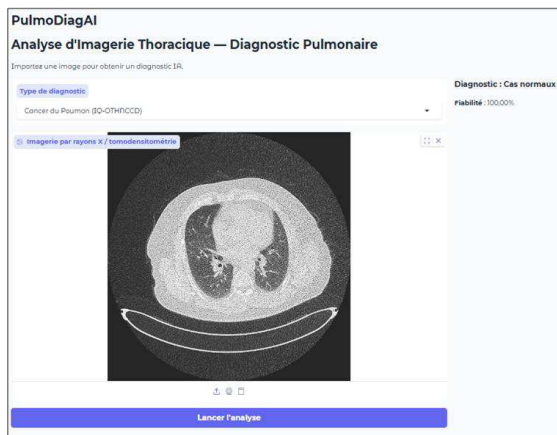


Figure 16: Normal lung prediction

Figures 11-16 present representative examples of the diagnostic results generated by the proposed PulmoDiagAI system.

For the COVID-19 application, the GUI successfully identifies:

- COVID-19 positive patients.
- Non-COVID pulmonary infections.
- Normal chest examinations.

Similarly, for lung cancer diagnosis, the interface accurately distinguishes:

- Malignant tumors.
- Benign tumors.
- Normal pulmonary tissues.

For every analyzed image, the system displays the predicted class together with the corresponding confidence probability, allowing clinicians to rapidly assess the reliability of the automated diagnosis.

Beyond the predicted class, the proposed CAD system provides a confidence score representing the posterior probability associated with the prediction.

Displaying prediction confidence offers several clinical advantages:

- Improves transparency of AI decisions.
- assists clinicians in assessing prediction reliability.
- Facilitates interpretation of uncertain cases.
- Supports clinical decision-making by providing an additional quantitative indicator.

Rather than replacing physician expertise, the proposed system acts as an intelligent second reader, assisting radiologists in interpreting chest medical images more rapidly and consistently.

4.4.1.4 Clinical Perspective

The deployment of the proposed PulmoDiagAI framework demonstrates the feasibility of translating high-performing deep learning models into an operational clinical decision-support tool. By combining high diagnostic accuracy, rapid inference, and an intuitive graphical interface, the developed application enables healthcare professionals to obtain immediate AI-assisted diagnostic recommendations while preserving a simple and user-friendly workflow.

The use of the optimized MobileNetV3 architecture further enhances the practicality of the system by ensuring low computational requirements without compromising diagnostic performance. Consequently, the proposed framework is well suited for integration into hospital workstations, portable diagnostic devices, telemedicine platforms, and future point-of-care screening systems, where fast and reliable pulmonary disease diagnosis is essential.

5. Conclusion

These pulmonary diseases, particularly lung cancer and COVID-19, remain among the leading causes of morbidity and mortality worldwide, highlighting the urgent need for accurate, rapid, and reliable computer-assisted diagnostic solutions. In this work, we proposed PulmoDiagAI, an intelligent deep learning-based computer-aided diagnosis (CAD) framework designed for the automated classification of pulmonary diseases from chest X-ray and CT images [26]. The proposed system integrates state-of-the-art convolutional neural networks with an optimized preprocessing pipeline to provide accurate, real-time diagnostic support for clinical practice.

A comprehensive comparative evaluation was conducted using five transfer learning architectures, namely VGG19, ResNet50, MobileNetV3, EfficientNet-B3, and DenseNet121, on two publicly available benchmark datasets: COVID-QU-Ex for COVID-19 detection and IQ-OTH/NCCD for lung cancer classification. Experimental results demonstrated that all evaluated models achieved excellent diagnostic performance, confirming the effectiveness of deep transfer learning for pulmonary image analysis.

For COVID-19 classification, EfficientNet-B3 achieved the best overall performance, reaching 97% classification accuracy and an outstanding AUC-ROC of 99.43%, while maintaining excellent precision, recall, and F1-score across all diagnostic categories. For lung cancer diagnosis, MobileNetV3 emerged as the most effective architecture, achieving an overall classification accuracy of 99%, whereas both MobileNetV3 and ResNet50 demonstrated perfect discrimination of malignant tumors (AUC = 1.000), eliminating false-negative malignant diagnoses. These findings underline the robustness of the proposed framework and its strong potential for supporting early detection of life-threatening pulmonary diseases.

Beyond quantitative performance evaluation, one of the major contributions of this work is the successful translation of the proposed deep learning framework into a fully operational clinical decision-support application [27-28]. The developed graphical user interface enables clinicians to upload chest radiographs or CT scans, automatically perform image preprocessing and classification, and instantly obtain both the predicted diagnosis and the corresponding confidence score. Owing to its outstanding balance between diagnostic accuracy, computational efficiency, and reduced memory footprint, MobileNetV3 was selected as the deployment model, making the proposed system suitable for real-time hospital environments, portable diagnostic devices, telemedicine platforms, and edge-computing healthcare applications.

The consistency observed between the confusion matrices, classification metrics, and ROC-AUC analyses further confirms the robustness and reliability of the proposed methodology. In particular, the perfect detection of malignant lung tumors across all evaluated architectures represents a clinically significant achievement, as minimizing false-negative cancer diagnoses remains one of the primary objectives of intelligent medical decision-support systems.

Although the obtained results are highly encouraging, several research directions remain open. Future work will investigate the integration of Vision Transformers (ViTs) and hybrid CNN–

Transformer architectures to further improve feature representation and classification performance. The development of a large-scale, clinically validated Algerian pulmonary imaging database is also envisioned to evaluate the proposed framework under real-world healthcare conditions and enhance its generalization capability across diverse patient populations. Additional improvements will include multimodal learning by combining medical images with clinical metadata and laboratory findings, ensemble learning strategies for further performance enhancement, and the extension of the system to detect a broader spectrum of respiratory diseases, including bacterial and viral pneumonias, tuberculosis, interstitial lung diseases, and other thoracic abnormalities.

In conclusion, this study demonstrates that PulmoDiagAI constitutes a robust, accurate, and clinically applicable artificial intelligence framework for automated pulmonary disease diagnosis. By combining high diagnostic performance with an intuitive deployment platform, the proposed system bridges the gap between deep learning research and practical clinical implementation. It has the potential to assist radiologists by providing rapid and reliable second opinions, reducing diagnostic workload, improving early disease detection, and ultimately contributing to more efficient, accessible, and personalized healthcare.

Statements and Declarations

- Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

- Competing Interests

On behalf of all authors, the corresponding author states that there is no conflict of interest.

- Author Contributions

A BOUZID-DAHO contributed to the conception, methodology, implementation, data analysis, and writing of the manuscript. All authors reviewed and approved the final version of the manuscript.

- Ethics Approval

This study does not involve human participants or animals. Therefore, ethical approval was not required.

- Consent to Participate

Not applicable.

- Data Availability

The datasets used in this study are publicly available:

- 1- COVID-QU-Ex:
<https://www.kaggle.com/datasets/anas-mohammedtahir/covidqu>
- 2- IQ-OTH/NCCD Lung Cancer:
<https://www.kaggle.com/hamdallak/th-e-iqothnccd-lung-cancer-dataset>

- Consent to Publish

Not applicable.

References

- [1] Y. Gu, H. Bai, M. Chen et al., "WSDC-ViT: A Novel Transformer Network for Pneumonia Image Classification Based on Windows Scalable Attention and Dynamic Rectified Linear Unit Convolutional Modules," *Scientific Reports*, vol. 15, N°. 27868, 2025, doi: 10.1038/s41598-025-12117-0
- [2] A. M. Rifai, A. A. Abdillah, M. R. Ramadhan et al., "Analysis for Diagnosis of Pneumonia Symptoms Using Chest X-ray Based on MobileNetV2 Models with Image Enhancement Using White Balance and Contrast Limited Adaptive Histogram Equalization (CLAHE)," *Biomedical Signal Processing and Control*, vol. 90, N°. 105857, 2024, doi: 10.1016/j.bspc.2023.105857
- [3] Bouzid-Daho, A., et al., "Textural analysis of bio-images for aid in the detection of abnormal blood cells," *Int. J. Biomed. Eng. Technol.* vol. 25, N°. 1, pp. 1–13, 2017. doi: 10.1504/IJBET.2017.10009312
- [4] M. Khalil et al., "The State of the Art for Artificial Intelligence in Lung Digital Pathology," *The Journal of Pathology*, vol. 257, N°. 4, pp. 413-429, 2022. doi: 10.1002/path.5895
- [5] Singh et al., "Multi-modal Deep Learning for Respiratory Disease Classification," *International Journal of Information Technology*, vol. 14, N°. 4, pp. 2049-2056, 2022. doi: 10.1007/s41870-022-00967-z
- [6] R. Javed, T. Abbas, A. H. Khan, A. Daud, A. Bukhari, and R. Alharbey, "Deep learning for lungs cancer detection: A review," *Artificial Intelligence Review*, vol. 57, N°. 197, 2024. doi: 10.1007/s10462-024-10807-1.
- [7] Bouzid-Daho, A., Siarry, P.: Detection, identification and tracking of anomaly in medical images. In: *Proc. IEEE ADACIS, Marrakech* 2023. doi: 10.1109/ADACIS59374.2023.10424339
- [8] V. Kumar, C. Prabha, P. Sharma, N. Mittal, S. S. Askar, and M. Abouhawwash, "Unified deep learning models for enhanced lung cancer prediction with ResNet-50-101 and EfficientNet-B3 using DICOM images," *BMC Medical Imaging*, vol. 24, N°. 63, 2024. doi: 10.1186/s12880-024-01241-4
- [9] W. El-Tarhouni et al., "Multi-classification of chest diseases using deep learning," *Computers in Biology and Medicine*, vol. 138, N°. 104348, 2021. doi: 10.1016/j.compbimed.2021.104348.
- [10] T.-W. Wang et al., "Systematic review and meta-analysis of deep learning applications in computed tomography lung cancer segmentation," *Radiotherapy and Oncology*, vol. 197, N°. 110344, 2024. doi: 10.1016/j.radonc.2024.110344.
- [11] A. Hiranman, S. Viriri, and M. Gwetu, "Lung tumor segmentation: A review of the state of the art," *Frontiers in Computer Science*, vol. 6, 2024. doi: 10.3389/fcomp.2024.1423693.
- [12] T. Du et al., "A non-enhanced CT-based deep learning diagnostic system for COVID-19 infection at high risk among lung cancer patients," *Frontiers in Medicine*, vol. 11, 2024. doi: 10.3389/fmed.2024.1444708.
- [13] M. A. Talukder, M. A. Layek, M. Kazi, M. A. Uddin, and S. Aryal, "Empowering COVID-19 detection: Optimizing performance through fine-tuned EfficientNet deep learning architecture," *Computers in Biology and Medicine*, vol. 168, N°. 107789, 2024. doi: 10.1016/j.compbimed.2023.107789.
- [14] C. Zhang et al., "Enhancing lung cancer diagnosis with data fusion and mobile edge computing using DenseNet and CNN," *Journal of Cloud Computing*, vol. 13, N°. 91, 2024. doi: 10.1186/s13677-024-00597-w.
- [15] A. A. Nafea, M. S. Ibrahim, M. M. Shwaysh, K. A. Abdul-Kadhim, H. R. Almamoori, and M. M. Al-Ani, "A Deep Learning Algorithm for Lung Cancer Detection Using EfficientNet-B3," *Wasit Journal of Computer and Mathematics Science*, vol. 2, N°. 4, pp. 68–76, 2023. doi: 10.31185/wjcms.209.
- [16] A. Taher et al., "IQ-OTH/NCCD Lung Cancer Dataset," *University of Thi-Qar, Iraq*

- [17] M. Rahimzadeh and A. Attar, "COVID-QU-Ex: A Three-Class Chest X-ray Dataset for COVID-19 Detection."
- [18] Mustapha, Benzorgat, Yatong Zhou, and Benzorgat Nawel. "Hybrid CNN with SGE Attention and Sequencer2D-L for Efficient Classification of Chest X-Rays and CT-Scans in COVID-19 and Pneumonia Detection." 2025 International Conference on Recent Advances in Mathematics and Informatics (ICRAMI). IEEE, 2025.
- [19] Hossain, Md Yearat, Md Mahbub Hasan Rakib, and Rashedur M. Rahman. "LitCovNet: Attention-Based Lightweight Convolutional Network for Covid and Lung Disease Classification from Chest X-ray Images." SN Computer Science 7.1 (2025): 57.
- [20] Fu, Xiaoyang, et al. "Explainable hybrid transformer for multi-classification of lung disease using chest X-rays." Scientific Reports 15.1 (2025): 6650.
- [21] Schraut, Jaiden Xuan, et al. "A multi-output network with U-net enhanced class activation map and robust classification performance for medical imaging analysis." Discover Artificial Intelligence 3.1 (2023): 1.
- [22] Al-Yasriy, Hamdalla F., et al. "Diagnosis of lung cancer based on CT scans using CNN." IOP conference series: materials science and engineering. Vol. 928. No. 2. IOP Publishing, 2020.
- [23] Kareem, Hamdalla F., et al. "Evaluation of SVM performance in the detection of lung cancer in marked CT scan dataset." Indonesian Journal of Electrical Engineering and Computer Science 21.3 (2021): 1731.
- [24] Al-Huseiny, Muayed S., and Ahmed Sattar Sajit. "Transfer learning with GoogLeNet for detection of lung cancer." Indonesian Journal of Electrical Engineering and computer science 22.2 (2021): 1078-1086.
- [25] Ghosh, Susmita, and Abhiroop Chatterjee. "Self-attention driven transfer learnt deep neural network for improved lung cancer detection from radiographic images." Heliyon 11.16 (2025).
- [26] Bouzid-Daho, A., Sofi, N., Siarry, P.: Extracted Haralick's texture features for abnormal blood cells. In: Advances in Multidisciplinary Medical Technologies, Springer, Cham, pp. 115–126 (2020). https://doi.org/10.1007/978-3-030-57552-6_9
- [27] Bouzid-Daho, A., Boughazi, M., Petit, E.: Detection of abnormal blood cells by segmentation and classification. Int. J. Med. Eng. Inform. 11(1), 57–70 (2019). <https://doi.org/10.1504/IJMEI.2019.10017842>
- [28] Bouzid-Daho, A., Boughazi, M., Hafs, T.: Segmentation by an improved mathematical morphology approach. Appl. Math. Inf. Sci. 14(5), 841–846 (2020). <https://doi.org/10.18576/amis/14050>.