Online-Journals.org

iJOE International Journal of Online and Biomedical Engineering

iJOE | eISSN: 2626-8493 | Vol. 22 No. 1 (2026) | OPEN ACCESS

https://doi.org/10.3991/ijoe.v22i01.58445

PAPER

DI-EffNet: A Dual-Attention Network for Binary ILD Classification from Imbalanced CT Data

Norhène Gargouri[1](✉), Nesrine Charfi[2], Alima Damak Masmoudi[3], Wiem Feki[4], Chifa Damak[5]

[1]Digital Research Center of Sfax, Sfax, Tunisia

[2]Higher Institute of Technological Studies (ISET) of Kef, Boulifa – University Campus, Le Kef, Tunisia

[3]Lab.CEM, Electrical Department at the National Engineering School, Sfax, Tunisia

[4]Department of Radiology, Hédi CHaker Hospital, Sfax, Tunisia

[5]Internal Medicine Clinic, ALCHIFA Building, Sfax, Tunisia

norhene.gargouri@crns.tn

ABSTRACT

Classifying computed tomography (CT) images for interstitial lung diseases (ILDs) is a significant research challenge. Recent studies have explored the effectiveness of pre-trained models to improve performance and accuracy. However, training on imbalanced datasets remains a major hurdle, often resulting in biased predictions. This study focuses on multi-label classification of thoracic lung diseases using CT images, specifically addressing class imbalance. To mitigate this issue, data augmentation methods were employed to create synthetic samples, and images from various sources were integrated, improving the models' generalization capabilities. In this paper, we introduce DI-EffNet, a novel dual-input neural network architecture that combines both local and global attention mechanisms to enhance the binary classification of ILDs on CT scans. Comparative experiments show that DI-EffNet significantly outperforms established deep learning models, achieving an accuracy of 99.75%, compared to 89.39% for ResNet-50, 93.9% for VGG-19, and 93.29% for EfficientNet B0. These results demonstrate that DI-EffNet provides a robust and effective solution for ILD detection, with strong potential to support clinical diagnosis.

KEYWORDS

interstitial lung disease (ILD), data augmentation, computed tomography (CT) imaging, imbalanced datasets, attention mechanisms

1 INTRODUCTION

Interstitial lung diseases (ILDs) represent a diverse group of pulmonary disorders characterized by varying degrees of inflammation and fibrosis, leading to significant morbidity and mortality worldwide [1]. Early and accurate diagnosis is crucial for effective management and improved patient outcomes. Computed tomography (CT) imaging plays a pivotal role in the detection and classification of ILDs, as it provides detailed visualization of lung structures. However, the manual interpretation of CT scans is time-consuming and subject to inter-observer variability, highlighting the need for automated and reliable diagnostic tools [2]. Recent advances in

Gargouri, N., Charfi, N., Masmoudi, A. D., Feki, W., Damak, C. (2026). DI-EffNet: A Dual-Attention Network for Binary ILD Classification from Imbalanced CT Data. *International Journal of Online and Biomedical Engineering (iJOE)*, 22(1), pp. 57–77. https://doi.org/10.3991/ijoe.v22i01.58445

Article submitted 2025-08-28. Revision uploaded 2025-10-30. Final acceptance 2025-10-30.

deep learning have shown promising results in medical image analysis, particularly through the use of pre-trained convolutional neural networks (CNNs). Despite these advancements, the classification of ILD from CT images remains challenging due to the inherent imbalance in available datasets. This imbalance can bias model training, leading to suboptimal performance, especially for underrepresented classes. To address these challenges, we propose a novel deep learning architecture, the DI-EffNet, designed specifically for binary ILD classification from imbalanced CT datasets. DI-EffNet integrates both local and global attention mechanisms to enhance feature extraction and improve classification accuracy. Additionally, we employ data augmentation techniques and incorporate images from diverse sources to mitigate the effects of class imbalance and enhance the model's generalization capabilities.

In this study, we evaluate the performance of DI-EffNet against established models such as ResNet-50, VGG-19, and EfficientNet B0. Our results demonstrate that DI-EffNet achieves superior performance, particularly when combined with strategies to balance the dataset. This work aims to contribute to the development of robust and accurate automated tools for ILD classification, ultimately supporting clinicians in the diagnostic process. The main contribution of this work lies in the development of DI-EffNet, a novel dual-input deep learning framework that integrates both segmentation and classification for ILD detection from CT images. By fusing features from raw images and their segmentation masks, DI-EffNet enhances diagnostic accuracy and robustness.

2 RELATED WORK

The landscape of deep learning architectures has undergone a series of transformative evolutions, one of the most notable being the introduction of DenseNet in 2017. Unlike traditional feedforward networks where each layer connects linearly to the next, DenseNet adopt a radically different philosophy: every layer is directly connected to all subsequent layers. In parallel, another architectural innovation was first introduced within the InceptionNet framework [3]. The Inception paradigm sought not merely to deepen the network, but to widen it [4]. By orchestrating parallel convolutions with varying receptive fields and combining them through dimensionality-reducing 1×1 convolutions, inception blocks enabled the network to extract multiscale features with minimal computational time [5]. The squeeze-and-excitation (SE) network introduces attention mechanisms that help the model highlight important channel features and reduce the impact of less useful ones. By learning how different channels relate to each other, SE blocks help the network concentrate on the most relevant information [6].

The convolutional block attention module (CBAM) [7] builds on adding attention not only to channels but also to spatial locations. CBAM first finds which channels and then which spatial areas are important, and then adjusts the feature maps to highlight these key parts. This helps the network focus better on what and where important information is found. K. Wang et al. [8] added the SE attention module to a DenseNet model to help in the detection of pneumonia, showing how useful these methods can be in medical diagnosis. Similarly, Chakraborty et al. [9] used CBAM attention modules to improve breast density classification from mammograms, demonstrating the flexibility of attention techniques in medical imaging. To sum up, deep learning models have changed a lot: from basic layers to DenseNet with many connections, from simple convolutions to Inception modules that look at different scales, and now to attention methods such as SE and CBAM that help the

network focus better. This shows that networks are not just getting bigger, but also becoming smarter.

In recent years, many CNN models have been developed to analyze chest CT scans, especially for lung-related problems. For example, Wang et al. [10] used chest CT images to successfully detect COVID-19. So, Vaidyanathan et al. [11] improved on this by using a 3D inception model that looks at 48 CT slices at once, advancing the detection of COVID-19 in three dimensions. Abdar et al. [12] used the popular VGG-16 model with transfer learning, which means they used knowledge from other datasets to help improve diagnosis. At the same time, Bakshi et al. [13], in 2022, created a detailed process using DeepLab v3 to accurately segment lung lesions, which helped in improving later classification steps. In the same context, new 3D dual-scale CNN was created to classify COVID-19 and regular pneumonia [14]. Ciompi et al. [15] combined several classifiers, including the OverFeat CNN, to classify lung nodules with strong results. Many CNN methods have also been used to classify different tissue types in ILDs. For example, Anthimopoulos et al. [16] designed a deep CNN with five convolution layers and special activations, reaching 85.5% accuracy on small image patches. Later, Christodoulidis et al. [17] improved this by using global pooling and transfer learning from other datasets, increasing accuracy by 2%. Wang et al. [18] created a more complex CNN that used multi-scale features and special filters, achieving 90% accuracy on the same patch size. In 2019, Kim et al. [19] developed a CNN with four convolution layers and two dense layers, reaching 95.12% accuracy on small patches. However, all these models work on small parts of the image, which limits their use in real clinical settings because they miss the full context of lung scans.

To resolve this problem, Gao et al. [20] used whole lung slices for ILD classification, with a CNN containing five convolution and four dense layers. This matched clinical needs better but had a lower accuracy of about 73%, limiting its reliability for clinical use. Shin et al. [21] tried to improve these results by using GoogleNet with transfer learning on full lung slices from CT scans; however, this approach requires large amounts of labeled data and significant computational resources. Anthimopoulos et al. [22] used a CNN with dilated convolution layers to better segment ILD patterns, but the complexity of the model can lead to longer training times and potential overfitting. Oh et al., in 2024, [23] have started using fully automated systems that include steps like removing noise, sampling pixels, extracting texture features, classifying patterns, and measuring disease areas at the pixel level. While promising, these pipelines can be sensitive to variations in image quality and may require extensive validation before clinical adoption.

Recent advancements in deep learning have significantly improved binary classification for ILD detection. Kumarganesh et al. [24] introduces an automated method for early identification of ILD using CT images by combining radiomic feature analysis with deep learning models. The approach optimizes radiomic features and employs an attention-based CNN, with both outputs merged for final classification. This method achieved a 92.3% accuracy for binary classification. In the same context and for the same dataset, [25] proposed a novel approach based on the Split Branch Identification Design SB-ID Net, a new deep learning architecture that combines parallel processing, multi-scale features, and dense connectivity to improve feature extraction and robustness. Multiple attention modules are used to highlight important features, enhancing the model's effectiveness. Ablation studies demonstrate the value of each component, while comparisons show the model outperforms current state-ofthe-art methods. Kumarganesh et al.'s method achieves strong binary classification results but depends heavily on radiomic feature extraction, which can vary

with differences in image segmentation and preprocessing steps. The combination of handcrafted and deep features might also reduce its adaptability when applied to different datasets or imaging techniques. Likewise, the SB-ID Net introduced by Bakshi et al. offers powerful performance through its complex design with multiple attention mechanisms and dense connections, but this complexity leads to higher computational demands and longer training times.

3 MATERIALS AND METHODS

This paper centers on the classification of ILD based on a new model. The key components of the workflow (see Figure 1) presented for accomplishing this objective are outlined in the following sections.

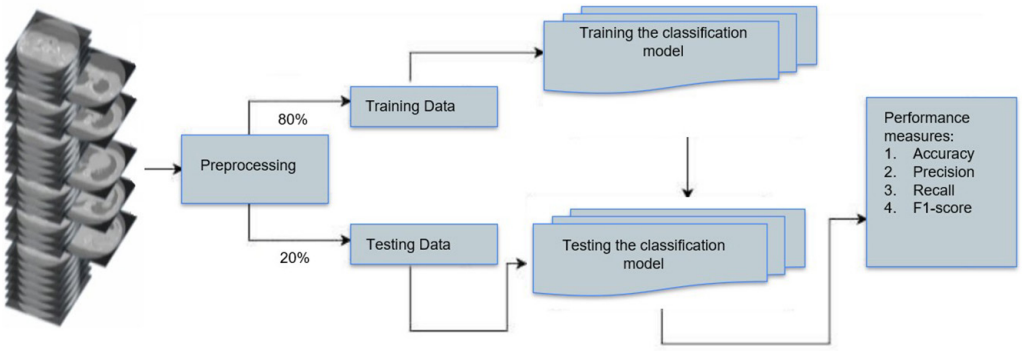

Fig. 1. Experimental framework of the classification model used in this study

3.1 Description of the database

The database used in this paper is MedGIFT, which specializes in ILDs [26]. It was developed as part of the TALISMAN project at the University Hospitals of Geneva. It contains 3D-annotated CT image series representing pathological lung regions, accompanied by clinical parameters linked to ILD diagnoses confirmed through histopathological examinations. Collected over a period of 38 months, it includes data from 128 patients affected by one of the 13 recognized histological forms of ILD (see Figure 2). The dataset comprises a total of 3,024 CT images, each paired with its corresponding segmentation mask, amounting to 3,024 masks in total. Using this database presented both a scientific opportunity and a technical challenge, requiring rigorous efforts in understanding, converting, and structuring the data for the success of the proposed model. Approximately 25% of the cases exhibited multiple disease patterns; these particular slices were excluded from our dataset to maintain clarity and consistency in the analysis.

3.2 Representation of the dataset problem

To tailor the MedGIFT database for the specific objectives of this study, the dataset was partitioned into two task-oriented subsets: one for lung region segmentation and another for image classification. Each subset required unique preprocessing steps and design considerations to meet the respective modeling goals. The following

sub-subsections detail the data preparation, organization, and utilization strategies employed for both the segmentation and classification tasks, highlighting their distinct structures and roles in training the proposed deep learning models.

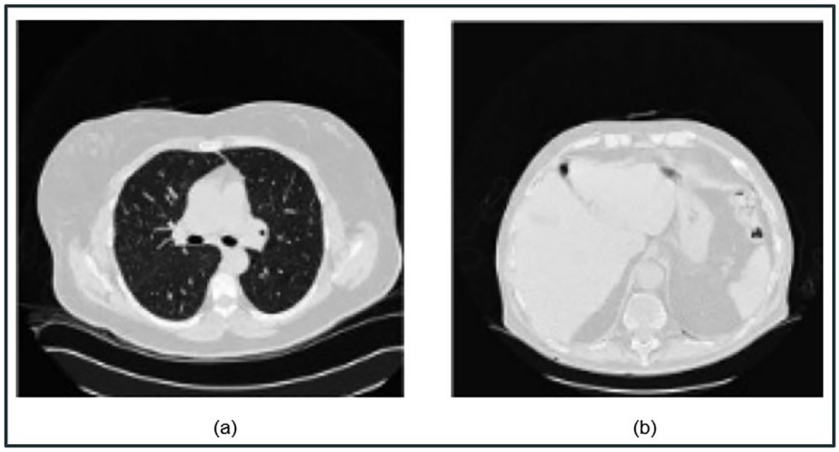

Fig. 2. CT slices example: (a) image with a visible part of the lung, (b) image without any lung structure

3.3 Pre-processing

Preparing the dataset for segmentation. For the segmentation task, we utilized a specialized subset from the MedGIFT database, which comprised 2,377 DICOM images of axial thoracic CT slices along with their corresponding segmentation masks. Each mask provided a binary ground truth for the anatomical regions of interest. To facilitate efficient access and training, the images and masks were each stored separately in dedicated folders in JPG format. The dataset was then divided into two parts: 80% of the images were used to train the U-Net model, while the remaining 20% were reserved for validation.

Preparing the dataset for classification. To classify thoracic CT scans as healthy or pathological, the same MedGIFT dataset from the segmentation task was utilized. Since only 195 healthy samples were available, an equal number of pathological images were randomly selected to create a balanced dataset of 390 scans. Balancing the classes in this way is crucial to prevent bias and support more effective training. To further enhance the dataset's variety and robustness of the classification model, we applied multiple data augmentation techniques. These augmentations simulate different real-world conditions and improve the model's generalization by artificially increasing dataset diversity. The transformations included:

- Rotations (30°): To simulate various image orientations during scanning
- Horizontal translation: To mimic shifts in patient positioning
- Zooming: To represent scale changes in anatomical structures
- Brightness adjustments: To reflect different image contrast conditions
- Horizontal flipping: To create mirror-image variations

Such augmentations (see Figure 3) diversify the data without the need for additional image collection, boosting the model's adaptability and reducing overfitting risk. After augmentation, the data was split again into 80% for training and 20% for validation, ensuring a wide range of training examples and consistent evaluation on unseen data.

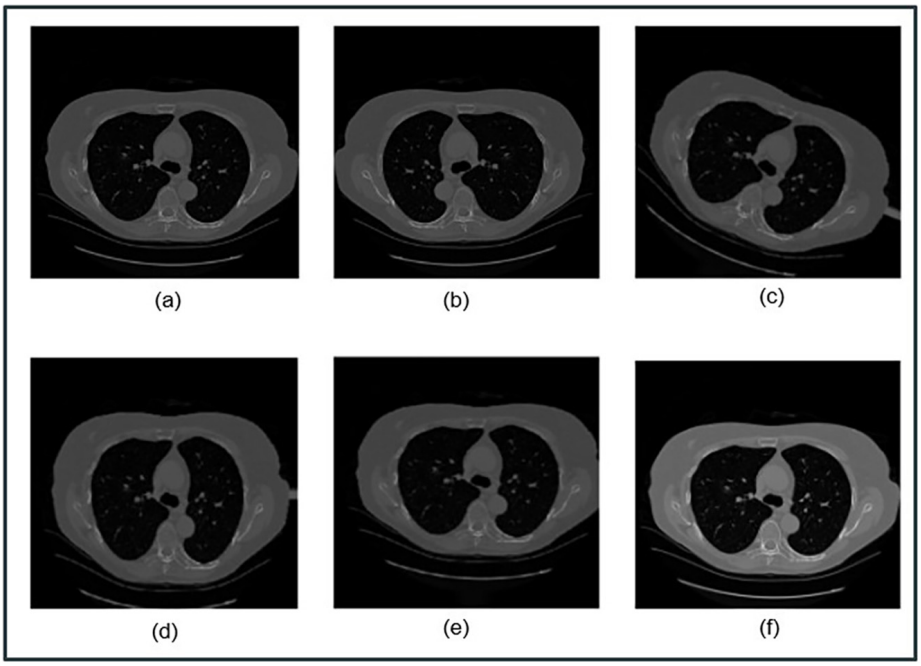

Fig. 3. Examples of data augmentation applied to a lung image from the MedGIFT database: (a) Original image, (b) Horizontal flips, (c) 30° Rotation, (d) Zoom operations, (e) Horizontal translation, (f) Brightness adjustments

4 IMAGE NORMALIZATION

To improve consistency across the dataset and enhance training stability, all pixel intensity values were normalized to a scale between 0 and 1. This step mitigates variations caused by different imaging settings and acquisition protocols, thereby supporting more efficient and reliable model development.

4.1 Proposed model: DI-EffNet

In this study, we introduce DI-EffNet, an innovative dual-input framework designed to classify pulmonary CT scans as either healthy or unhealthy. Our method brings together segmentation and classification components, combining thorough image pre-processing, precise lung segmentation using a U-Net architecture, and a dual-path classification strategy powered by EfficientNet models. This integrated approach aims to enhance the accuracy and reliability of distinguishing between normal and diseased lung images within an unbalanced CT dataset.

Segmentation architecture at the outset, all CT images are converted to grayscale and uniformly resized to 256×256 pixels. To isolate lung regions from the raw scans, we used a U-Net segmentation model, trained using paired CT images and their corresponding masks from the MedGift dataset.

In this study, we used a U-Net model [27] to segment lungs from chest X-ray images. Lung segmentation is an important step in computer-aided diagnosis systems because it helps focus on the lung area and improves the accuracy of subsequent analyses. U-Net is a popular deep learning model for medical image segmentation, known for its encoder-decoder design and skip connections that preserve important spatial information.

U-Net has four main parts: an encoder that extracts features at different levels, a bottleneck that captures the most abstract information, a decoder that reconstructs the spatial resolution, and an output layer that produces a mask showing the lung area.

The encoder has four blocks, each consisting of two convolution layers with 3×3 filters and ReLU activation, followed by max-pooling to reduce the spatial size while increasing feature depth. The number of filters increases from 64 to 512 across these blocks. The bottleneck includes two convolution layers with 1024 filters to extract high-level features.

The decoder mirrors the encoder structure but in reverse. It upsamples the feature maps and combines them with the corresponding encoder features through skip connections, thus preserving fine details. Each decoder block has two convolution layers with filters decreasing from 512 to 64. Finally, a 1×1 convolution with a sigmoid activation outputs a pixel-wise probability map for lung tissue classification.

We trained the model on 256×256 grayscale CXR images using the Adam optimizer with a learning rate of 0.001. The loss function was binary cross-entropy, and the training was conducted for 60 epochs with a batch size of 4 (see Figure 4).

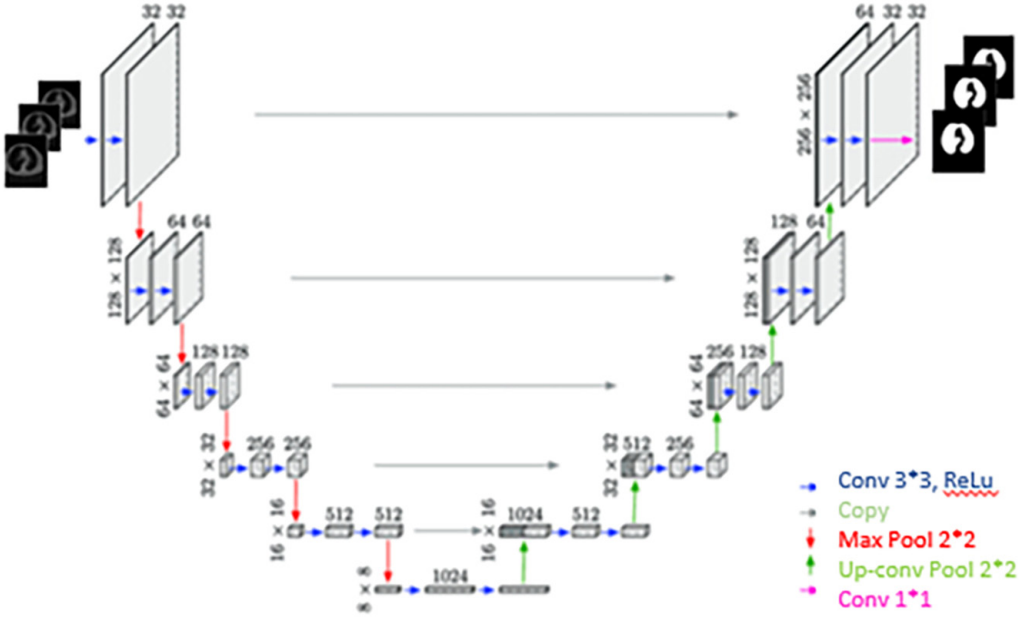

Fig. 4. Diagram illustrating the segmentation model architecture

The segmentation model was configured with several key hyper-parameters that influence both the learning process and generalization capability (refer to Table 1).

Table 1. Segmentation model hyper-parameters

Hyperparameter	Value
Input size	256×256×1
Learning rate	0.001
Loss function	Binary Crossentropy
Optimizer	Adam
Number of epochs	60
Batch size	4

After the training phase, the U-Net model generates binary masks that accurately delineate the lung regions in both healthy and diseased CT images. These segmented regions are then extracted and stored separately to serve as focused inputs for the classification stage. Dual paths are used in the remainder of this paper, we focus exclusively on the lung regions extracted from CT scans. To exploit the clearer and more localized lung textures, we use a deeper version of the EfficientNet B0 model (see Figure 5). We apply similar data augmentation techniques and follow a two phase training strategy, including fine-tuning the final layers. Additionally, regularization and dropout techniques are used to reduce overfitting and enhance the model's generalization. EfficientNet B0 is a convolutional neural network (CNN) model introduced by Google AI in 2019 as part of the EfficientNet family [28]. It was designed to provide high classification accuracy while maintaining computational efficiency, significantly reducing memory usage and inference time without sacrificing performance. The model integrates traditional convolutional layers with mobile inverted bottleneck convolution (MBConv) blocks, originally developed in MobileNetV2. These blocks are called "inverted bottlenecks" because they first expand the number of channels before compressing them again, opposite to traditional bottlenecks. EfficientNet B0 also incorporates squeeze-and excitation (SE) modules, which function as attention mechanisms by dynamically reweighting channel-wise features [29]. This allows the network to emphasize important features and improves overall performance. The architecture comprises several stages optimized for computational efficiency. The applied parameters to an input of size 224×224×3 are Kernel sizes are indicated as k3×3 or k5×5, and S1 and S2 refer to stride values of 1 (same spatial size) and 2 (halved spatial size), respectively (refer to Table 2). The final fully connected (FC) layer follows global pooling and outputs the classification prediction.

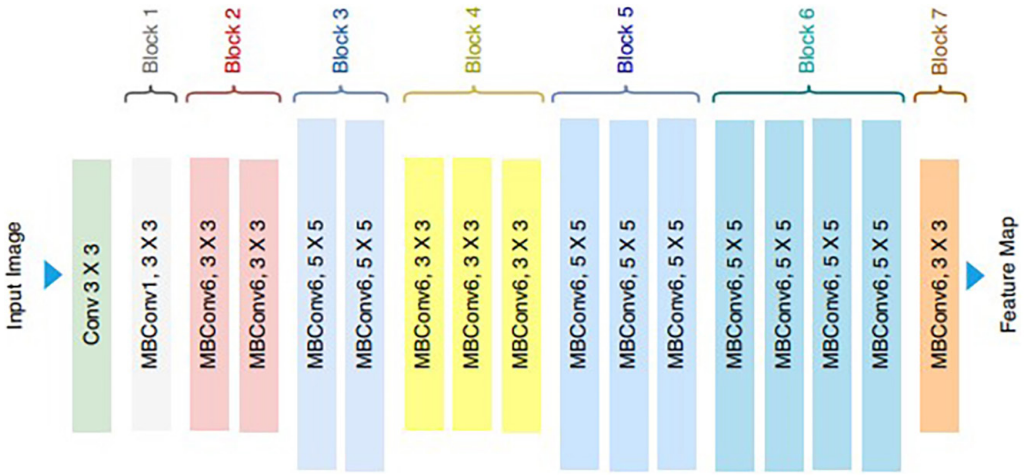

Fig. 5. EfficientNet B0 architecture based on MBConv as core building blocks

EfficientNet B0 offers several advantages for pulmonary texture analysis in medical imaging. It is lightweight and efficient, achieving strong accuracy with relatively few parameters, making it suitable for resource-constrained systems and datasets such as those used in this study.

In this paper, we propose two classification pipelines:

1. Path A without segmentation: We directly classify the original CT images using an EfficientNet B0 model. The dataset is divided into training and validation sets with an 80/20 split. Various data augmentation techniques are applied to

improve generalization. Training is conducted in two steps: initially freezing the base layers to train the classifier head, followed by fine-tuning deeper layers.

2. Path B with segmentation: In this path, we use only the segmented lung regions extracted from CT scans. A deeper EfficientNet B0 model is employed to take advantage of the focused lung features. The same augmentation and training strategies are used, along with dropout and regularization to minimize overfitting.

To make our dual-path architecture even more effective for diagnosing lung disease, we introduce a fusion-based classification strategy. Instead of analyzing the original CT image and its segmentation mask separately, this method brings together information from both sources. By combining these perspectives, the model can better highlight the areas of the lungs that matter most for diagnosis, while still considering the overall structure of the chest.

In practice, the process works as follows: both the CT image and its segmentation mask are processed using the same deep learning backbone (EfficientNet B0). This ensures that both types of input are treated consistently and that the network learns similar kinds of features from each. After features are extracted from each path, they are merged; specifically, the outputs from both branches are concatenated. This fusion happens before the final classification layers. By integrating information from both the raw image and the mask, the network is encouraged to pay special attention to abnormal lung regions identified by segmentation, while also keeping the broader anatomical context in mind. This approach helps the model develop a more nuanced understanding of the images, which can lead to more accurate and reliable classification results. This fusion strategy not only maintains a consistent architecture but also enriches the information available to the model, making it better equipped to distinguish between healthy and diseased lungs. The end result is a smarter, more focused diagnostic tool that leverages the strengths of both global and local image features.

To train our dual-branch fusion model, which brings together both the original CT images and their segmentation masks, we selected a set of hyperparameters designed to promote strong learning and reliable performance. These choices were informed by initial experiments and established practices in deep learning for medical imaging. The goal was to find a balance between effective training and the ability to generalize well to new data.

Table 2. Hyperparameters used for training the model in the proposed fusion-based approach

Hyperparameters	Value
Input image size	128×128×3
Batch size	16
Number of epochs	15
Optimizer	Adam
Learning rate	0.0001 (1e–4)
Loss function	Binary Crossentropy
Dropout rate	0.5
Number of neurons in Dense layer	128
Final activation function	Sigmoid

5 RESULTS AND DISCUSSION

All experiments for this project were performed on a laptop designed for the execution of computational tasks (refer to Table 3). The chosen hardware made it possible to process data efficiently, run image processing routines smoothly, and train deep learning models without major slowdowns. This setup also helped ensure that the results could be reliably reproduced.

Table 3. Specifications of the used computing environment

Component	Specification
Laptop Brand	DESKTOP-TPKNPLJ
Processor	12th Gen Intel(R) Core(TM) i7-12650H 2.30 GHz
Graphics Card	NVIDIA GeForce RTX
Memory	16,0 Go (15,7 Go utilisable)
Hard Drive	512 GB SSD
Operating System	Windows 10 Pro

5.1 Evaluation metrics

We used various performance metrics to validate the obtained results. These metrics are based on the following concepts:

- True Positive (TP): Number of elements correctly identified as positive
- False Positive (FP): Number of elements incorrectly identified as positive
- False Negative (FN): Number of elements incorrectly identified as negative
- True Negative (TN): Number of elements correctly identified as negative
- Accuracy (Acc): Accuracy measures the proportion of correctly classified pixels both positive and negative relative to the total number of pixels

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

- Dice Coefficient: The Dice coefficient measures the similarity between the predicted segmentation and the ground truth mask. It is particularly sensitive to under-segmentation, heavily penalizing relevant areas that are not detected.

$$Dice = \frac{2\left|\text{Predictions} \cap \text{Ground Truth}\right|}{\left|\text{Predictions}\right| + \left|\text{Ground Truth}\right|} = \frac{2TP}{2TP + FP + FN} \tag{2}$$

- Jaccard Index (Intersection Over Union): This index measures the ratio between the intersection and the union of the predicted pixel set and the ground truth mask. It is used to assess the quality of overlap between prediction and ground truth.

$$IOU = \frac{\left|\text{Predictions} \cap \text{Ground Truth}\right|}{\left|\text{Predictions} \cup \text{Ground Truth}\right|} = \frac{TP}{TP + FP + FN} \tag{3}$$

- Precision: Measures the ratio between the number of true positives and the sum of true positives and false positives. It indicates how many of the elements predicted as positive are actually correct.

$$Precision = \frac{TP}{TP + FP} \tag{4}$$

- Recall (Sensitivity): Measures the ratio between the number of true positives and the sum of true positives and false negatives.

$$Recall = \frac{TP}{TP + FN} \tag{5}$$

- F1 Score: Harmonic mean between precision and recall.

$$F1\,Score = 2.\frac{Precision.Recall}{Precision + Recall} \tag{6}$$

- Receiver operating characteristic (ROC) curve: A graph used to evaluate the performance of a binary classifier. It plots "sensitivity" (true positive rate) against "1 – specificity" (false positive rate) for various model decision thresholds.
  The ROC curve axes are describes as follows:
  Vertical Axis (Y): True Positive Rate (TPR), also called Sensitivity:

$$TPR = \frac{TP}{TP + FN} \tag{7}$$

  Horizontal Axis (X): False Positive Rate (FPR):

$$FPR = \frac{FP}{FP + TN} \tag{8}$$

  A ROC curve near the upper left corner indicates excellent model performance, while a curve close to the 45° diagonal suggests random prediction. A curve below this diagonal implies worse than random performance, and flipping the predicted classes may improve results.
- Area under the ROC curve (AUC): AUC ranges from 0 to 1, where 1 represents a perfect model, 0.5 indicates a random model, and values below 0.5 reflect ineffective performance.
- Confusion Matrix: A 2 * 2 matrix (refer to Table 4) that compares the model predictions to the ground truth.

Table 4. Binary confusion matrix

Prediction\Reality	Actual Values	
	Positive (1)	Negative (0)
Positive (1)	True Positive (TP)	False Positive (FP)
Negative (0)	False Negative (FN)	True Negative (TN)

5.2 Segmentation results

The proposed model was trained for 60 epochs. Visual analysis of the results clearly shows the progression of learning as well as the stability of the model. The training accuracy curve (see Figure 6) starts at 88.62% and steadily improves to reach 99.39%, illustrating a consistent enhancement of the model's performance over iterations. Encouragingly, the validation accuracy follows a similar trajectory, closely aligned with the training curve. This consistency indicates that the model does not suffer from overfitting and generalizes well to unseen data. The analysis of the loss function further supports this observation. The associated curve decreases progressively and smoothly, indicating that the model effectively reduces the error between predictions and ground truth values. A regular decrease, without abrupt fluctuations or stagnation, generally reflects well-controlled learning and appropriate hyperparameter choices. Segmentation specific metrics such as the Dice coefficient and IoU also show favorable trends. The Dice score gradually increases throughout the epochs, demonstrating the model's improving ability to accurately detect regions of interest, ensuring good overlap between predicted masks and ground truth ones. Similarly, the IoU values continuously improve, indicating increasingly precise spatial interpretation of segmented objects.

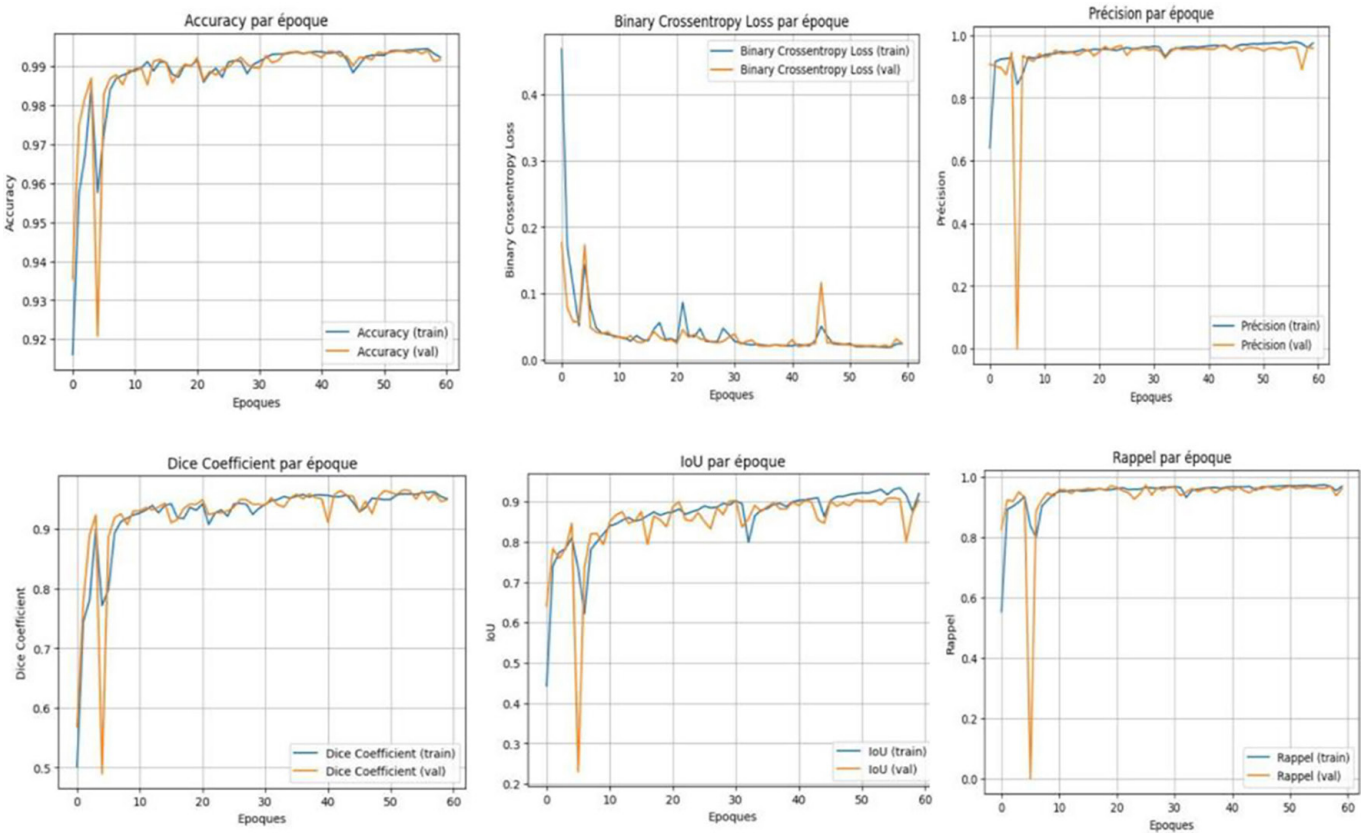

Fig. 6. Evolution of the model performance during training

The resulting curves reveal overall consistency among the different performance metrics. The simultaneous increase in accuracy, decrease in loss, and improvement in Dice and IoU scores reflect effective and balanced learning. No signs of divergence or overfitting are observed, reflecting a well-designed architecture and a rigorously

controlled training process. This harmonious convergence of indicators constitutes a strong empirical validation of the reliability of the U-Net model.

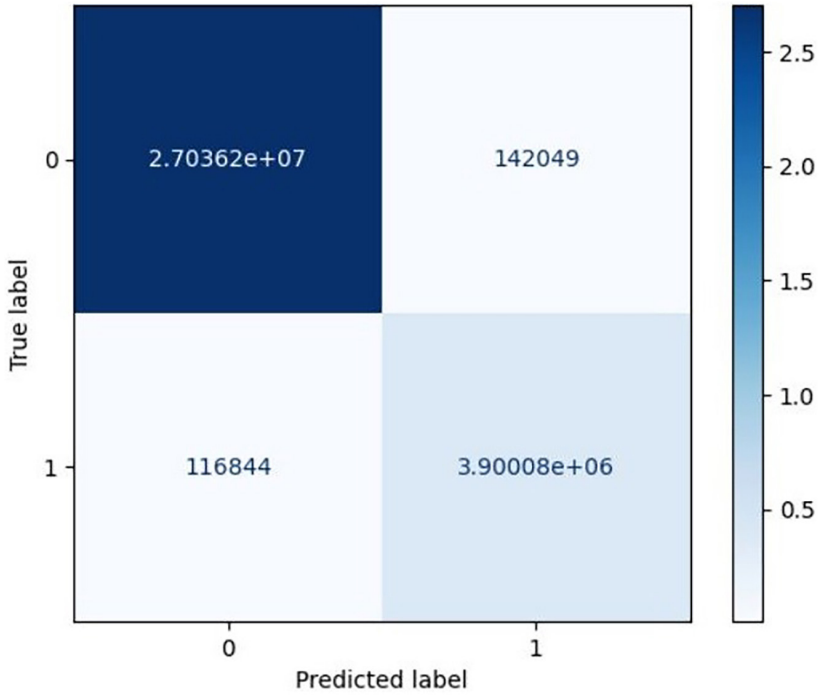

Fig. 7. Confusion matrix of the segmentation model

The confusion matrix (see Figure 7) corresponds to the evaluation of the binary segmentation model applied to pulmonary CT images, performed at the pixel level. In this context:

- Class 1 represents pixels belonging to the pulmonary regions.
- Class 0 represents pixels (background, anatomical structures other than the lung, etc.).

Analysis of this matrix highlights several significant results:

- The model accurately identifies the majority of non-pulmonary pixels, as evidenced by the very low false positive rate.
- It also effectively detects pulmonary regions, with a large number of true positives, demonstrating its ability to segment the relevant structures well.

These results confirm the model's performance and robustness on the validation set, underlining its ability to effectively separate pulmonary areas from other tissues in CT images.

After training the U-Net model, a testing phase was conducted to evaluate its performance on pulmonary images from heterogeneous sources. The model was evaluated on an international dataset to verify its generalization capability. It achieved an accuracy of 97.11%, demonstrating high performance on both healthy and pathological cases.

The results highlight its ability to accurately extract pulmonary structures (see Figure 8), even in the presence of significant contrast variations or pathological abnormalities, confirming its robustness to data diversity.

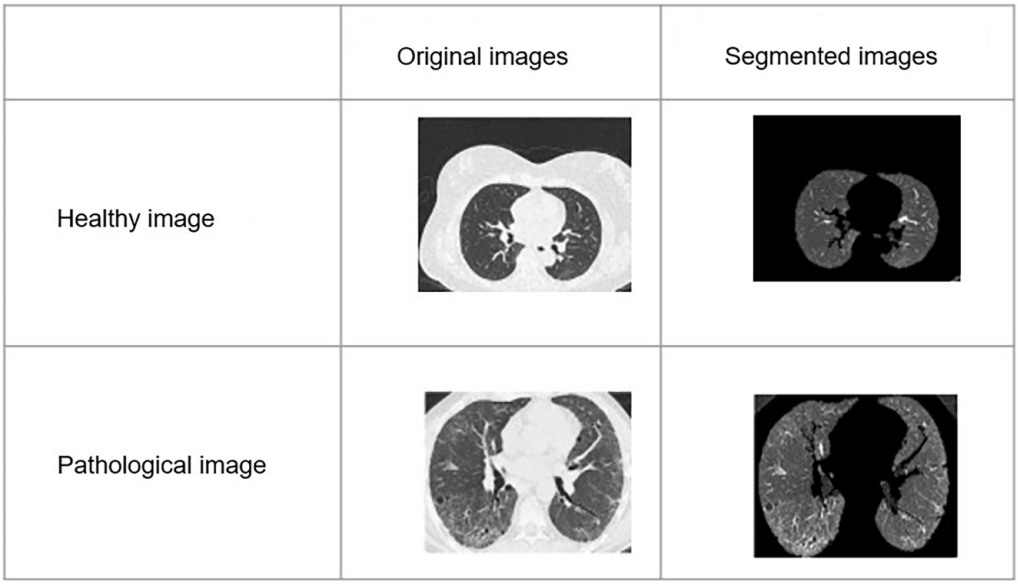

Fig. 8. Segmentation results of CT images examples

5.3 Classification results

Classification performance with and without segmentation. In this study, binary classification was conducted to distinguish healthy lungs from diseased ones using thoracic CT images. The model used is based on the EfficientNet B0 architecture, known for its excellent balance between performance and computational efficiency in computer vision tasks. This choice was motivated by the need to ensure high accuracy while minimizing resource costs, a critical factor for large-scale clinical applications. Two approaches were compared: one without prior segmentation, and another incorporating a segmentation step using the U-Net model.

Classification performance without segmentation. In this configuration, binary classification was performed to differentiate healthy from diseased lungs. The model is based on the EfficientNet B0 architecture (refer to Table 5).

Table 5. Hyperparameters of the classification model without segmentation

Hyperparameter	Value
Input image size	128×128×3
Batch size	16
Training epochs	30
Optimizer	Adam (learning rate = 10^{-5})
Loss function	Binary Crossentropy
Label smoothing	0.1
Dropout rate	0.4
Neurons in dense layer	128, ReLU activation
Final activation	Sigmoid activation

The model was trained for 15 epochs, achieving an accuracy of 93.29% and an AUC of 97.75%. These results demonstrate strong model performance. The high accuracy suggests that most predictions are correct, which is a useful global metric. The results confirm effective model training and satisfactory generalization capacity. Segmentation plays a crucial role in the preprocessing of medical images by helping to highlight key anatomical structures, filter out irrelevant background details, and ultimately enhance the accuracy of the classification process. VGG19 achieved the highest accuracy at 93.90%, showing strong performance in classifying the data. EfficientNet B0 followed closely with 93.29% accuracy, outperforming ResNet-50, which scored 89.39%. While VGG19 leads in accuracy, EfficientNet B0 stands out for offering an excellent balance between accuracy and efficiency (refer to Table 6). Its design prioritizes optimized parameters and scalability, making it lighter and faster than VGG19. This combination of competitive accuracy and computational efficiency makes EfficientNet B0 a promising choice, especially for real-time applications or environments with limited resources.

Table 6. Comparison of the accuracy for different architectures without segmenation

MModel	Accuracy (%)
ResNet-50	89.39
VGG19	93.90
EfficientNet B0	93.29

Classification performance with segmentation. In this setting, CT images were first segmented using the U-Net segmentation model to isolate pulmonary textures from the generated masks. These segmented images were then fed into the same classification model parameters. The model was also trained for 15 epochs and achieved an accuracy of 84.62% and an AUC of 93.14%. Although slightly lower than the results obtained without segmentation, the performance remains satisfactory. Given the importance of segmentation in identifying key anatomical structures, the following section introduces a new method that combines both the segmentation results and the original CT images. This integrated approach aims to improve the accuracy of classification by leveraging detailed structural information from the segmented regions alongside the raw image data.

6 CLASSIFICATION PERFORMANCE OF THE PROPOSED APPROACH

To better leverage the information contained in the segmentation masks, we proposed an innovative strategy that fuses features extracted from both the raw CT image and its corresponding segmentation mask 9. More specifically, the EfficientNet B0 model is applied independently to the original image and the mask; the resulting representations from each branch are then concatenated before the final classification stage.

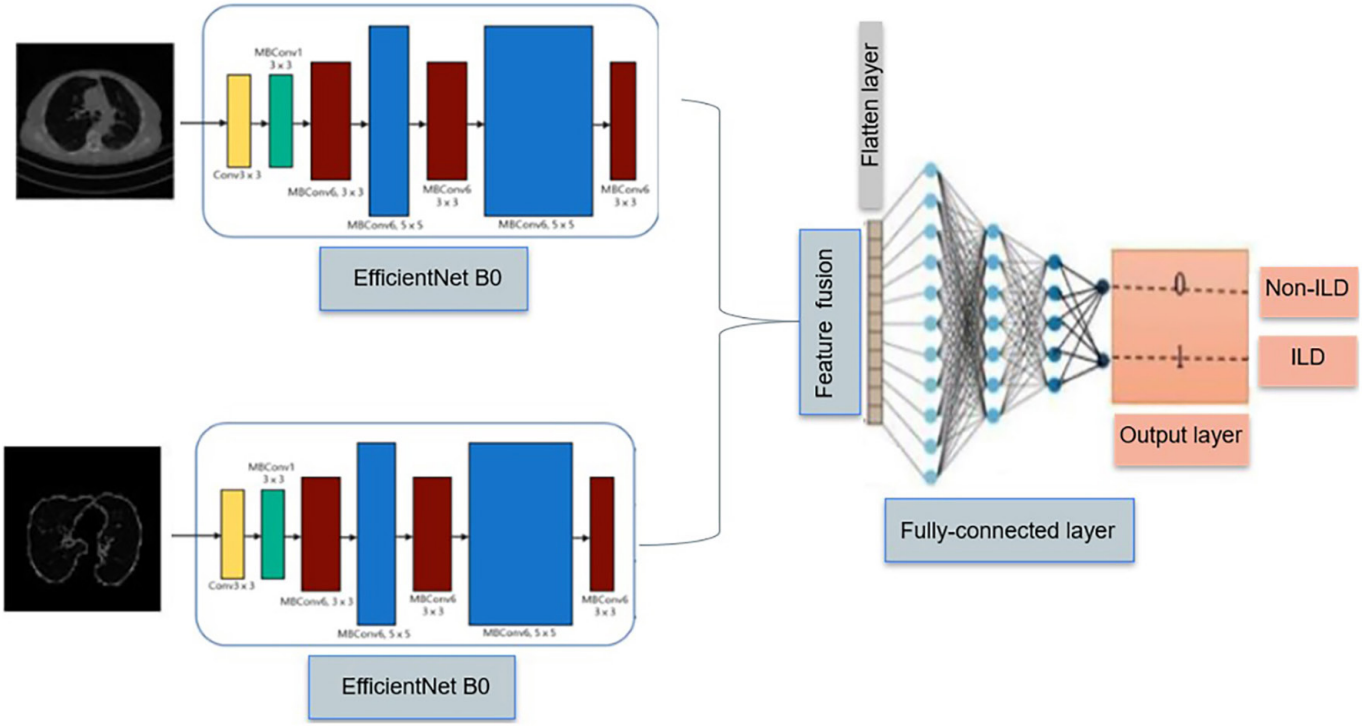

Fig. 9. Detailed overview of DI-EffNet

This approach enables the model to combine global image information with localized regions of interest identified through segmentation, offering both contextual and focused insight (see Figure 9). The goal is to enhance the model's ability to identify pathological cases by incorporating both general texture cues and localized pulmonary indicators. The performance comparison of the EfficientNet B0 model under three configurations: 1) without segmentation, 2) with U-Net-based segmentation, and 3) the proposed feature fusion approach highlights the substantial performance improvements enabled by the fusion strategy in both accuracy and AUC (refer to Table 7).

Table 7. Performance comparison of the EfficientNet B0 model under different input strategies

Configuration	Acc (%)	AUC (%)
Without segmentation	93.29	97.75
With segmentation	84.62	93.14
Proposed approach (image + mask fusion)	97.12	99.75

The model was retrained for 15 epochs using both the CT images and their corresponding segmentation masks as input. This fusion of information guided the learning process toward more relevant regions, resulting in a significant boost in overall performance. Specifically, both accuracy and AUC saw notable improvements using the new approach. These results emphasize the relevance of the proposed method, which effectively combines global image features with localized cues from segmentation masks, thereby enhancing the model's ability to discriminate between healthy and pathological cases (see Figure 10).

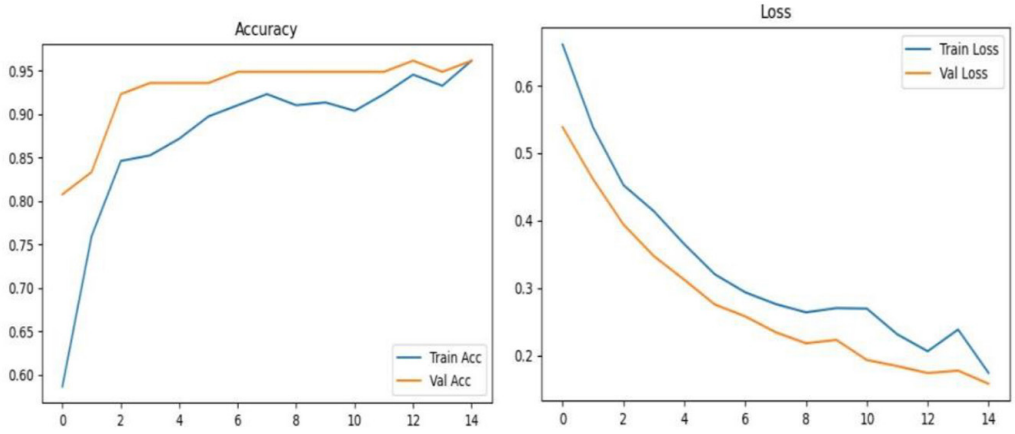

Fig. 10. Training curves of accuracy and loss during the learning process

The model performs very well, with high accuracy and low loss. There is no sign of overfitting, and training is stable. This is a sign of well-trained model.

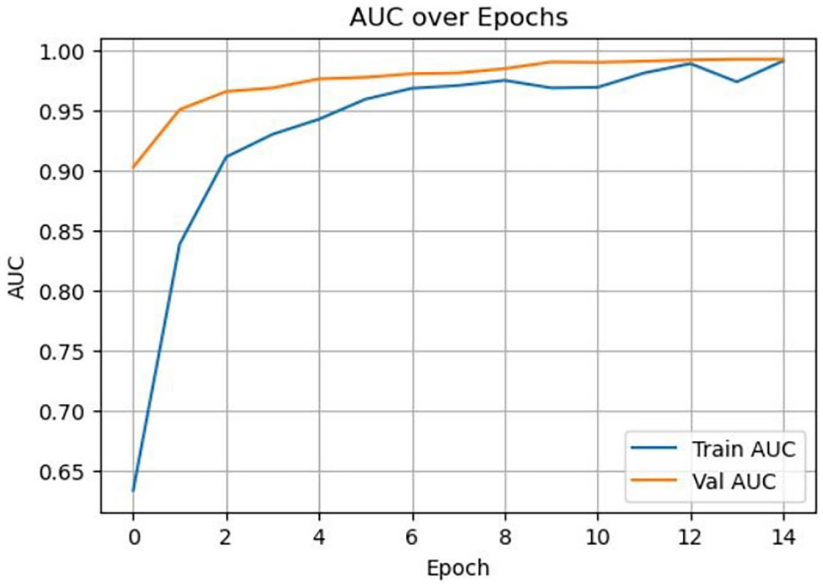

Fig. 11. ROC curve of the proposed model

The training AUC increases steadily from approximately 0.63 in the first epoch to nearly 0.99 by the final epoch, indicating progressive learning and improved class discrimination for reduced data and unbalanced classes (see Figure 11). The performance of two recent ILD classification methods were compared with the proposed DI-EffNet approach, all tested on the Medgift dataset. Kumarganesh et al. combined radiomic features with an attention-based CNN and an RBFNN neural network, achieving 92.3% (refer to Table 8).

Table 8. Comparative analysis of the proposed model vs. recent approaches in binary classification

Reference	Kumarganesh et al. [24]	Bakshi et al. [25]	Proposed Approach
Year	2025	2025	2025
Training-test plitting	70% for training	80% for training	80% for training a
	15% for Validation 15% for test	20% for testing	20% for testing
Dataset	Medgift	Medgift	Medgift
Technique	Attention CNN RBFNN	SB IDNet	DI-EffNet
Accuracy (%)	92.3	82.27	99.7

Bakshi et al. developed the SB-ID Net with parallel branches and dense connections, but despite its complexity, it only reached 82.27% accuracy, possibly due to overfitting or suboptimal training without a separate validation set. The proposed DI-EffNet model achieved a remarkable 99.7% accuracy on an unbalanced dataset, suggesting better learning and feature representation. Overall, while earlier methods made important progress, the DI-EffNet shows significant improvements and could be valuable in clinical settings. The result of the segmentation and prediction was approved by the experts (see Figure 12).

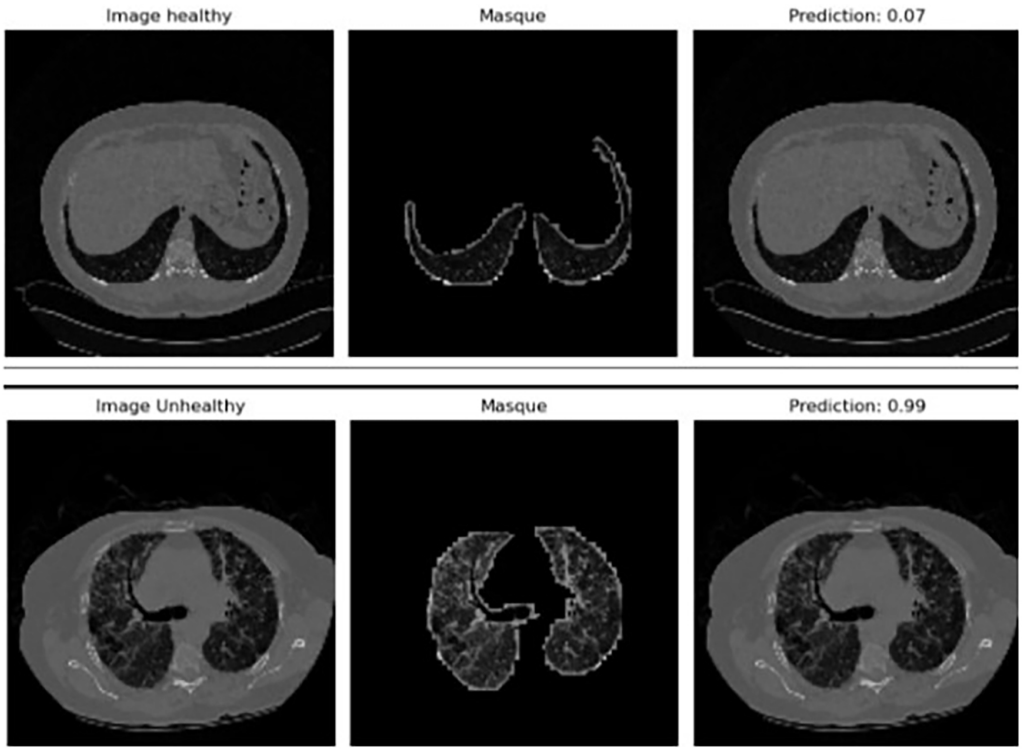

Fig. 12. Results of segmentation and classification

7 CONCLUSION

In this study, we addressed the complex task of identification of ILD patterns in high-resolution CT scans using a slice-based analysis approach. Our key innovation,

DI-EffNet, is a deep learning model featuring a specialized attention mechanism, which highlights important diagnostic features by focusing on segmented input images to reduce noise. To improve training quality and clinical relevance, we carefully re-annotated the public dataset with the help of expert radiologists, ensuring the model focuses on subtle yet critical imaging details. When compared to other advanced neural networks, DI-EffNet showed superior performance by combining raw CT slices with U-Net based segmentations, enabling it to capture complex textures and abnormalities associated with ILD more effectively. This integrated approach led to significant improvements in accuracy. In future work, we plan to integrate this attention mechanism within a Vision Transformer framework to simplify the process and improve generalization. We aim also to explore a 3D convolutional model to analyze entire scans at once, which could better capture spatial relationships in the data, although this requires more extensive data and computational power.

8 ACKNOWLEDGEMENTS

I express my sincere gratitude to Pr. Zaineb Mnif, Professor of Radiology at the Radiology Department of Polyclinique Alya, for her valuable guidance, insightful discussions, and continuous support throughout this study. Her expertise and collaboration have been instrumental in advancing this study.

9 REFERENCES

[1] M. Wijsenbeek, A. Suzuki, and T. M. Maher, "Interstitial lung diseases," *The Lancet*, vol. 400, no. 10354, pp. 769–786, 2022. https://doi.org/10.1016/S0140-6736(22)01052-2

[2] S. L. F. Walsh and D. M. Hansell, "High-resolution CT of interstitial lung disease: A continuous evolution," *Seminars in Respiratory and Critical Care Medicine*, vol. 35, no. 2, pp. 129–144, 2014. https://doi.org/10.1055/s-0034-1371526

[3] C. Szegedy *et al.*, "Going deeper with convolutions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 1–9. https://doi.org/10.1109/CVPR.2015.7298594

[4] F. Salvetti, B. Bertagni, I. Contardo, R. Gardner, J. Rudolph, and R. Minehart, "AI-driven avatars in medical training: Personalized feedback for enhanced learning," *International Journal of Advanced Corporate Learning (iJAC)*, vol. 18, no. 3, pp. 46–59, 2025. https://doi.org/10.3991/ijac.v18i3.52595

[5] F. Alebeisat, A. M. A. Awwad, A. Qatawneh, and S. Al-Suhemat, "A real-time heart attack detection and warning system for drivers using neural network," *International Journal of Interactive Mobile Technologies (iJIM)*, vol. 19, no. 20, pp. 183–203, 2025. https://doi.org/10.3991/ijim.v19i20.55789

[6] D. Chakraborty, S. Palit, and U. Bhattacharya, "Deep classification of mammographic breast density: DCBARNet," in *2023 38th International Conference on Image and Vision Computing New Zealand (IVCNZ)*, 2023, pp. 1–6. https://doi.org/10.1109/IVCNZ61134.2023.10344251

[7] S. Woo, J. Park, J.-Y, Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 3–19. https://doi.org/10.1007/978-3-030-01234-2_1

[8] K. Wang, P. Jiang, J. Meng, and X. Jiang, "Attention-based DenseNet for pneumonia classification," *IRBM*, vol. 43, no. 5, pp. 479–485, 2022. https://doi.org/10.1016/j.irbm.2021.12.004

[9] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 7132–7141. https://doi.org/10.1109/CVPR.2018.00745

[10] S. Wang *et al.*, "A deep learning algorithm using CT images to screen for coronavirus disease (COVID-19)," *European Radiology*, vol. 31, pp. 6096–6104, 2021. https://doi.org/10.1007/s00330-021-07715-1

[11] A. Vaidyanathan *et al.*, "An externally validated fully automated deep learning algorithm to classify COVID-19 and other pneumonias on chest computed tomography," *ERJ Open Research*, vol. 8, no. 2, 2022. https://doi.org/10.1183/23120541.00674-2021

[12] A. K. Abdar *et al.*, "Automatic detection of coronavirus (COVID-19) from chest CT images using VGG16-based deep learning," in *2020 27th National and 5th International Iranian Conference on Biomedical Engineering (ICBME)*, 2020, pp. 212–216. https://doi.org/10.1109/ICBME51989.2020.9319326

[13] S. Bakshi, S. Palit, U. Bhattacharya, K. Gholami, N. Hussain, and D. Mitra, "A novel CNN-based approach for distinguishing between COVID and common pneumonia," in *International Conference on Image and Vision Computing*, New Zealand, 2022, pp. 330–344. https://doi.org/10.1007/978-3-031-25825-1_24

[14] A. Singh, V. P. Gopi, A. Thomas, and O. Singh, "Dual-scale CNN architecture for COVID-19 detection from lung CT images," *Biomedical Engineering: Applications, Basis and Communications*, vol. 35, no. 3, p. 2350012, 2023. https://doi.org/10.4015/S1016237223500126

[15] F. Ciompi *et al.*, "Automatic classification of pulmonary peri-fissural nodules in computed tomography using an ensemble of 2D views and a convolutional neural network out-of-the-box," *Medical Image Analysis*, vol. 26, no. 1, pp. 195–202, 2015. https://doi.org/10.1016/j.media.2015.08.001

[16] M. Anthimopoulos, S. Christodoulidis, L. Ebner, A. Christe, and S. Mougiakakou, "Lung pattern classification for interstitial lung diseases using a deep convolutional neural network," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1207–1216, 2016. https://doi.org/10.1109/TMI.2016.2535865

[17] S. Christodoulidis, M. Anthimopoulos, L. Ebner, A. Christe, and S. Mougiakakou, "Multisource transfer learning with convolutional neural networks for lung pattern analysis," *IEEE Journal of Biomedical and Health Informatics*, vol. 21, no. 1, pp. 76–84, 2016. https://doi.org/10.1109/JBHI.2016.2636929

[18] Q. Wang, Y. Zheng, G. Yang, W. Jin, X. Chen, and Y. Yin, "Multiscale rotation-invariant convolutional neural networks for lung texture classification," *IEEE Journal of Biomedical and Health Informatics*, vol. 22, no. 1, pp. 184–195, 2017. https://doi.org/10.1109/JBHI.2017.2685586

[19] G.B. Kim *et al.*, "Comparison of shallow and deep learning methods on classifying the regional pattern of diffuse lung disease," *Journal of Digital Imaging*, vol. 31, pp. 415–424, 2018. https://doi.org/10.1007/s10278-017-0028-9

[20] M. Gao *et al.*, "Holistic classification of CT attenuation patterns for interstitial lung diseases via deep convolutional neural networks," *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, vol. 6, no. 1, pp. 1–6, 2018. https://doi.org/10.1080/21681163.2015.1124249

[21] H.-C. Shin *et al.*, "Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1285–1298, 2016. https://doi.org/10.1109/TMI.2016.2528162

[22] M. Anthimopoulos *et al.*, "Semantic segmentation of pathological lung tissue with dilated fully convolutional networks," *IEEE Journal of Biomedical and Health Informatics*, vol. 23, no. 2, pp. 714–722, 2018. https://doi.org/10.1109/JBHI.2018.2818620

[23] J. H. Oh, G. H. J. Kim, and J. W. Song, "Interstitial lung abnormality evaluated by an automated quantification system: Prevalence and progression rate," *Respiratory Research*, vol. 25, no. 1, p. 78, 2024. https://doi.org/10.1186/s12931-024-02715-3

[24] S. Kumarganesh *et al.*, "Aggregated approach for interstitial lung diseases classification using attention-based CNN and radial basis function neural network," *Systems and Soft Computing*, vol. 7, p. 200228, 2025. https://doi.org/10.1016/j.sasc.2025.200228

[25] S. Bakshi, S. Palit, U. Bhattacharya, and S. Baksi, "Identification of interstitial lung disease: Breaking barriers with SB-ID Net," in *International Conference on Neural Information Processing*, Springer, 2024, pp. 137–152. https://doi.org/10.1007/978-981-96-6954-7_10

[26] A. Depeursinge *et al.*, "Building a reference multimedia database for interstitial lung diseases," *Computerized Medical Imaging and Graphics*, vol. 36, no. 3, pp. 227–238, 2012. https://doi.org/10.1016/j.compmedimag.2011.07.003

[27] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, Springer, 2015, pp. 234–241. https://doi.org/10.1007/978-3-319-24574-4_28

[28] S. Nigam, R. Jain, V. K. Singh, S. Marwaha, A. Arora, and S. Jain, "EfficientNet architecture and attention mechanism-based wheat disease identification model," *Procedia Computer Science*, vol. 235, pp. 383–393, 2014. https://doi.org/10.1016/j.procs.2024.02.067

[29] Y. Shen and Y. Shang, "Online teaching effect evaluation and analysis using combined weighting technique," *International Journal of Emerging Technologies in Learning (iJET)*, vol. 19, no. 3, pp. 67–81, 2024. https://doi.org/10.3991/ijet.v19i03.47667

10 AUTHORS

Norhène Gargouri is a researcher at the Lab. SMARTS, Digital Research Center of Sfax (CRNS), Tunisia, and currently serves as an Associate professor in Image Processing. Her research interests include medical imaging, computer-aided diagnosis systems and deep learning (E-mail: norhene.gargouri@crns.tn).

Nesrine Charfi is an Assistant at the Higher Institute of Technological Studies (ISET) of Kef, Boulifa – University Campus. Her research interests include artificial intelligence, biometrics, medical image processing, and computer vision (E-mail: charfi.nesrine@gmail.com).

Alima Damak Masmoudi is a Professor at Faculty of Science of Sfax, Tunisia. Her research interests include artificial intelligence, biometrics, computer-aided diagnosis systems, and deep learning for healthcare applications (E-mail: alima.damak@fss.usf.tn).

Wiem Feki is an Associate Professor and researcher at the Faculty of Medicine of Sfax, Tunisia, and a radiologist at CHU Sfax. Her research interests include medical imaging, radiology, and computer-aided diagnosis systems (E-mail: waima_feki@yahoo.fr).

Chifa Damak is a Doctor of Internal Medicine with a private medical practice. Her professional interests include internal medicine, patient care, and clinical diagnosis (E-mail: chifa.damak@example.org).