

# Progressive and Combined Deep Transfer Learning for pneumonia diagnosis in chest X-ray images

Mamar Khaled<sup>1</sup>, Djamel Gaceb<sup>1</sup>, Fayçal Touazi<sup>1</sup>, Ahmed Otsmane<sup>1</sup> and Farouk Boutoutaou<sup>1</sup>

*1 LIMOSE Laborator, University M'Hamed Bougara of Boumerdes, Boumerdès, 35000, Algeria*

## Abstract

Pneumonia is a life-threatening disease that occurs in the lungs and is caused by a bacterial or viral infection. However, it is very difficult to diagnose it by simply looking at chest x-rays, because it is necessary to improve diagnostic accuracy. This study aims to simplify the process of detecting and classifying pneumonia for both experts and patients, using a dataset containing 5247 CXR images. Five different pretrained CNNs: AlexNet, VGG-16, ResNet-50, DenseNet-121 and InceptionV3 were used separately or together for transfer learning in a progressive way. Firstly, they are pretrained on ImageNet dataset, and secondly, on a radiographic images which concerns another disease (available in medium size with a nature close to our base). These models are refined according to different fine-tuning levels and strategies. A weighted classifier-based approach is introduced to combine their weighted prediction. The results obtained show the possibility of moving easily from the classification of a disease to another using a progressive transfer learning, which has a limited number of images by taking advantage of the knowledge already acquired on another very large base.

## Keywords

Pneumonia, deep learning; progressive transfer learning; medical image processing; computer-aided diagnosis.

## 1. Introduction

Medical imaging has a key role in the detection and classification of diseases. Although chest x-rays (CXR) have lower resolution than magnetic resonance imaging (MRI) or computed tomography (CT), and they adopt low-cost and easy to-use technology. Pneumonia is an acute respiratory infection that affects the lungs. This disease is considered the leading cause of infant mortality worldwide. About 1.4 million children die of pneumonia every year, or 18% of the total number of children who die under the age of five. There are mainly two types of pneumonia: viral and bacterial. Generally, viral pneumonia tends to be mild while bacterial pneumonia is more severe, especially in children. Inflammations seen as white patches on chest x-ray are due to various abnormalities such as pneumonia, tuberculosis, pneumothorax, pleural effusion, etc. hence the ambiguity in the diagnosis of pneumonia from chest X-ray images, either by radiologists or by computer-aided diagnosis (CAD) systems. This article studies the application of deep transfer learning in medical imaging to help diagnose pneumonia, it is organized as follows: Section 2 expose some related works; Section 3 explains the proposed method; section 4 details the experiments and results performed on the system, and Section 5 presents the discussion and conclusions.

## 2. Related works

Computer-aided diagnosis (CADE or CADx) has become very popular these days. Various methods and works have been proposed in recent years to improve the precision of detection of pulmonary diseases on CXRs.

## 2.1. Machine learning based methods

One of the primary use cases for machine learning in healthcare is in the early detection and effective diagnosis of disease. Certain diseases such as hereditary or genetic disorders and certain types of cancer are difficult to identify at an early stage by a specialist. However, the use of intelligent computing solutions through machine learning can well detect these. These solutions have continued to evolve using innovative learning algorithms on datasets representing different diseases with varying complexity. In general, these systems consist of two main parts: a discriminant feature extractor (which must be chosen by computer vision and health experts, or dedicated algorithms) and a classifier (based on machine learning using a representative image base).

The authors of [1] and [2], used the Random Forest classifier (RF), the first was devoted to colon cancer, leukemia, breast cancer and lung cancer, while the second was devoted to breast cancer. The RF classifier was combined with a most relevant features selector where they obtained better classification rates by comparing their model with 15 classifiers. Several medical diagnostic works are based on the SVM classifier, such as the work of El-Naqa et al. [3] which reached a rate of 94% for the detection of microcalcification (MC) clusters in digital mammograms. Ghumbre et al. [4] have combined the SVM classifier with a Sequential Minimal Optimization (SMO) algorithm for diagnosing heart disease. Chaurasia and Pal [5], compared different classification techniques (naive Bayesian classifiers, non-linear SVM with RBF kernel, RBF neural networks and decision trees) for the diagnosis of breast cancer. The results obtained showed the efficiency of the SVM-RBF classifier with an accuracy of 96.84%. Christobel and Sivaprakasa [6] have obtained an accuracy of 97.13% using SVM classifier, which was compared to the Naive Bayes, KNN classifiers.

The detection of pneumonia by chest X-ray has been an open problem for many years, the main limitation being the scarcity of publicly available data. Yao and al. [7] achieved an accuracy of 80% in the pneumonia detection task with the SVM classifier. While Naydenova and al. [8] compared the SVM classifier with RF and Logistic Regression for automated detection of childhood pneumonia in resource-constrained settings. Chandra and al. [9] developed an approach based on Multi Layers Perceptron (MLP), Random Forest (RF), sequential minimal optimization (SMO), regression classification and logistic regression for the pneumonia detection on Chest Xray. They segmented lung regions from chest X-ray images and extracted eight (08) statistical features from these regions. They obtained an accuracy rate of 95.39% for the MLP classifier. Kuo et al. [10], used 11 features to detect pneumonia in 185 schizophrenic patients, using decision tree, SVM and logistic regression (LR) classifiers. The highest accuracy rate (94.5%) was for the decision tree classifier. Along the same lines, Yue et al. [11] used 6 features with LR and RF classifiers for predicting hospital stay in patients with pneumonia associated with SARS-CoV-2 infection; the best AUC value they got was 97%. Sousa and al. [12] used a pneumonia detection algorithm with five classifiers KNN, Naive Bayes, Multi Layers Perceptron, Decision Tree and SVM, combined with different dimensionality reduction techniques. Their algorithm achieved an accuracy of 96%.

## 2.2. Deep learning-based methods

Deep learning has quickly established itself as a standard in several fields, often with better performance compared to conventional machine learning algorithms (presented previously). Recently, a number of researchers have come up with different artificial intelligence (AI) based solutions for different medical problems. Convolutional Neural Networks (CNNs) have enabled researchers to achieve positive results in broad medical problems such as detection of thyroid nodules, breast cancer, detection and segmentation of brain tumors, classification of diseases in x-ray images, etc. The authors of [13] proposed a hybrid method for the diagnosis of thyroid nodules on ultra-sound images. They combined two pretrained CNNs with different convolutional layers and different fully connected layers. The features maps of the two networks obtained after refinement are merged and used as input to the softmax classifier. This method has been validated on 15,000 ultrasound images. Gulshan and al. [14] applied a DCNN for the automatic detection of diabetic retinopathy (DR) on the EyePACS-1 dataset with more than 10,000 images. Haloi [15] implemented a five-layer CNN with drop-out

mechanism for early-stage DR detection on two different datasets (Retinopathy Online Challenge (ROC) and Massidor). Rakhlin and al [16] classified images of H&E stained breast tissue. For each image, 20 crops of 400×400 pixels and 650×650 were extracted. Then, the pretrained ResNet-50, InceptionV3 and VGG-16 networks were used as features extractors. Then, they were combined into a single features vector. A Light GBM classifier with 10-fold cross validation was used to classify deep features. Van and al. [17], used the CNN InceptionV3 model for multi-class classification of breast cancer images. Due to the voluminousness histology images, they used Inception V3 for patch-level classification. The predictions were then passed through an ensemble fusion framework involving majority voting, their proposed ensemble classifier included Gradient Boosting Machine (GBM) and Logistic Regression to get the final prediction per frame. The refinement achieved an accuracy of 87.50%.

In the article [18], the authors carried out a comparative study of different CNN networks in the diagnosis of pulmonary diseases. They collected a sample of 357 images of the different diseases, and a healthy sample (100 images), and having analyzed 38 non-image related features, including complaints of cough, weakness, chest pain and high body temperature. The neural network with the Levenberg Marquardt algorithm achieved the highest accuracy for the detection of pneumonia (91.67%). In 2018, some researchers highlighted a visualization technique in conjunction with CNNs to locate and detect Regions Of Interest (ROIs) that can be used to identify pneumonia and distinguish between bacterial and viral types in pediatrics. In this sense, we find the work of the authors of [19] who evaluated the performance of different CNN architectures: Sequential CNN, Inception CNN, Residual CNN and VGG16 on a dataset of 5232 pediatric chest radiographs. Their approach was based on a visualization technique to define the ROI. The customized VGG16 achieved 96.2% accuracy in pneumonia detection and 91.8% in its classification. Xianghong and al. [20] used a 08-layer FCN model with transfer learning to segment anatomical lung regions. The model was trained and tested on both JSRT (241 images) and MC (138 images) datasets. After segmentation, an AlexNet DCNN model was used to classify the lung regions. A binary classification [21] was performed using the SVM-RBF kernel by combining DCNN and manual functionalities. Features extraction by DCNN with transfer learning achieved accuracy ( $0.8048 \pm 0.0202$ ) and better sensitivity ( $0.7755 \pm 0.0296$ ). Rajpurka and al. [22] proposed the Chexnet algorithm for pneumonia detection and tested its accuracy against 4 radiologists, using an F1-score metric. The algorithm worked as the same as an experienced radiologist; however, it exceeded the performance of an average radiologist. Jaiswal and al. [23] proposed an identification model inspired by the Mask-RCNN model that included critical modifications to the training process and a new post-processing step that merges the bounding boxes of multiple models. Ayan and al. [24] used two CNN network models: Xception and VGG16 for the diagnosis of pneumonia on the Kermany dataset (5856 images). Relying on transfer learning and fine-tuning on the Xception model, hence the use of pretrained ImageNet weights before the start of training and the last 10 first layers were frozen. The test results showed that the Vgg16 network outperforms the Xception network. Varshani and al. [25] used ResNet50, DenseNet-121 and DenseNet169 as the optimal CNN models for the features extraction step and different classifiers such as RF, SVM, etc. The best results favored the SVM classifier. In 2020, Vikash and al. [26], used AlexNet, DenseNet121, Inception V3, GoogLeNet and Res-Net18 pe-trained on the ImageNet basis. Then they came up with an ensemble model that combines the outputs of all these models into a prediction vector, and a majority vote is used to choose the final prediction. This combination achieved a test accuracy of 96.39%, with an area under the ROC curve of 99.34% and a sensitivity of 99.62%. Rahman and al. [27] attempted to automatically diagnose different classes of pneumonia (bacterial and viral), on the Kaggle Chest X-Ray pneumonia dataset, it includes 5247 X-ray images.

Data training is done using the pretrained algorithms: AlexNet, ResNet18, DenseNet201 and SqueezeNet. They observed that DenseNet201 outperforms the other three CNNs, achieving 98% accuracy in pneumonia detection and accuracy of 93.3% in the differentiation between the two etiological variants. Hashmi and al. [28], presented a method combining Five CNNs: ResNet18, Dense-Net121, InceptionV3, Xception and MobileNetV2 for the automatic pneumonia detection, using data augmentation, after that the five CNNs were refined to pneumonia classification. Then the predictions were combined, using a weighted classifier to calculate the final prediction. Their model achieved an accuracy of 98.857%, and furthermore, a high F1-score of 99.002 and an AUC score of 99.809.

In 2021, the authors of [29], proposed a novel end-to-end Deep Transfer Learning framework using deep convolutional neural network that detects and classifies three types of pneumonia from chest X-ray scans. In the same year, Alqudah and al. [30] constructed two hybrid models, namely CNN-KNN and CNN-SVM, while using a 10-fold cross-validation methodology. The proposed CNN has been trained, validated and tested using a large chest X-ray image dataset of 5852 images. The hybrid CNN-KNN model achieved an accuracy of 94.03%, while the CNN-SVM model achieved an accuracy of 93.9%. In the work of [31], we meet a deep learning-based approach to diagnose and classify pneumonia from Chest X-ray images using transfer learning based on three pre-trained architectures (ResNet50, InceptionV3 and InceptionResNetV2).

Polat and al. [32] proposed a classification approach based on extracting relevant features from digital chest radiographs. They used a binary CNN and a three-class CNN to detect pneumonia in 5840 pediatric CXR images. They also used a minimum distance classifier for classification. They conducted three different parametric studies and found that the proposed method achieved 100% accuracy in detecting pneumonia, 92% in distinguishing between two types of pneumonia, and 90% in distinguishing normal, bacterial, or viral pneumonia.

### 3. Proposed approach

CNNs perform better on large datasets; however, our target dataset is very small. This requires the use of data augmentation step to enrich the small base and the transfer learning strategy to take advantage of the consistency of a pretrained network on a largest dataset. This strategy can improve significantly the classification rates.

The proposed approach is based on transfer learning of five CNN architectures (AlexNet, ResNet-50V2, DenseNet-121, VGG-16 and InceptionV3) for pneumonia detection from chest x-rays. To solve the dataset size issue, we also propose a new strategy of transfer learning, called progressive transfer learning, that adopts a pretrained CNN with a large amount of acceptable medical images (See Figure 1, 2). In order to evaluate the performance of the proposed approach, three transfer learning approaches have been developed, using different strategies (See Figure 1, 2, 3 and 4). First, the five models are pretrained on the ImageNet dataset, and second, are pretrained and refined on an intermediate Tuberculosis image database (Kermany dataset). Then we reused the pretrained model progressively on these two datasets to learn how to classify pneumonia on our dataset (Curated Xray dataset). The third approach consists of combining the five CNN models, to take advantage of the performance of each approach.

#### 1. Approach 1: Direct transfer learning from ImageNet to Curated Xray dataset:

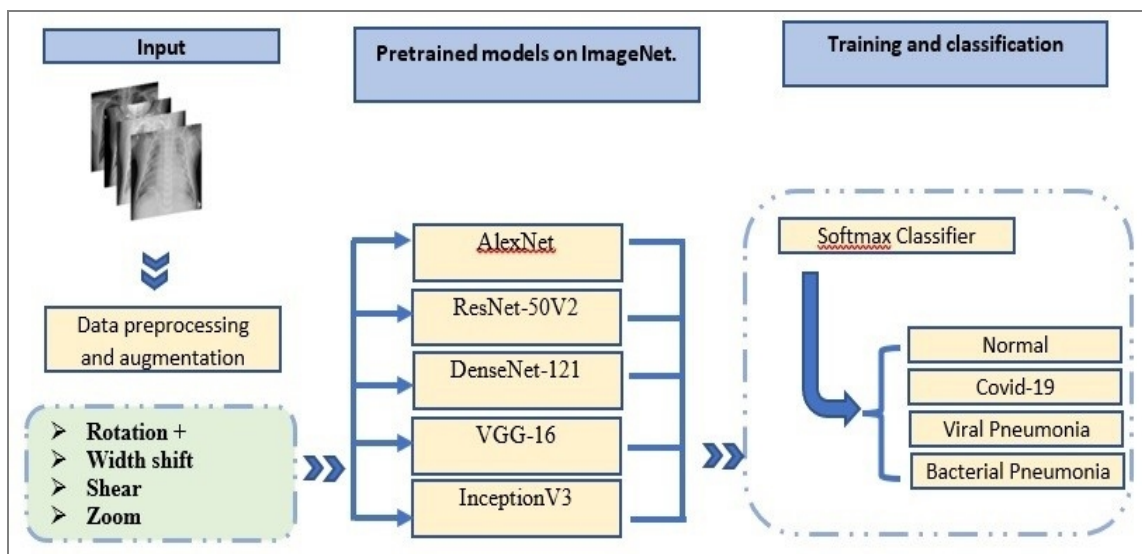
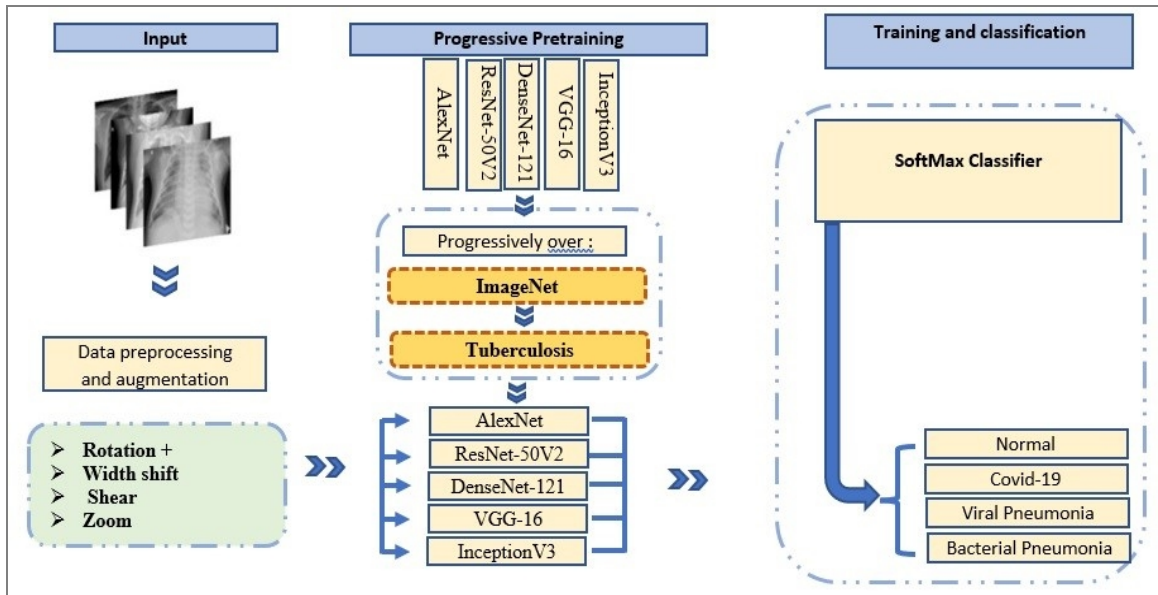


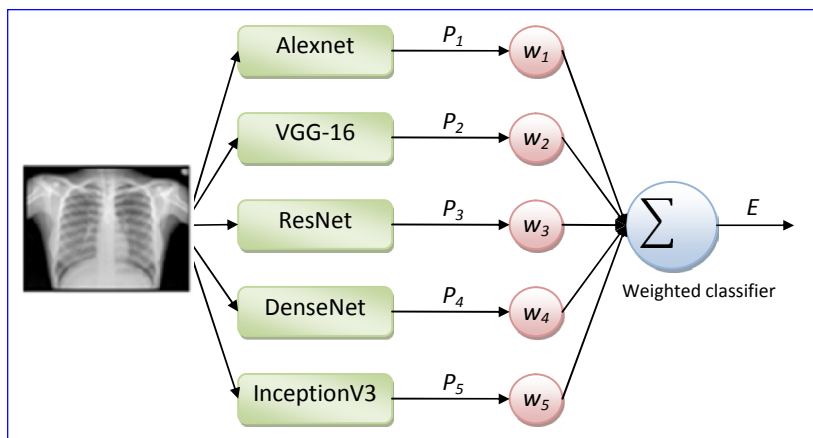
Figure 1: Direct transfer learning from ImageNet to Curated Xray dataset.

2. **Approach 2:** Progressive transfer learning, form ImageNet to Kermany dataset, then to Curated Xray dataset



**Figure 2:** Progressive transfer learning, form ImageNet to Kermany dataset, to Curated Xray dataset

3. **Approach 3:** The combined architecture (called AVRDIS) is based on the combination by weighting of the five CNN models (AlexNet, VGG-16, ResNet-50, DenseNet-121 and InceptionV3) using the Softmax classifier (See Figure 3).



**Figure 3:** Weighted classifier model of proposed combined approach.

After feature extraction, different supervised classifiers (SVM, RF, Gaussian Naïve Bayes : GNB and softmax) were used for the pneumonia classification task in combination with CNN part of feature extraction. The proposed method is carried out in three steps:

- the first consists in extracting the features of the radiological images of pneumonia; such that the latter are passed as inputs to the five CNN models, the output is a feature vector which depends on this image.
- The second step consists in passing this feature vector as input of one of the four classifiers : SVM, RF, GNB and softmax.
- For the last step, the predictions of these models were combined, using a weighted classifier to calculate the final prediction.

In summary, this strategy consists in passing as input for each architecture preprocessed radiological images of pneumonia, then providing as output the following predictions: PNEUMONIA

or NORMAL for the first base (binary classification); Bacterial Pneumonia, Viral Pneumonia, Normal and Covid-19 for the second base (4-class multiple classification).

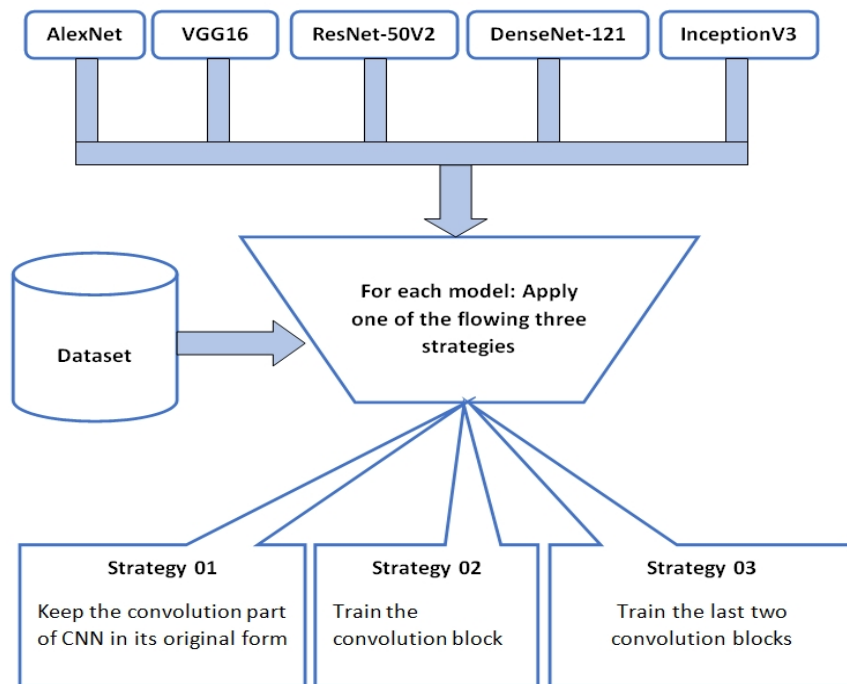
Weighted average assembly is a powerful classifier fusion mechanism. However, the choice of weights to be assigned to the respective basic learners plays a central role in ensuring the success of the whole. In this work, we adopted a weighted mean ensemble technique for better classification. To find the best combination of weights that gives the maximum precision, we have implemented a method that consists of calculating the final precision for each combination of weights, and the weights with the best precision found will be returned.

A weight ( $w_k$ ) takes a value between 0 and 1, each model, after being refined, returned the probabilities for each class label, i.e. 2 classes in the form of a matrix (P). The weights ( $w_k$ ) are multiplied by the corresponding basic learners probabilities  $P_k$  to calculate the weighted average probability ensemble, as shown in Equation (1).

$$E = \sum_{k=1}^5 W_k \times P_k \quad (1)$$

The convolution layers of the five CNNs can be trained according to the following three transfer learning strategies (See Figure 4).

- Strategy 1: Using a pretrained model as a feature extractor, and for that one has to freeze all the convolution blocks to keep the convolution part of CNN in its original form, and then use its outputs to feed an updated classifier, this strategy is generally used in the case of small databases where there is a lack of calculation.
- Strategy 2: Building on the idea that the lower layers of a CNN refer to general features (problem independent), while the upper layers refer to specific features (problem-dependent), it serves to replace the last layer Fully-Connected by the new randomly initialized classifier, and we fix the parameters of the last convolution block of the pretrained network.
- Strategy 3: It has the same principle as the previous strategy, except that instead of training only one block, the last two convolution blocks are trained. The AlexNet architecture is not affected by this strategy because it contains only 5 convolution layers.



**Figure 4:** The three Training strategies.

## 4. Experimentations and results

### 4.1. Datasets

To evaluate the proposed approach, we used two different datasets: 1) Kermamy dataset containing 5840 images labeled in two categories (Normal: healthy, and pneumonia: viral or bacterial). This dataset is subdivided into 90% for training task and 10% for test task. 2) Curated Xray dataset with a total of 18417 images labeled in 4 categories (Normal: healthy, Bacterial Pneumonia, viral Pneumonia, COVID19), subdivided into 91% for training task and 9% for test task.

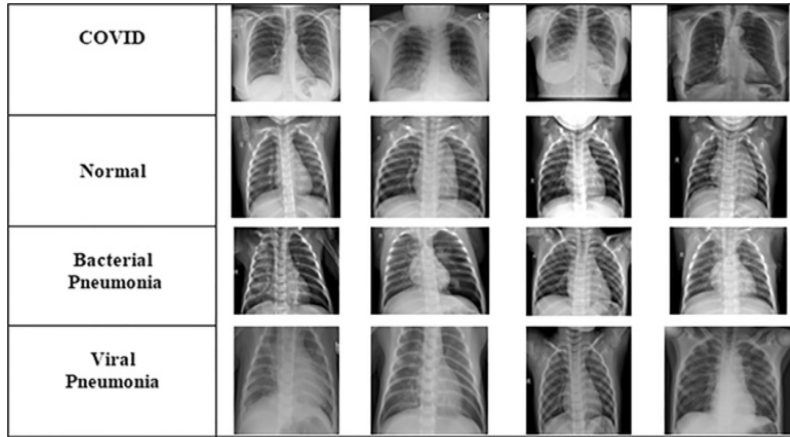


Figure 5: Curated Xray image samples [33].

### 4.2. Evaluation metrics

The proposed models were evaluated by using some evaluation metrics such as accuracy, recall, precision, and f-score. The metrics formulas are given below:

- $Accuracy = (TP+TN)/(TP+FN+TN+FP)$  (2)

- $Precision = TP/(TP+FP)$  (3)

- $Recall = TP/(TP+FN)$  (4)

- $F1-Score = 2 \times (precision \times recall) / (precision + recall)$  (5)

TP, TN, FN, FP represents the number of true positive, true negative, false negative, false positive respectively.

### 4.3. Results of approach 1 : Direct transfer learning method from ImageNet dataset

#### 4.3.1. To Kermamy dataset

Tables 1 and 2 summarize the results obtained by the 5 models without and with image augmentation respectively using the Kermamy database, including the transfer of learning based on ImageNet dataset and refined with the three fine-tuning strategies quoted previously. Expressed in terms of accuracy, precision, recall and F1-score. And we notice that the rates obtained with image augmentation surpass the rates obtained without data augmentation. These results confirm that this methodology is relevant to apply to our problem. The parameters used in data augmentation are: (Rotation = 20°, Width shift = 0.1, Shear = 0.2, Zoom=0.2).

**Table 1**

Models performances without data augmentation

	Strategy One			
	Accuracy	Precision	Recall	F1-score
VGG-16	75.00	71.99	98.20	83.08
ReseNet-50V2	63.30	63.04	99.74	77.25
DenseNet-121	80.28	76.43	98.97	86.25
InceptionV3	82.37	79.04	97.69	87.38
AlexNet	80.60	77.28	97.96	86.29
	Strategy Two			
	Accuracy	Precision	Recall	F1-score
VGG-16	81.57	77.44	99.48	87.09
ReseNet-50V2	79.32	77.35	94.61	85.12
DenseNet-121	78.06	74.37	99.74	85.21
InceptionV3	79.80	75.88	99.23	86.00
AlexNet	78.52	74.71	99.23	84.24
	Strategy Three			
	Accuracy	Precision	Recall	F1-score
VGG-16	75.48	71.90	99.74	83.56
ReseNet-50V2	78.68	75.04	98.71	85.27
DenseNet-121	80.44	76.27	99.74	86.44
InceptionV3	80.28	76.43	98.97	86.25
AlexNet	-	-	-	-

**Table 2**

Models performances with data augmentation

	Strategy One			
	Accuracy	Precision	Recall	F1-score
VGG-16	81.89	80.17	94.5	86.69
ReseNet-50V2	90.86	89.92	96.15	92.93
DenseNet-121	85.89	84.31	95.12	89.39
InceptionV3	84.45	83.67	93.33	88.24
AlexNet	86.69	88.08	91.02	89.53
	Strategy Two			
	Accuracy	Precision	Recall	F1-score
VGG-16	91.34	90.19	96.66	93.31
ReseNet-50V2	93.42	92.45	97.43	94.55
DenseNet-121	89.74	88.26	96.41	92.15
InceptionV3	93.38	88.01	97.94	92.71
AlexNet	85.09	81.13	99.23	89.27
	Strategy Three			
	Accuracy	Precision	Recall	F1-score
VGG-16	91.98	89.90	98.20	93.87
ReseNet-50V2	92.94	93.46	95.38	94.41
DenseNet-121	89.42	87.15	97.43	92.00
InceptionV3	94.87	94.08	97.94	95.97
AlexNet	-	-	-	-

Tables 3 and 4 show the performances of different classifiers (Softmax, SVM, RF and GNB ) on the Kermamy dataset (Table 3 without data augmentation and Table 4 with data augmentation). This experiment allowed us to evaluate the impact of the choice of classifier type on the performance of the network. we can notice that machine learning algorithms (such as SVM and RF) perform better than a simple supervised softmax classifier.



**Table 3**

Classifiers performances (Softmax, SVM, RF and GNB ) with data augmentation.

	Strategy One			
	Accuracy	Precision	Recall	F1-score
VGG-16 (Softmax)	75.00	71.99	98.20	83.08
VGG-16 (SVM)	83.79	86.12	83.79	82.67
VGG-16 (RF)	84.92	86.84	84.92	84.04
VGG-16 (GNB)	85.25	85.14	85.25	85.16
	Strategy Two			
	Accuracy	Precision	Recall	F1-score
VGG-16 (Softmax)	81.57	77.44	99.48	87.09
VGG-16 (SVM)	88.16	89.09	88.16	87.74
VGG-16 (RF)	86.87	88.30	86.70	86.25
VGG-16 (GNB)	86.22	87.56	86.22	85.57
	Strategy Three			
	Accuracy	Precision	Recall	F1-score
VGG-16 (Softmax)	75.48	71.90	99.74	83.56
VGG-16 (SVM)	88.16	87.68	88.16	89.41
VGG-16 (RF)	86.54	87.40	86.54	86.04
VGG-16 (GNB)	87.84	88.74	87.84	87.40

**Table 4**

Classifiers performances (Softmax, SVM, RF and GNB) with data augmentation

	Strategy One			
	Accuracy	Precision	Recall	F1-score
VGG-16 (Softmax)	81.89	80.17	94.35	86.69
VGG-16 (SVM)	88.00	88.15	88.00	87.78
VGG-16 (RF)	86.78	87.02	86.87	86.60
VGG-16 (GNB)	79.90	93.09	79.90	80.22
	Strategy Two			
	Accuracy	Precision	Recall	F1-score
VGG-16 (Softmax)	91.34	90.19	96.66	93.31
VGG-16 (SVM)	90.11	90.62	96.11	89.86
VGG-16 (RF)	90.92	91.03	90.92	91.03
VGG-16 (GNB)	89.14	89.33	89.14	89.14
	Strategy Three			
	Accuracy	Precision	Recall	F1-score
VGG-16 (Softmax)	91.98	89.90	98.20	93.87
VGG-16 (SVM)	93.35	93.39	93.35	93.90
VGG-16 (RF)	92.54	92.52	92.54	92.51
VGG-16 (GNB)	92.41	91.87	91.41	91.48

### 4.3.2. To Curated Xray dataset

Table 5 displays the found metric values of the five models for the three strategies, with data-augmentation. This experiment was conducted with the aim of evaluating our 5 models on a broader basis and with more classes (4 classes including viral, bacterial, Covid and normal pneumonia). The results obtained by the models (using previous experience) with image augmentation using the Curated X-ray image dataset are summarized in Table 5. The DenseNet model outperformed all other models with an F1-score equal to 92.49%.

**Table 5**

Models performances with data augmentation on Curated Xray dataset .

	Strategy One			
	Accuracy	Precision	Recall	F1-score
VGG-16	72.79	73.50	68.27	64.00
ReseNet-50V2	84.06	84.00	81.75	82.50
DenseNet-121	82.47	81.25	80.50	80.50
InceptionV3	90.07	89.50	88.75	89.25
AlexNet	79.65	79.00	77.50	80.00
	Strategy Two			
	Accuracy	Precision	Recall	F1-score
VGG-16	87.99	87.50	87.00	87.00
ReseNet-50V2	91.91	92.25	90.50	91.00
DenseNet-121	88.23	88.25	86.25	86.75
InceptionV3	89.95	89.25	90.23	89.90
AlexNet	81.86	81.75	78.75	82.00
	Strategy Three			
	Accuracy	Precision	Recall	F1-score
VGG-16	87.13	88.00	83.75	84.50
ReseNet-50V2	87.13	88.00	83.75	84.50
DenseNet-121	87.75	87.61	97.94	92.49
InceptionV3	88.84	89.00	86.50	87.25
AlexNet	-	-	-	-

#### 4.4. Results of approach 2: Progressive transfer learning

Table 6 shows the models performance using progressive transfer learning on the Kermany dataset to Curated Xray dataset, with data augmentation (using three fine-tuning strategies).

**Table 6**

Models performances, using progressive transfer learning on the Kermany dataset to Curated Xray dataset, with data augmentation.

	Strategy One			
	Accuracy	Precision	Recall	F1-score
VGG-16	82.53	83.53	89.74	86.52
ReseNet-50V2	87.50	94.31	85.12	89.48
DenseNet-121	89.42	87.85	96.41	91.93
InceptionV3	82.37	87.43	83.84	85.60
AlexNet	90.70	87.89	98.71	92.99
	Strategy Two			
	Accuracy	Precision	Recall	F1-score
VGG-16	90.06	87.27	98.46	92.53
ReseNet-50V2	93.58	96.79	92.82	94.76
DenseNet-121	93.10	92.00	97.43	94.64
InceptionV3	91.50	92.87	93.58	93.23
AlexNet	92.78	92.17	99.66	94.36
	Strategy Three			
	Accuracy	Precision	Recall	F1-score
VGG-16	93.42	94.62	94.87	94.75
ReseNet-50V2	92.94	96.01	92.56	94.25
DenseNet-121	88.30	85.14	98.46	91.31
InceptionV3	91.34	94.21	91.79	92.98
AlexNet	/	/	/	/

#### 4.5. Comparison between the two approaches

The tables presented above allow us to see how the right choice of the training dataset and the transfer of learning helps to improve performance. As well as the increase in data that has played a crucial role in this improvement and confirms all its performance in the medical field where available data is scarce. Table 7 shows a comparison of the two approaches (progressive TL and direct TL) using F1-score metric. These results show that progressive transfer learning (Approach 2) offers better performance compared to direct transfer learning from ImageNet (Approach 1).

**Table 7**

Comparison between the two approaches, based on F1-Score metric.

	Progressive (%)	ImageNet (%)
AlexNet	94,36	89,53
VGG-16	94,75	93,87
ResNet	94,76	94,55
DenseNet	94,64	92,15
InceptionV3	93,23	95,97

#### 4.6. Approach 3 : AVRDIS Combined model

Approach 3 consists of combining the 5 models. This makes it possible to assess the presence of the complementarities of different models in the pneumonia classification. The following table (Table 8 and Figure 6), shows a comparison of the results of approach 3 (called AVRDIS, See Figure 2) and the results of each of the 5 models (with its best configuration). This comparison showed the superiority of the approach based on the combination of models. This confirms the presence of complementarities between the 5 models. This combination gave the best performance and achieved a test accuracy of 96.79%, and an F1-score equal to 97.44%. Table 8 shows the results for each model as well as for the proposed method.

**Table 8**

Comparison of the different architectures with the AVRDIS.

	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
VGG-16	93,42	94,72	94,87	94,75
ResNet	93,58	96,76	92,86	94,76
DenseNet	93,10	92,00	97,43	94,64
InceptionV3	94,87	94,08	97,94	95,97
AlexNet	92,87	92,17	99,66	94,36
AVRDIS	96,15	95,52	98,46	96,96

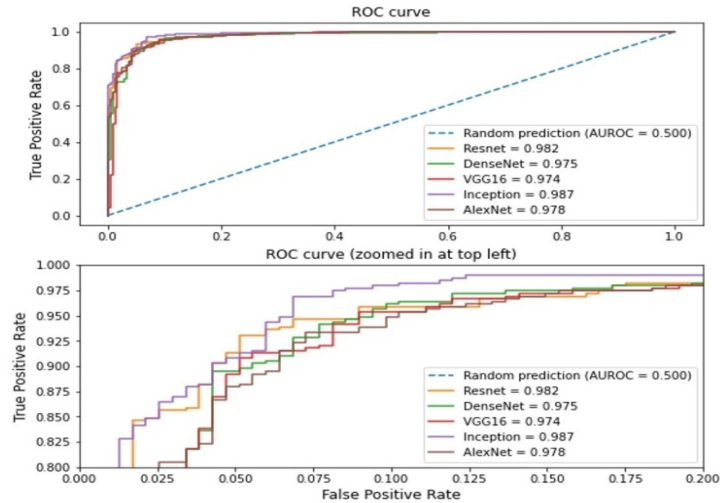


Figure 6: ROC curves obtained via the five models, and the zoom on the upper part.

#### 4.7. Comparison of the AVRDIS approach to existing approaches in the literature

Table 9 gives us a comparison of the combined model (AVRDIS) with some existing methods using the metrics mentioned above. It confirms the superiority of the proposed approach. This final model is the most relevant to apply it to our problem with an accuracy of 96.76%, a sensitivity of 97.47% and an F1-score of 97.44%.

Table 9

Comparison of the different existed architectures with the proposed approach AVRDIS.

	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
Ayan and al. [24]	87,00	91,00	94,00	84,00
Polat and al. [32]	92	-	-	-
Alqudah and al.[30]	94,03	94,22	96,68	-
Vikash and al.[26]	96	93,28	99,62	-
Proposed method (AVRDIS)	<b>96,15</b>	<b>95,52</b>	<b>98,46</b>	<b>96,96</b>

## 5. Conclusion

This paper deals with the problem of the classification of pneumonia on radiographic images, since they are considered the main cause of infant mortality in the world. The goal was to develop a robust and automatic approach, based on deep learning. In this context, we adopted the transfer learning approach and used a combination of five pre-trained architectures: AlexNet, VGG-16, ResNet, DenseNet121 and InceptionV3, initially trained on the ImageNet dataset which consists of 14 million natural images. Our experiments revealed that the scenario of progressive transfer learning from the tuberculosis image dataset to classify pneumonia improves the performance of the studied models. In addition, a comparative study of supervised softmax classifier with different classifiers showed that these latter offered better results thanks to the power of RF, SVM and Naive Bayes classifiers combined with CNN models. Finally, we used a model combining five pretrained CNN models. The

latter outperformed all other models with a good performance in terms of F1-score (97.44%) and accuracy (96.76%). Although many methods have been developed to work on this dataset, the proposed methodology has obtained better results. We observed that performance could be further improved by increasing the size of the databases using a data augmentation approach and by using other more advanced models, such as Vision Transformers (ViT) without or with combination with a CNN. The coupling with a metaheuristic approach (like PSO) would allow better piloting and optimization of the combination of classifiers. This will pave the way for our future works to explore this line of research.

## 6. References

- [1] A. Ozcift, Enhanced Cancer Recognition System Based on Random Forests Feature Elimination Algorithm, *Journal of medical systems*, 2011, Vol. 36, Num 4, pp. 2577-2585
- [2] C. Nguyen, Y. Wang, H.N. Nguyen; Random Forest classifier combined with feature selection for breast cancer diagnosis and prognostic. 2013, *journal of Biomedical Science and Engineering*. Vol 06, pp. 551-560.
- [3] I. El-Naqa, Y. Yang, M. Wernick, N. Galatsanos, R. Nishikawa, A support vector machine approach for detection of micro calcifications, *IEEE Transactions on Medical Imaging*. Vol. 21, pp. 1552-1563.
- [4] S. Ghumbre, C. Patil, A. Ghatol, Heart Disease Diagnosis using Support Vector Machine. 2011.
- [5] V. Chaurasia and S. Pal, A Novel Approach for Breast Cancer Detection using Data Mining Techniques. 2017, *International Journal of Innovative Research in Computer and Communication Engineering*. Vol. 2, issue 1.
- [6] A. Christobel et Y. Sivaprakasa, An Empirical Comparison of Data Mining Classification Methods. 2011, *journal, international journal of computer information systems* Vol. 3, n° 2.
- [7] J. Yao, A. Dwyer, R. Summers; Computer-aided diagnosis of pulmonary infections using texture analysis and support vector machine classification. 2011, *journal of Academic Radiology*. Vol. 18, pp 306-314.
- [8] E. Naydenova, A. Tsanas, C. Casals-Pascual, M. De Vos, Smart diagnostic algorithms for automated detection of childhood pneumonia in resource-constrained settings. 2015, *IEEE Global humanitarian Technology Conference*. pp. 377–384.
- [9] T. B. Chandra and k. Verma, Pneumonia Detection on Chest Xray Using Machine Learning Paradigm. 2020, *Actes de la 3e conference internationale sur la vision par ordinateur et le traitement d'images*. p.2133
- [10] K.M. Kuo, P.C. Talley, C.H Huang, L.C Cheng, Predicting Hospital-acquired pneumonia among schizophrenic patients: a machine learning approach. 2019, *PMC article*, PMID : 30866913, DOI : 10.1186/s12911-019-0792-1
- [11] H. Yue, Q. Yu, C. Liu, Y. Huang, Z. Jiang, C. Shao, Machine learning-based CT radiomics method for predicting hospital stay in patients with pneumonia associated with SARS-CoV-2 infection: a multicenter study. 2020, *PMC article*, PMID : 32793703 DOI : 10.21037/atm-20-3026
- [12] R.T. Sousa, O. Marques, F. Alphonsus, A.M.N. Soares, I.I.G. Sene, L.L.G. de Oliveira, E.S. Spoto; Comparative performance analysis of machine learning classifiers in detection of childhood pneumonia using chest radiographs. 2013, *journal of Procedia Computer Science*. Vol. 18, pp. 2579–2582.
- [13] M. Jinlian, F. Wu, J. Zhu, D. Xu and D. Kong, A pretrained convolutional neural network-based method for thyroid nodule diagnosis. 2016, *Ultra-sonics*, pp. 221-230
- [14] V. Gulshan, L. Peng, M. Coram, M. Stumpe, D. Wu, Aru-nachalam Narayanaswamy, Subhachini Venu-gopalan; Development and validation of a deeplearning algorithm for detection of diabetic retinopathy in retinal fundus photographs. 2016, *article*. DOI : <http://jamanetwork.com/article.aspx?doi=10.1001/jama.2016.17216>.
- [15] M. Haloi, Improved Microaneurysm Detection using Deep Neural Networks. 2015, *article*. DOI : <https://doi.org/10.48550/arXiv.1505.04424>

- [16] A. Rakhlin, A. Shvets, V. Iglovikov, A. Kalinin, Deep convolutional neural networks for breast cancer histology image analysis. 2018, International Conference Image Analysis and Recognition, pp. 737–744.
- [17] Y. S. Vang, Z. Chen, X. Xie, Deep learning framework for mul-ti-class breast cancer histology image classification. 2018, International Conference Image Analysis and Recognition, pp. 914–922.
- [18] O. Er, N. Yumusk, F. Temurtas; Chest diseases diagnosis using artificial neural networks. 2010, journal of Expert Systems with Applications, pp. 7648-7655.
- [19] S. Rajaraman, S. Candemir, I. Kim, G. Thoma, S. Antani, Visualization and interpretation of convolutional neural network predictions in detecting pneumonia in pediatric chest radiographs. 2018, article. Doi : <https://doi.org/10.3390/app8101715>
- [20] X. Gu, L. Pan, H. Liang, R. Yang, Classification of Bacte-rial and Viral Childhood Pneumonia Using Deep Learning in 71 Chest Radiography. 2018, In Proceedings of the 3rd international conference on Mobile and image Processing, pp. 88–93.
- [21] A.H. Alharbi, H.A. Hosni Mahmoud, Pneumonia Transfer Learning Deep Learning Model from Segmented X-rays. *Healthcare* 2022, 10, 987. <https://doi.org/10.3390/healthcare10060987>
- [22] P. Rajpurkar, J. Irvin, R.L. Ball, K. Zhu, B. Yang, H. Mehta, T. Duan, D. Ding, A. Bagul, C. P. Langlotz, Deep learning for chest radiograph diagnosis: A retrospective comparison of the CheXNeXt algorithm to practicing radiologists. 2018, PMC article, PMID: 30457988, DOI: <https://doi.org/10.1371/journal.pmed.1002686>.
- [23] A. K. Jaiswal, P.T.S. Kumar, D. Gupta, A. Khanna, J. Rodrigues; Identifying pneumonia in chest X-rays: A deep learning approach. 2019, *Journal of Measurement*. Vol. 145, pp 511-518.
- [24] E. Ayan and H. M. Unver, Diagnosis of Pneumonia from Chest X- Ray Images Using Deep Learning. 2019, Conference, Scientific Meeting on Electrical-Electronics & Biomedical Engineering and Computer Science (EBBT). pp. 1–5.
- [25] D. Varshni, K. Thakral, L. Agarwal, R. Nijhawan, A. Mittal, Pneumonia Detection Using CNN based Feature Extraction. 2019, International Conference on Electronics, Communication and Computing Technologies (ICECCT), pp. 1-7.
- [26] V. Chouhan, S.K. Singh, A. Khamparia, D. Gupta, P. Tiwari, C. Moreira; A Novel Transfer Learning Based Approach for Pneumonia Detection in Chest X-ray Images. 2020, *journal of Applied Sciences*. Vol. 10, pp 559.
- [27] T. Rahman, M. E. H. Chowdhury, A. Khandakar, R. Islam, Z. B. Mahboub, Transfer Learning with Deep Convolutional Neural Network (CNN) for Pneumonia Detection Using Chest X-ray. 2020, article. Doi : <https://doi.org/10.3390/app10093233>
- [28] M. F. Hashmi, S. Katiyar, A.G. Keskar, N.D. Bokde, Z.W. Geem; Efficient Pneumonia Detection in Chest Xray Images Using Deep Transfer Learning. 2020, PMC article, PMID: 32575475, DOI: <https://doi.org/10.3390%2Fdiagnostics10060417>
- [29] Y. Brima, M. Atemkeng, S. Tankio Djiokep, J. Ebiele, F. Tchakounté, Transfer Learning for the Detection and Diagnosis of Types of Pneumonia including Pneumonia Induced by COVID-19 from Chest X-ray Images. *Diagnostics* 2021, 11, 1480. <https://doi.org/10.3390/diagnostics11081480>
- [30] A.M. Alqudah, S. Qazan, I.S. Masad, Artificial Intelligence Framework for Efficient Detection and Classification of Pneumonia Using Chest Radiography Images. 2021, *Journal of Medical and biological Engineering*, pp. 599–609.  
A. Manickam, J. Jiang, Y. Zhou, A. Sagar, R. Soundrapandiyam, R.D. Samuel, (2021). Automated pneumonia detection on chest X-ray images: A deep learning approach with different optimizers and transfer learning architectures. <https://doi.org/10.1016/j.measurement.2021.109953>
- [31] O. Polat, Z. Dokur, T. Olmez, Determination of Pneumonia in X-ray Chest Images by Using Convolutional Neural Network. 2021, *journal of Electrical Engineering and Computer Sciences*, pp. 16151627.  
D. Avola, A. Bacciu, L. Cinque, A. Fagioli, M. R. Marini, R. Taiello. Study on transfer learning capabilities for pneumonia classification in chest-x-rays images. *Comput Methods Programs Biomed.* 2022 Jun;221:106833. doi: 10.1016/j.cmpb.2022.106833. Epub 2022 Apr 22. PMID: 35537296; PMCID: PMC9033299.