Fast COVID-19 and Pneumonia Classification Using Chest X-ray Images

Juan Eduardo Luján-García 1,*, Marco Antonio Moreno-Ibarra 1,*, Yenny Villuendas-Rey 2,*, and Cornelio Yáñez-Márquez 1,*

1 Centro de Investigación en Computación, Instituto Politécnico Nacional, Mexico City 07700, Mexico; jeduardolujan5@gmail.com
2 Centro de Innovación y Desarrollo Tecnológico en Cómputo, Instituto Politécnico Nacional, Mexico City 07700, Mexico
* Correspondence: mmorenoi@ipn.mx (M.A.M.-I.); yvilluendasr@ipn.mx (Y.V.-R.); coryanez@gmail.com (C.Y.-M.)

Received: 18 July 2020; Accepted: 21 August 2020; Published: 26 August 2020

Abstract: As of the end of 2019, the world suffered from a disease caused by the SARS-CoV-2 virus, which has become the pandemic COVID-19. This aggressive disease deteriorates the human respiratory system. Patients with COVID-19 can develop symptoms that belong to the common flu, pneumonia, and other respiratory diseases in the first four to ten days after they have been infected. As a result, it can cause misdiagnosis between patients with COVID-19 and typical pneumonia. Some deep-learning techniques can help physicians to obtain an effective pre-diagnosis. The content of this article consists of a deep-learning model, specifically a convolutional neural network with pre-trained weights, which allows us to use transfer learning to obtain new retrained models to classify COVID-19, pneumonia, and healthy patients. One of the main findings of this article is that the following relevant result was obtained in the dataset that we used for the experiments: all the patients infected with SARS-CoV-2 and all the patients infected with pneumonia were correctly classified. These results allow us to conclude that the proposed method in this article may be useful to help physicians decide the diagnoses related to COVID-19 and typical pneumonia.

Keywords: COVID-19; pneumonia; classification; deep learning; convolutional; network

1. Introduction

COVID-19 is a newly known disease that can be misdiagnosed as common pneumonia. In 2020, COVID-19 has become a major pandemic due to the easy propagation though the air and contact with contaminated objects and people. According with the World Health Organization (WHO), as of 30 June 2020, there have been more than 10 million cases of COVID-19 and more than half a million confirmed deaths [1]. COVID-19 is caused by the SARS-CoV-2 virus infecting the lung and the respiratory system. This is a major concern for the medical field due to the fuzzy symptoms that present early contagious people. Patients with COVID-19 can develop symptoms that belong to common flu, pneumonia, and other respiratory diseases in the first four to 10 days since they were infected [2].

It is possible to support the diagnosis of respiratory and lung diseases through computer-assisted diagnosis (CAD). Within CAD, we can find techniques that obtain images from the internals of the body, such as chest X-ray (CXR) images.

The use of computed tomography (CT) as well as magnetic resonance imaging (MRI) is also useful. Nowadays, the most effective image examination for diagnosing COVID-19 on infected patients is the CT images, due to their high sensitivity compared to CXRs. Nonetheless, only patients with severe complications are hospitalized, and CT Scan Units are limited in hospitals. On the other hand,
CXRs can also provide useful images to help to visualize COVID-19-infected patients. Radiologists have found radiological features that could be useful for screening COVID-19 in CXR images [3]. However, these features can be confused with atypical pneumonia and other pulmonary manifestations.

Even though CXRs are not the most accurate examination, CXR images can be widely used because they are not as expensive as CT units and do not require extensive patient preparation to perform an examination. Furthermore, there are portable X-ray units that can be moved to non-critical hospital areas, even to specific facilities to analyze suspicious COVID-19 cases.

According to the Radiology Society of North America (RSNA), there are some specific manifestations that can be found in CT scans, such as peripheral and bilateral consolidations with “crazy paving” pattern and peripheral and bilateral ground-glass opacification (GGO) signs [4].

Moreover, radiological features can be observed in CXRs from patients five days before acquiring COVID-19. The findings on CXRs are airspace opacities, such as consolidations or GGO, both unilateral or bilateral [5]. Two examples of CXR images are shown in Figure 1: a healthy patient and a person infected with COVID-19.

![Figure 1. Examples of CXR images: (a) shows a healthy patient without any anomaly within the thorax; (b) shows a COVID-19-infected patient, bilateral opacities are found. Original images were obtained from (https://github.com/ieee8023/covid-chestxray-dataset), a public COVID-19 dataset. Annotations were manually performed by us with a doctor assistant.](image)

On the other hand, pneumonia is an infectious disease that also affects the lungs, and according to WHO, it is considered a leading cause of death in children [6]. Pneumonia can be caused by fungus, bacteria, or a virus attack.

Pneumonia causes pain in the chest and limits the oxygen intake of the infected patient. In the same way, pneumonia presents radiological features, such as consolidations by fluid accumulation [6]. Two examples of CXR images are shown in Figure 2: a healthy patient and a person with pneumonia caused by bacteria.

There are several machine-learning (ML) techniques that can help develop CAD tools. One of the best current methods for computer vision (CV) tasks are the convolutional neural networks (CNN) [7]. A lot of techniques have been used in the medical field to assist CAD tasks, such as lesion segmentation [8], brain tumor segmentation [9,10], automatic size calculation of the heart [11], and classification among several thorax diseases [12–15].

In this research, we propose to use the Xception [16] CNN with pretrained weights on ImageNet. The use of pretrained models allows us to use transfer learning (TL) to obtain new retrained models to classify, in this case, among COVID-19, pneumonia, and healthy patients.
2. Materials and Methods

CNN have proved to be one of the best techniques to solve CV problems. Even since a CNN won the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) in 2012 [17], a lot of new models for CV tasks have been developed and established new milestones on the ILSVRC.

Nowadays, we can find general robust models like VGG-16 [18], Inception-V3 [19], ResNet [20], Xception [16], and DenseNet [21], which are considered as baselines for the use of CNN in many tasks.

2.1. Convolutional Neural Networks and Chest Diseases

The medical field and CAD have presented big advances thanks to deep learning (DL) and CNN. Moreover, TL has been widely used in order to reuse pre-trained models on ImageNet [22].

Although, CAD applications have been actively published after the release of the Chest-Xray14 dataset (CX14) in 2017, by Wang et al. [12]. The CX14 contains posterior–anterior images from the chest, with a total of 108,948 images. The CX14 represented a big milestone for CAD, CV, and DL applications for multiple chest diseases classification [13,23,24].

Other CAD applications have been developed for pneumothorax detection [25], cardiomegaly detection [26], and synthetic CXRs creation for CNN training [27].

2.2. Convolutional Neural Networks, Pneumonia, and COVID-19

Pneumonia classification has also been aborded by multiple research groups. For example, Rajpurkar et al. used a DenseNet model in CX14 to achieve radiologist-level diagnosis of pneumonia [28].

Kermany et al. [29], presented a pneumonia dataset with only 5232 images and established a baseline result for it. Additionally, on Kermany’s dataset, multiple works had been presented; the three latest top results are by Lian and Zheng [28], Luján et al. [30], and Chouhan et al. [31]. On the RSNA Challenge [32], the top result was presented by Sirazitdinov et al. [33].

By July 2020, a large number of scientific papers related to the COVID-19 were published. Most of the investigations used CT images in conjunction with CNN models to classify sick patients and try to differentiate them from typical pneumonia. Nonetheless, the main problem was the lack of public datasets that would allow us to perform diverse experiments with different techniques. Furthermore, the small number of datasets that contain images from patients with COVID-19 lack a great number of samples. The last statement represents a big challenge to many of the state-of-the-art classification
models that are related with DL and other ML algorithms that need several examples to work effectively. Therefore, is a great challenge to present alternatives to classify and screen new COVID-19, even using traditional techniques.

One of the first attempts to classify COVID-19 patients was presented by El-Din et al. [34] in which they compared multiple CNN baselines to select the best model for classification. Butt et al. [35] presented a ResNet variant to classify and localize the patterns from the SARS-CoV-2 virus in COVID-19. Ardakani et al. [36] implemented and compared 10 models of the best-known CNN models to classify CT images, obtaining outstanding results compared with medical radiologists.

A similar approach to the proposed work was presented by Ozturk et al. [37] in which they used the Darknet model with pretrained weights from the YOLO system [38], achieving the best results using CXR images until the publication of their article. Other approaches presented by Zhanh et al. [39] include the use of a CNN as feature extractor and two different multilayer perceptron (MLP) as anomaly detection and confidence prediction; a DenseNet121 model presented by Cohen et al. [40] established a baseline for an early version of the same dataset used in this research.

The proposed method aims to use the Xception pre-trained network on ImageNet and combining two datasets from both pneumonia and COVID-19 to classify among healthy and sick patients. We use preprocessing techniques and cost-sensitive learning (CSL), as in previous research to avoid the undesired effect produced by data imbalance problems.

3. Materials and Methods

In this section, we describe the datasets we have used to carry out these research experiments. A class imbalance problem is presented, and both classification algorithms and screening techniques are presented. Finally, performance measures are described.

3.1. Pneumonia Dataset

The pneumonia dataset selected to perform the experiments in this article is the one published in 2018 by the team of Kermany et al. [29]. This dataset contains 5856 CXR images of children up to five years old (https://data.mendeley.com/datasets/rscbbr9sj/2).

The total set of images is made up of two disjoint sets: a training set with 5232 images and a test set containing 624 images. The training set includes 3883 images of pneumonia-infected patients and 1349 images of healthy patients. Otherwise, the test set contains 390 images of pneumonia-infected patients and 234 images of healthy patients.

3.2. COVID-19 Dataset

Due to the recent outbreak of COVID-19, there are not too many publicly available datasets that allow researchers to try different classification and segmentation methods for this new disease. Nonetheless, the Montreal University has opened a regularly updated dataset for COVID-19 and other respiratory diseases to contribute to diagnosis tools.

The COVID-19 Image Data Collection, presented by Cohen et al. [41], contained 539 images from different respiratory diseases, such as pneumonia, SARS, MERS, among others (until before 1 June 2020).

In this dataset, only 287 CXRs were specifically from patients with COVID-19. Most of them are from adult patients and include people from 12 to 87 years old and different nationalities. Only COVID-19 images were used in this research from this dataset (https://github.com/ieee8023/covid-chestxray-dataset).

3.3. Class Imbalance

Considering the two datasets, we have used the images of each one to get a new dataset in which we can set three different classes: COVID-19, HEALTHY, and PNEUMONIA.
Moreover, we can also observe that the number of examples of each class are not the same. Therefore, we can know how imbalanced our new dataset is. We computed the imbalance ratio (IR) as in Equation (1):

\[
IR = \frac{\text{majority class}}{\text{minority class}}.
\]

If \( IR > 1.5 \), we can consider a dataset imbalanced [42]. Then, the IR for our dataset is computed in Equation (2):

\[
IR = \frac{|\text{PNEUMONIA}|}{|\text{COVID − 19}|} = \frac{3883}{287} = 13.52.
\]

3.4. Convolutional Network

François Chollet published, in 2016, one of the most effective CNN models known to date: it is the Xception model [16], which we have chosen as the basis for the experimental section of this article.

This CNN is structured by 36 convolutional layers, and its main difference is the use of depth-wise separable convolutions (DWSC) and residual connections just as in ResNet models. DWSC obtain the same results as traditional convolutions but perform fewer operations when using large filter sizes.

Outperforming in operations models, such as VGG, Inception, and ResNet, deeper configurations can be used to extract features from the images.

The original Xception model contained a Softmax layer of 1000 neurons and a global average as a pooling layer. We adapted the Xception model for three class classification, as shown in Figure 3.

\[
\sigma(z)_i = \frac{e^{z_i}}{\sum_{k=1}^{k} e^{z_k}}. \tag{3}
\]

in which \( z = x^T w \) for \( i = 1, 2, \ldots, k \) and \( z = (z_1, z_2, \ldots, z_k) \in \mathbb{R}^k \).
3.5. Screening and Localization

We have implemented the Gradient-Weighted Class Activation Mapping (Grad-CAM) [44] algorithm in order to visualize the activation of the last convolutional layer of the Xception network. The last convolutional layer is the one that provides the final features’ values for the logistic layer to compute a probabilistic output. The gradients of this layer are used to generate the Grad-CAM of an input image. Therefore, Grad-CAM provides a coarse localization map that indicates the most important regions in the image (radiological features) for a specific concept—in this case, COVID-19 or PNEUMONIA class.

3.6. Performance Measures

The confusion matrix allows us to compute metrics of the performance of our classification task. In a binary classification problem, the confusion matrix includes [45]:
- true positives (tp)
- true negatives (tn)
- false positives (fp), and
- false negatives (fn).

Figure 4 represents an example of confusion matrix.

Figure 4. Confusion matrix example.

On the other hand, in a multiclass problem, a common solution is to evaluate each class separately. That is, each class needs to be evaluated considering the tp and fn of the class of interest and the tn and fp from the other classes.

In this research, when evaluating a single class like COVID-19, the tn and fp are the sum of the instances of HEALTHY and PNEUMONIA classes.

Although there are a large number of performance measures in pattern classification problems, one of the most used is accuracy, which is defined in Equation (4) for multiclass problems:

\[
Average\ Accuracy = \frac{\sum_{i=1}^{l} \frac{tp_i + tn_i}{tp_i + fp_i + fn_i + tn_i}}{l},
\]

where, \(l\) is the total of classes. Individual Accuracy is computed per class without dividing the result by \(l\).

On the contrary, when we have imbalanced datasets, it is recommended to use different performance measures that are not biased to the general result of the classification algorithm. Some of these performance measures include sensitivity (also called recall) and precision, which are derived directly from the confusion matrix, in addition to the F1-Score, which is defined from recall and precision. The quality of the classification can be assessed as either macro or micro average. In general,
macro average treats classes as equal [45]. Therefore, both individual and macro-average measures will be used in this research. Precision, recall, and F-1 Score are detailed in Equations (5)–(7), respectively.

\[
\text{Precision}_M = \frac{\sum_{i=1}^{l} \frac{tp_i}{tp_i + fp_i}}{l}
\]  
\[
\text{Recall}_M = \frac{\sum_{i=1}^{l} \frac{tp_i}{tp_i + fn_i}}{l}
\]  
\[
F_1\text{-Score}_M = 2 \frac{\text{Precision}_M \text{Recall}_M}{\text{Precision}_M + \text{Recall}_M}
\]

Individual measures are computed similar to the last equations but without dividing the values by \(l\). In addition, it is common to analyze the classification results with graphics tools, such as the reception operating characteristic curve (ROC curve) and its general score—the area under the curve (AUC). Therefore, in Section 5, we will present the performance measures of our classification results and the ROC curve associated with them.

4. Proposed Method

4.1. Image Preprocessing

It is common that CNN uses square images as input. Our CXRs from the two different datasets come in a variety of sizes that depend on the original equipment and preprocessing of the X-ray machines. Consequently, in order to feed the network, we would need to resize the images to a square form, causing a distortion on the images, as shown in Figure 5.

![Figure 5. Distortion due to resizing the images to square shape. (a) Shows the original shape of a CXR of a patient with COVID-19; (b) shows the resized image to squared shape. Original images are from (https://github.com/ieee8023/covid-chestxray-dataset).](image)

We wanted to avoid distortion on the input images in order to evade the suppression of useful data of the images. As a result, we have used a preprocessing technique presented by Pasa et al. [46] to extract the central region and eliminating black bars. Figure 6 shows a preprocessed image to which the following operations have been performed:

1. If some black band appears at the edges, they are removed.
2. The size of the image is transformed until the smallest border measures 299 pixels.
3. Extract the central region of 299 × 299 pixels.
With previous experience, we follow the next hyperparameters (Table 2), obtaining the current presented trained model. We also implemented early stopping to prevent the network from becoming overfitting.

At last, normalization was performed over all the sets of images. We transformed the distribution of the dataset to a normal distribution with media $\mu = 0$ and standard deviation $\sigma = 1$. We achieved normalization by computing the mean and standard deviation values from the training set partition, and then, adding the mean and dividing training, validation, and test sets by the computed standard deviation. Partitions will be explained in Section 4.4.

### 4.2. Cost Sensitive Learning

We can apply a penalty to the function that scores each class in a classification task, this process is known as cost-sensitive learning (CSL) [47]. In this research, the cost function is categorical cross-entropy [48]. CSL helps avoid bias in classification when using CNNs. Therefore, CSL helps us as a method for solving imbalance problem, presented in Section 3.3. In this case, weights were obtained considering the number of training examples of each class, using a Sci-Kit Learn [49] function that obtains the numbers to balance the number of examples based on the principles of logistic regression. Table 1 includes the weights.

**Table 1.** Matrix of weights to penalize cost function.

| Class    | Weight |
|----------|--------|
| COVID-19 | 6.42   |
| HEALTHY  | 1.36   |
| PNEUMONIA| 0.47   |

### 4.3. Data Augmentation and Hyperparameter Tuning

Deep-learning models, such as CNN, often present better results when used as much images, if possible. As a result, data augmentation is useful to generate more data based on the original training images. We performed the following data augmentation operations at training time:

- Random rotation of ±10 degrees.
- Zoom on a range of ±10%.
- Horizontal flipping.

On the other hand, hyperparameters used for training a neural network are not obtained by a specific rule [50]. Moreover, it is necessary to adjust the parameters considering the validation data. With previous experience, we follow the next hyperparameters (Table 2), obtaining the current

---

![Figure 6. Preprocessing applied to all images of the datasets. (a) Shows the original CXR from a patient with COVID-19 (https://github.com/ieee8023/covid-chestxray-dataset); (b) shows the preprocessed image to a squared shape. Original images are from (https://github.com/ieee8023/covid-chestxray-dataset).](image-url)
presented trained model. We also implemented early stopping to prevent the network from overfitting; we stopped the training when the score of the loss function did not improve after 10 epochs.

Table 2. Hyperparameters of the network.

| Cost Function                  | Learning Rate (LR) | Optimizer | Epochs | Batch Size | LR Decay       |
|-------------------------------|--------------------|-----------|--------|------------|----------------|
| Categorical cross entropy     | $1 \times 10^{-3}$ | Adam      | 100    | 32         | 10 times after a plateau |
|                               | $\beta_1 = 0.9$   |           |        |            |                |
|                               | $\beta_2 = 0.999$ |           |        |            |                |

The categorical cross-entropy is a generalization of the cross-entropy loss function. It is used when we have a multiclass classification task. The loss is computed for each class independently, and then, the results are summed. Categorical cross entropy is defined in following Equation (8) [48].

$$H(p, q) = -\sum_k p_k \log(q_k). \quad (8)$$

where, $k$ is the number of the class, and $q$ is the Softmax function or predicted probability of class $k$ (Equation (8)). Moreover, if we apply the weights for CSL, presented in the last subsection, we will have Equation (9).

$$H_{csl}(p, q) = -\sum_k W_k p_k \log(q_k). \quad (9)$$

where $W_k$ are the weights for COVID-19, HEALTHY, and PNEUMONIA classes, respectively as shown in Table 1.

4.4. Dataset Partition

A simple validation strategy was followed to obtain training, validation, and test sets. First, 32 randomly selected examples where selected from the COVID-19 class, which represents the 10% of the complete data. Then, with the left images, a hold-out 80–20 validation method was applied. Similarly, with the original training images of HEALTHY and PNEUMONIA patients, hold-out 80–20 was performed. Although, test examples for both HEALTHY and PNEUMONIA were the official ones, presented by Kermany et al. [51] in the original dataset. Final partitions are summarized in Table 3.

Table 3. Number of examples per partition.

| Partition | COVID-19 | HEALTHY | PNEUMONIA |
|-----------|----------|---------|-----------|
| Training  | 229      | 1079    | 3106      |
| Validation| 58       | 270     | 777       |
| Test      | 32 $^1$  | 234     | 390       |

$^1$ Randomly selected before hold-out for training and validation sets.

5. Results

In this section, we preset the general performed methodology for the experiments. Performance on validation set is presented. Finally, classification performance and screening of the diseases are shown.

5.1. Experimental Framework

All experiments were implemented using Python 3.6 programming language, using Jupyter Notebook and OpenCV [52] as our main graphic processor software. We preprocessed all the available images and, then, performed hold-out validation to obtain the partitions presented in Table 3. We have used a fixed seed for replication purposes of the algorithm. Hold-out validation was only performed once, without changes on the partitions. In carrying out the research whose results are presented in this article, we have taken advantage of the free online Linux platform called Google Collaboratory.
Therefore, at the time of conducting the experiments, we had 25 GB of RAM, in addition to a Nvidia Tesla K80 model GPU that included 12 GDDR VRAM.

Summary of the proposed methodology is as follows:

- Resizing and cropping of all images with the proposed method.
- Hold-out as validation method to obtain training, validation, and test sets.
- Normalization of the images.
- Model selection by Xception training and validation.
- Performance evaluation of the model on the test set.
- Grad-CAM generation for test examples.

5.2. Validation Set Results

We trained the Xception network with pretrained weights on ImageNet and, then, selected the best model using the score of the loss function over the validation set. Moreover, we also computed the Accuracy, Precision, Recall, and ROC AUC over the validation set. Last scores are shown in Figure 7.

![Figure 7. Metrics for validation set on training time: (a) Shows the value of the loss function; (b) shows the accuracy, precision, recall, and ROC AUC macro-averaged scores.](image)

In addition, we also compared the proposed method with some of the most important CNN baselines, such as VGG16, ResNet50, and DenseNet121. Figure 8 shows the score of the loss function for each model.

![Figure 8. Metrics for validation set on training time for all CNN models.](image)

Furthermore, we measured the training time per epoch and example. Our model obtained the best loss on the validation set on the sixth epoch, but it continued training until the epoch 16. Table 4 shows the time measures and the validation loss score for all CNN models. It was found that the proposed method took less than 12 min to obtain a competitive model to classify COVID-19.
Table 4. Time measurement of network training.

| Model       | Best Epoch | Validation Loss | Average Training Time (Epoch) | Average Training Time (Example) | Convergence Time (Best Model) | Training Total Time |
|-------------|------------|-----------------|-------------------------------|---------------------------------|-----------------------------|---------------------|
| VGG16       | 34         | 0.79751         | 80 s                          | 0.0181 s                        | 2720 s                      | 3520 s              |
| ResNet50    | 7          | 0.19316         | 78 s                          | 0.0177 s                        | 546 s                       | 1248 s              |
| DenseNet121 | 8          | 0.08312         | 74 s                          | 0.0168 s                        | 592 s                       | 1258 s              |
| Proposed method | 6       | 0.05619         | 100 s                         | 0.0226 s                        | 700 s                       | 1600 s              |

Bold numbers represent the best results among all the following tables.

5.3. Test Set Classification Results

Our best model, the one with the lowest validation loss score, was selected to classify our test set. Results of classification are shown in the confusion matrix of Figure 9.

![Confusion Matrix](image1)

**Figure 9.** Results obtained for the classification of the test set using the proposed model. (a) Shows the confusion matrix of the three classes; (b) shows ROC curves scores.

In Figure 9, we can observe the values of each individual CLASS. In the same way, ROC curves and their AUC were computed. Individual and macro-average (MA) metrics are condensed in Table 5.
Table 5. Performance measures of the classification task with the proposed model.

| Class       | Accuracy | Precision | Recall | F1-Score | ROC Curve AUC |
|-------------|----------|-----------|--------|----------|---------------|
| COVID-19    | 1.00     | 0.94      | 1.00   | 0.97     | 1.00          |
| HEALTHY     | 0.86     | 0.99      | 0.62   | 0.76     | 0.97          |
| PNEUMONIA   | 0.86     | 0.82      | 1.00   | 0.90     | 0.97          |
| Macro-averaged | 0.91  | 0.92      | 0.87   | 0.88     | 0.98          |

1 Average accuracy (also known as balanced accuracy) of the model.

Moreover, we also computed the macro-averaged metrics and plotted the ROC curves for each CNN model. Metrics are presented in Table 6, and ROC curves are shown in Figure 10.

Table 6. Macro-average measures for all CNN models.

| Model        | Average Accuracy | MA Precision | MA Recall | MA F1-Score | MA ROC Curve AUC |
|--------------|------------------|--------------|-----------|-------------|------------------|
| VGG16        | 0.46             | 0.60         | 0.53      | 0.30        | 0.81             |
| ResNet50     | 0.88             | 0.77         | 0.85      | 0.79        | 0.95             |
| DenseNet121  | 0.88             | 0.83         | 0.85      | 0.80        | 0.97             |
| Proposed method | 0.91   | 0.92         | 0.87      | 0.88        | 0.98             |

Figure 10. Macro-averaged ROC curves scores for all CNN models.

Finally, we applied t-distributed Stochastic Neighbor Embedding (t-SNE) [53] using Sci-Kit Learn [49] to represent the high-dimensional final features vector of the Xception network, which contains 2048 features for each example. Figure 11 shows the prediction of the test set in a two-dimensional plane.

(a) (b)

Figure 11. t-SNE of predicted outputs for test set: (a) shows the instances using random initialization; (b) shows the instances using principal component analysis (PCA) for initialization.
5.4. Disease Screening

As mentioned in Section 3.5, Grad-CAM was used to provide the possible localization of the manifestations of both COVID-19 and PNEUMONIA. The maps are showed as a heatmap, where the intense red color represents the most important area (extracted features) from which the classification decision is taken by the network. Figures 12 and 13 show several examples of correctly classified images of both COVID-19 and PNEUMONIA classes, respectively.

![Figure 11](image1.png)
![Figure 11](image2.png)

Figure 11. t-SNE of predicted outputs for test set: (a) shows the instances using random initialization; (b) shows the instances using principal component analysis (PCA) for initialization.

![Figure 12](image3.png)
![Figure 12](image4.png)

Figure 12. Grad-CAM of some COVID-19 test images: (a,c,e) show the original and preprocessed images; (b,d,f) show their corresponding heatmaps.

**Figure 12.** Grad-CAM of some COVID-19 test images: (a,c,e) show the original and preprocessed images; (b,d,f) show their corresponding heatmaps.
Figure 13. Grad-CAM of some pneumonia test images: (a,c,e) show the original and preprocessed images; (b,d,f) show their corresponding heatmaps.

We did not perform further comparison of localization of the diseases with other state-of-the-art papers due to the daily updates of the dataset. Therefore, the number of examples on different dates would not be the same.

6. Discussion

In this section, we highlight the advantages of the proposed method and evaluate the performance of both the classification algorithm and the visualization technique.

On the validation set (Table 4), the proposed method obtained the minor loss score of 0.05619 among all CNN models, belonging to the sixth epoch of the training phase. Therefore, our method
trained over all the dataset and obtained the best model with an average of 700 s. This means that we were able to obtain a good model for the classification of pneumonia, healthy, and COVID-19 patients in less than 12 min. In addition, the total training time of the proposed model, took under 27 min, thanks to the condition of early stopping. On the other hand, even though DenseNet121 finished the training before our method, it did not obtain the best classification result. Moreover, the differences on performance are of less than six minutes of difference on training a full CNN model for COVID-19 classification. Therefore, experiments with the hyperparameters presented allow us to generate up to two solutions to the classification of CXRs in one hour. Furthermore, the time taken on the proposed method in processing only one example (inference of one CXR) is approximately 22.6 milliseconds, making our model time-efficient for the classification of new CXR images.

On the other hand, in the test set, we computed the performance measures using the confusion matrix (Figure 9a) and condensed the results in the Table 5. The proposed method has obtained an accuracy of 1.00; a precision of 0.96; a sensitivity (or recall) of 1.00; also, a F1-score of 0.97 and an AUC of 1.00 for the COVID-19 class. For the HEALTHY class, our model obtained the following scores: accuracy of 0.86; precision: 0.99; recall: 0.62; F1-Score: 0.76; AUC: 0.97. Moreover, for the PNEUMONIA class, we obtained the following scores: accuracy of 0.86; precision of 0.82; recall of 1.00; a F1-Score of 0.90; an AUC of 0.97. Finally, the average accuracy was 0.91; a macro-average precision of 0.92; macro-average recall of 0.82; macro-average recall of 0.87; macro-average of 0.88 for the F1-Score; macro-average score of 0.98 for the AUC.

When comparing with other baseline CNN models, we found (from Table 6) that our method achieved the best scores among all models, making our proposal a better option compared with most common architectures, such as DenseNet121, in computer vision tasks for medical applications. On the other hand, we would highlight that despite the fact of most pneumonia images are from children, and most COVID-19 images are mainly taken of adults, the proposed method does not discriminate by the “age” feature. We could observe from experiments and comparisons with baseline CNN that the proposed method extracts useful features from the texture and patterns within the CXRs to achieve a better classification score. On the contrary, baseline CNN models also misclassified a lot of patients from the “only children” dataset, which contains the samples for HEALTHY and PNEUMONIA classes. Moreover, we showed that an evident linear separability is found by our method for the COVID-19 class compared with HEALTHY and PNEUMONIA (Figure 11). Nonetheless, some overlaps exist between HEALTHY and PNEUMONIA patients, the proposed method effectively classifies most instances of the last two classes from children’s image datasets.

From the scores of Table 5, we can observe that we achieved more than 90% of precision and more than 80% of recall. Moreover, for individual measurements, we obtained a 100% of recall on both COVID-19 and pneumonia-infected patients. One of the main findings of this article is that the following relevant result was obtained in the dataset that we used for the experiments: all the patients infected with SARS-CoV-2 and all the patients infected with pneumonia were correctly classified. The results show more than 85% of F1-Score, which indicates a good relationship between precision and recall. The AUC value indicates the benefits of the model understudy that allow it to avoid, in general, false classification. We obtained a good value for AUC: a score of more than 97% considering the three classes, and 100% on COVID-19 individually. Again, the proposed method obtained the best scores compared with other baseline CNN models.

On the other hand, we argue that we obtained favorable results on the Grad-CAM representations for screening and visualization of the diseases. From Figure 12, we can observe that in the first example, the heatmap indicates the bilateral consolidations on both pulmonary fields. Moreover, the second and third examples show consolidations on both pulmonary fields with a GGO pattern on the vassal region of the left lung from both patients, which is a characteristic the COVID-19 pattern. However, the visualization is not perfect because in the three examples of Figure 12, we mentioned that there existed bilateral GGO patterns. Nonetheless, our network sometimes only detected unilateral
manifestations. Furthermore, sometimes the networks generate the heatmaps around the main bronchi area due to inflammation and dilatation that are not always exclusive manifestations of COVID-19.

In the same way, Grad-CAM for the PNEUMONIA class was generated. In Figure 13, we can observe that the three patients show fluid accumulation and consolidations of typical pneumonia. Therefore, our model correctly identified manifestations of pneumonia in the right lung of each patient. Therefore, generated heatmaps resulted in a useful tool for fast screening of pneumonia disease from CXR images.

7. Conclusions

In this paper, we have presented an effective and fast method to automatically classify three types of chest X-ray images: healthy people, patients with pneumonia, and patients infected with SARS-CoV-2 virus. We have used two different datasets to train the proposed method: for COVID-19, most of the samples were obtained from adult patients, including people from 12 to 87 years old; for HEALTHY and PNEUMONIA instances, we have used a dataset that contains images from children under five years old. The proposed method showed to be effective to extract features from the diseases, despite the fact of age differences. Furthermore, our method proves to be fast and obtain good performance under 12 min, in addition to obtaining good (and, in some cases, excellent) values in performance measures, such as macro-average precision, recall, F1-Score, and AUC compared with other state-of-the art baseline models. When performance measures were calculated for the three cases individually, the results were excellent. Specifically, we were successful in correctly classifying all the patients infected with SARS-CoV-2 and all the patients infected with pneumonia. This allows us to conclude that the proposed method may be useful to help physicians with the classification and visualization of COVID-19 and typical pneumonia.

Author Contributions: Conceptualization, J.E.L.-G., Y.V.-R., and C.Y.-M.; validation, M.A.M.-I.; formal analysis, J.E.L.-G., Y.V.-R., and C.Y.-M.; investigation, M.A.M.-I.; writing—original draft preparation, J.E.L.-G.; writing—review and editing, Y.V.-R., and C.Y.-M.; visualization, J.E.L.-G.; supervision, Y.V.-R., and C.Y.-M. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Acknowledgments: The authors gratefully acknowledge the Instituto Politécnico Nacional (Secretaría Académica, Comisión de Operación y Fomento de Actividades Académicas, Secretaría de Investigación y Posgrado, Centro de Investigación en Computación, and Centro de Innovación y Desarrollo Tecnológico en Computo), the Consejo Nacional de Ciencia y Tecnología (CONACYT), and Sistema Nacional de Investigadores for their economic support to develop this work. In addition, we are grateful to Ana Rosa Ambriz for her medical guidance. Finally, we thank anonymous reviewer #1 for suggesting t-SNE use.

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

References

1. World Health Organization Coronavirus Disease 2019. Available online: https://www.who.int/emergencies/diseases/novel-coronavirus-2019 (accessed on 30 June 2020).
2. World Health Organization Coronavirus. Available online: https://www.who.int/health-topics/coronavirus#tab=tab_1 (accessed on 30 June 2020).
3. Kooraki, S.; Hosseiny, M.; Myers, L.; Gholamrezanezhad, A. Coronavirus (COVID-19) outbreak: What the department of radiology should know. J. Am. Coll. Radiol. 2020, 17, 447–451. [CrossRef] [PubMed]
4. Simpson, S.; Kay, F.U.; Abbara, S.; Bhalla, S.; Chung, J.H.; Chung, M.; Henry, T.S.; Kanne, J.P.; Kligerman, S.; Ko, J.P; et al. Radiological society of North America expert consensus statement on reporting chest ct findings related to COVID-19. Endorsed by the society of thoracic radiology, the American college of radiology, and RSNA. Radiol. Cardiothorac. Imaging 2020, 2, e200152. [CrossRef] [PubMed]
5. Wong, H.Y.F.; Lam, H.Y.S.; Fong, A.H.-T.; Leung, S.T.; Chin, T.W.-Y.; Lo, C.S.Y.; Lui, M.M.-S.; Lee, J.C.Y.; Chiu, K.W.-H.; Chung, T.; et al. Frequency and Distribution of Chest Radiographic Findings in COVID-19 Positive Patients. Radiology 2020, 201160. [CrossRef] [PubMed]
6. Sutton, D. Textbook of Radiology and Imaging, 7th ed.; Chirchill Livingstone: London, UK, 2003.
7. Suzuki, K. Overview of deep learning in medical imaging. *Radiol. Phys. Technol.* **2017**, *10*, 257–273. [CrossRef] [PubMed]

8. Wang, G.; Li, W.; Zuluaga, M.A.; Pratt, R.; Patel, P.A.; Aertsen, M.; Doel, T.; David, A.L.; Deprest, J.; Ourselin, S.; et al. Interactive Medical Image Segmentation Using Deep Learning With Image-Specific Fine Tuning. *IEEE Trans. Med. Imaging* **2018**, *37*, 1562–1573. [CrossRef]

9. Chen, S.; Ding, C.; Liu, M. Dual-force convolutional neural networks for accurate brain tumor segmentation. *Pattern Recognit.* **2019**, *88*, 90–100. [CrossRef]

10. Li, H.; Li, A.; Wang, M. A novel end-to-end brain tumor segmentation method using improved fully convolutional networks. *Comput. Biol. Med.* **2019**, *108*, 150–160. [CrossRef]

11. Li, Z.; Hou, Z.; Chen, C.; Hao, Z.; An, Y.; Liang, S.; Lu, B. Automatic cardiothoracic ratio calculation with deep learning. *IEEE Access* **2019**, *7*, 37749–37756. [CrossRef]

12. Wang, X.; Peng, Y.; Lu, L.; Lu, Z.; Bagheri, M.; Summers, R.M. ChestX-ray: Hospital-Scale chest X-ray database and benchmarks on weakly supervised classification and localization of common thorax diseases. In *Advances in Computer Vision and Pattern Recognition*; Springer: Berlin, Germany, 2019; pp. 369–392.

13. Allaouzi, I.; Ben Ahmed, M. A Novel approach for multi-label chest X-ray classification of common thorax diseases. *IEEE Access* **2019**, *7*, 64279–64288. [CrossRef]

14. Irvin, J.; Rajpurkar, P.; Ko, M.; Yu, Y.; Ciurea-Ilcus, S.; Chute, C.; Marklund, H.; Haghgoo, B.; Ball, R.; Shpanskaya, K.; et al. CheXpert: A large chest radiograph dataset with uncertainty labels and expert comparison. In Proceedings of the AAAI Conference on Artificial Intelligence, Honolulu, HI, USA, 27 January–1 February 2019.

15. Chen, B.; Li, J.; Guo, X.; Lu, G. DualCheXNet: Dual asymmetric feature learning for thoracic disease classification in chest X-rays. *Biomed. Signal Process. Control* **2019**, *53*, 101554. [CrossRef]

16. Chollet, F. Xception: Deep learning with depthwise separable convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 26 June–1 July 2016.

17. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems*; Curran Associates, Inc.: New York, NY, USA, 2012; pp. 1097–1105.

18. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. In Proceedings of the International Conference on Learning Representations, Banff, AB, Canada, 14–16 April 2014.

19. Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlens, J.; Wojna, Z. Rethinking the inception architecture for computer vision. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015.

20. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 26 June–1 July 2016; pp. 770–778.

21. Huang, G.; Liu, Z.; Van Der Maaten, L.; Weinberger, K.Q. Densely connected convolutional networks. In *Advances in Neural Information Processing Systems*; Curran Associates, Inc.: New York, NY, USA, 2012; pp. 1097–1105.

22. Bakator, M.; Radosav, D. Deep learning and medical diagnosis: A review of literature. *Multimodal Technol. Interact.* **2018**, *2*, 47. [CrossRef]

23. Baltruschat, I.M.; Nickisch, H.; Grass, M.; Knopp, T.; Saalbach, A. Comparison of deep learning approaches for multi-label chest X-Ray classification. *Sci. Rep.* **2019**, *9*, 6381. [CrossRef] [PubMed]

24. Xu, S.; Wu, H.; Bie, R. CXNet-m1: Anomaly detection on chest X-Rays with image-based deep learning. *IEEE Access* **2019**, *7*, 4466–4477. [CrossRef]

25. Blumenfeld, A.; Greenspan, H.; Konen, E. Pneumothorax detection in chest radiographs using convolutional neural networks. In *Medical Imaging 2018: Computer-Aided Diagnosis*; Mori, K., Petrick, N., Eds.; SPIE: Bellingham, WA, USA, 2018; Volume 10575, p. 3.

26. Que, Q.; Tang, Z.; Wang, R.; Zeng, Z.; Wang, J.; Chua, M.; Gee, T.S.; Yang, X.; Veeravalli, B. CardioXNet: Automated detection for cardiomegaly based on deep learning. In Proceedings of the 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Honolulu, HI, USA, 17–21 July 2018; pp. 612–615.

27. Salehinejad, H.; Colak, E.; Dowdell, T.; Barfett, J.; Valaee, S. Synthesizing chest X-Ray pathology for training deep convolutional neural networks. *IEEE Trans. Med. Imaging* **2019**, *38*, 1197–1206. [CrossRef]

28. Liang, G.; Zheng, L. A transfer learning method with deep residual network for pediatric pneumonia diagnosis. *Comput. Methods Programs Biomed.* **2020**, *187*, 104964. [CrossRef] [PubMed]
29. Kermany, D.S.; Goldbaum, M.; Cai, W.; Valentim, C.C.S.; Liang, H.; Baxter, S.L.; McKeown, A.; Yang, G.; Wu, X.; Yan, F.; et al. Identifying medical diagnoses and treatable diseases by image-based deep learning. *Cell* 2018, 172, 1122–1131. [CrossRef] [PubMed]

30. Luján-García, J.E.; Yáñez-Márquez, C.; Villuendas-Rey, Y.; Camacho-Nieto, O. A transfer learning method for pneumonia classification and visualization. *Appl. Sci.* 2020, 10, 2908. [CrossRef]

31. Chouhan, V.; Singh, S.K.; Khamparia, A.; Gupta, D.; Tiwari, P.; Moreira, C.; Damaševičius, R.; de Albuquerque, V.H.C. A novel transfer learning based approach for pneumonia detection in chest X-ray images. *Appl. Sci.* 2020, 10, 559. [CrossRef]

32. Radiological Society of North America RSNA Pneumonia Detection Challenge. Available online: https://www.rsna.org/c/rsa-pneumonia-detection-challenge (accessed on 31 May 2019).

33. Sirazhidinov, I.; Kholiavchenko, M.; Mustafaev, T.; Yixuan, Y.; Kuleev, R.; Ibragimov, B. Deep neural network ensemble for pneumonia localization from a large-scale chest x-ray database. *Comput. Electr. Eng.* 2019, 78, 388–399. [CrossRef]

34. Ardakani, A.A.; Kanafi, A.R.; Acharya, U.R.; Khadem, N.; Mohammadi, A. Application of deep learning technique to manage COVID-19 in routine clinical practice using CT images: Results of 10 convolutional neural networks. *Comput. Biol. Med.* 2020, 121, 103795. [CrossRef]

35. Hemdan, E.E.-D.; Shouman, M.A.; Karar, M.E. COVIDX-Net: A framework of deep learning classifiers to diagnose COVID-19 in X-Ray images. *arXiv* 2020, arXiv:2003.11055. Available online: https://arxiv.org/abs/2003.11055 (accessed on 5 August 2020).

36. Butt, C.; Gill, J.; Chun, D.; Babu, B.A. Deep learning system to screen coronavirus disease 2019 pneumonia. *Appl. Intell.* 2020. [CrossRef]

37. Ozturk, T.; Talo, M.; Yildirim, E.A.; Baloglu, U.B.; Yildirim, O.; Rajendra Acharya, U. Automated detection of COVID-19 cases using deep neural networks with X-ray images. *Comput. Biol. Med.* 2020, 121, 103792. [CrossRef] [PubMed]

38. Redmon, J.; Farhadi, A. YOLOv3: An incremental improvement. *arXiv* 2018, arXiv:1804.02767. Available online: https://arxiv.org/abs/1804.02767 (accessed on 6 July 2020).

39. Zhang, J.; Xie, Y.; Liao, Z.; Pang, G.; Verjans, J.; Li, W.; Sun, Z.; He, J.; Li, Y.; Shen, C.; et al. Viral pneumonia screening on chest X-ray images using confidence-aware anomaly detection. *arXiv* 2020, arXiv:2003.12338. Available online: https://arxiv.org/abs/2003.12338 (accessed on 5 August 2020).

40. Cohen, J.P.; Dao, L.; Morrison, P.; Roth, K.; Bengio, Y.; Shen, B.; Abbasi, A.; Hoshmand-Kochi, M.; Ghassemi, M.; Li, H.; et al. Predicting COVID-19 pneumonia severity on chest X-ray with deep learning. *arXiv* 2020, arXiv:2005.11856. Available online: https://arxiv.org/abs/2005.11856 (accessed on 5 August 2020). [CrossRef]

41. Cohen, J.P.; Morrison, P.; Dao, L. COVID-19 image data collection. *arXiv* 2020, arXiv:2003.11597. Available online: https://arxiv.org/abs/2003.11597 (accessed on 1 June 2020).

42. Fernández, A.; García, S.; del Jesus, M.J.; Herrera, F. A study of the behaviour of linguistic fuzzy rule based classification systems in the framework of imbalanced data-sets. *Fuzzy Sets Syst.* 2008, 159, 2378–2398. [CrossRef]

43. Goodfellow, I.; Bengio, Y.; Courville, A. *Deep Learning*; MIT Press: Cambridge, MA, USA, 2016.

44. Selvaraju, R.R.; Cogswell, M.; Das, A.; Vedantam, R.; Parikh, D.; Batra, D. Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization. *arXiv* 2016, arXiv:1610.02391. Available online: https://arxiv.org/abs/1610.02391 (accessed on 1 June 2020).

45. Sokolova, M.; Lapalme, G. A systematic analysis of performance measures for classification tasks. *Inf. Process. Manag.* 2009, 45, 427–437. [CrossRef]

46. Pasa, F.; Golkov, V.; Pfeiffer, F.; Cremer, D.; Pfeiffer, D. Efficient deep network architectures for fast chest X-Ray tuberculosis screening and visualization. *Sci. Rep.* 2019, 9, 6268. [CrossRef] [PubMed]

47. López, V.; Fernández, A.; Moreno-Torres, J.G.; Herrera, F. Analysis of preprocessing vs. cost-sensitive learning for imbalanced classification. Open problems on intrinsic data characteristics. *Expert Syst. Appl.* 2012, 39, 6685–6608. [CrossRef]

48. Murphy, K.P. *Machine Learning: A Probabilistic Perspective*, 1st ed.; MIT Press: Cambridge, MA, USA, 2012.

49. Pedregosa, F.; Michel, V.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Vanderplas, J.; Cournapeau, D.; Pedregosa, F.; Varoquaux, G.; et al. Scikit-learn: Machine learning in python Gaël Varoquaux bertrand thirion vincent dubourg alexandre passos PEDREGOSA, VAROQUAX, GRAMFORT ET AL. Matthieu Perrot. *J. Mach. Learn. Res.* 2011, 12, 2825–2830.
50. Hernández-Castañeda, Á.; Calvo, H.; Gelbukh, A.; Flores, J. Cross-domain deception detection using support vector networks. Soft Comput. 2017, 21, 585–595. [CrossRef]

51. Kermany, D.; Zhang, K.; Goldbaum, M. Labeled Optical Coherence Tomography (OCT) and Chest X-Ray Images for Classification. Available online: https://data.mendeley.com/datasets/rscbjbr9sj/2 (accessed on 7 October 2019).

52. Bradski, G. The OpenCV library. Dr. Dobb’s J. Softw. Tools 2000, 120, 122–125.

53. Van Der Maaten, L.; Hinton, G. Visualizing Data using t-SNE. J. Mach. Learn. Res. 2008, 9, 2579–2605.

© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).