

# Comparison of GANs for Covid-19 X-ray classification

Luiz Felipe Cavalcanti<sup>1</sup>, Lilian Berton<sup>1</sup>

<sup>1</sup>Institute of Science and Technology – Federal University of Sao Paulo (UNIFESP)  
São José dos Campos 12247-014 – SP – Brazil

lberton@unifesp.br

**Abstract.** *Image classification has been applied to several real problems. However, getting labeled data is a costly task, since it demands time, resources and experts. Furthermore, some domains like disease detection suffer from unbalanced classes. These scenarios are challenging and degrade the performance of machine learning algorithms. In these cases, we can use Data Augmentation (DA) approaches to increase the number of labeled examples in a dataset. The objective of this work is to analyze the use of Generative Adversarial Networks (GANs) as DA, which are capable of synthesizing artificial data from the original data, under an adversarial process of two neural networks. The GANs are applied in the classification of unbalanced Covid-19 radiological images. Increasing the number of images led to better accuracy for all the GANs tested, especially in the multi-label dataset, mitigating the bias for unbalanced classes.*

## 1. Introduction

In the “Big Data” era, researches in Artificial Intelligence (AI) and Machine Learning (ML) have stood out since several algorithms in this area have been successful in pattern recognition, organization, classification, and content recommendation [Russell and Norvig 2009, Nguyen et al. 2019, Afonso et al. 2019, Sano et al. 2019]. Applications in many areas have made progress for companies, start-ups, research institutes, and governments. A good AI application results in a positive cycle, as more users will use the system, thus generating more data that makes it possible to continuously improve the models.

The development of AI took place mainly in the area of ML, specifically using Deep Neural Networks [LeCun et al. 2015]. In contrast to expert systems that use inference from a knowledge base by applying first-order logical rules, neural networks are characterized as connectionist artificial intelligence, building an internal representation of the input data. Especially, deep learning is characterized by computing these representations in terms of the previous ones, composing a function of greater complexity that can obtain better results.

The Generative Adversarial Networks (GANs) [Goodfellow et al. 2014, Goodfellow et al. 2016] are deep neural networks composed of two networks that confront each other intending to learn to imitate some data distribution. They have been applied in various domains such as images, music, speech, etc. GANs belong to the set of generative models, which produce new content. While discriminative models classify data, given a set of attributes they predict a label to which that data belongs.

GANs can also be used for *Data Augmentation* (DA), which consists of computational techniques in order to increase the number of labeled examples in a dataset

and thus improve the classification results [Fawzi et al. 2016, Antoniou et al. 2017]. One area that has benefited from GANs and deep neural networks is image classification. In [Frid-Adar et al. 2018], the authors proposed the use of traditional DA methods and the application of GANs to generate synthetic data. The tests were performed using a computer tomography database consisting of 182 images of liver lesions. The results showed a significant improvement. In [Tanaka and Aranha 2019] the authors used the GANs, the SMOTE, and ADASYN techniques to generate new data. The three datasets used were not balanced. The accuracy and precision results found with the classifier using balanced synthetic data from GANs were superior to those found with the original dataset.

In data classification, the algorithm will generally be biased in classifying the examples according to the class with the highest occurrence. Common methods for dealing with class imbalance are *under-sampling* and *over-sampling*, which will decrease the set of observations with the majority classification, or increase the observations with the minority classification, respectively [Russell and Norvig 2009]. In this work, we will use the GANs to perform *over-sampling*, synthesizing examples with less quantity to equalize with the ones with greater occurrence. Thus, we aim to investigate the use of GANs for DA in the treatment of unbalanced classes in Covid-19 images, studying and comparing different GAN models.

Covid-19 or SARS-Cov-2 spread rapidly in 2020 and has surpassed the number of deaths compared to SARS-Cov-1 [Gandhi et al. 2020, Garcia and Berton 2021]. It has been cataloged as a worldwide pandemic by the World Health Organization (WHO) since March 11, 2020. In many places, radiological tests may be more accessible, such as computed tomography (CT) and chest X-rays, which detect the presence of pulmonary alterations caused by the disease. We aim to improve the accuracy of results by employing GANs in classifying these images. In our application, we considered COVID-19 against normal X-ray images and COVID-19 against other diseases in multi-class classification.

Next, in Section 2, some related works are summarized. In Section 3 the materials and methods used in the work are presented. Section 4 brings the results of the experiments and finally, Section 5 brings the final considerations.

## 2. Related work

Some authors summarized the applications of GANs in medical applications. For example, [Yi et al. 2019] surveyed 150 articles, most of them explored GANs as data augmentation to alleviate the data scarcity and overfitting problem. Other works performed image captioning, cross-modality translation, automatic magnetic resonance protocol generation have also been discussed. [Kazemini et al. 2020] analyzed 79 innovative GAN-based papers in medical domains regarding image de-noising, reconstruction, segmentation, data simulation, detection, or classification.

Regarding Covid-19 applications, some recent papers used GANs in order to increase the training data volumes. [Goel et al. 2021] used the Whale Optimization Algorithm (WOA) to optimize the hyperparameters of GAN's generator to expand computed tomography images of Covid-19. The proposed method was tested in datasets consisting of Covid-19 and non-Covid-19 images. The accuracy considering an InceptionV3 DL network was 99.22%.

[Loey et al. 2020] employed conditional GAN augmentation to improve the accu-

racy of multi-class classification to identify 4 classes (Covid, normal, bacterial pneumonia, and viral pneumonia). The accuracy was 66.7%, 80.56% and 69.46% using AlexNet, GoogleNet and ResNet18. Considering a binary classification they achieved 100% of accuracy.

[Menon et al. 2020] presented a novel Mean Teacher + Transfer GAN (MTT-GAN), they argue the mean teacher algorithm makes the gradient descent converge more consistently and to a better global optimum than Adam optimization alone. They employed transfer learning to both the generator and discriminator from the Kaggle pneumonia X-ray dataset, not ImageNet. The results are better than a baseline GAN. The accuracy in a multi-class dataset was 83.45% and 84.91% on VGG-19 and AlexNet. Considering a binary classification they achieved 99.26% of accuracy.

[Quan et al. 2021] proposed the X-ray Projected Generative Adversarial Network (XPGAN) for projecting 3D CT volumes with golden ground truth labels to generate more realistic X-ray images. They trained a deep learning model to classify Covid-19 chest X-ray images. In a binary dataset, they achieved 82.3% of f1-score.

Most papers proposed a modification to create their own GAN and did not compare to other models. Here, we aim to compare four GANs models from literature for data augmentation and evaluate their efficiency for binary and multi-class classification of Covid-19 X-ray. Although in binary classification our results are lower than reported in these works, the multi-class results are higher.

### **3. Material and methods**

This section presents the datasets considered, the GANs and CNNs employed and their parameters settings. We also present the computer environment and the experimental set up to running the experiments.

#### **3.1. Datasets**

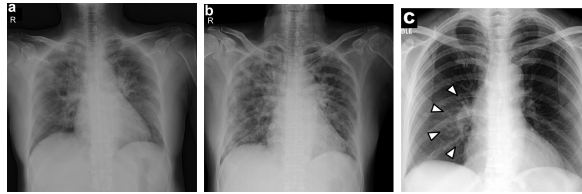
The datasets used are radiological images, publicly available on the internet, and are described below.

##### **3.1.1. Covid-19 X-ray images**

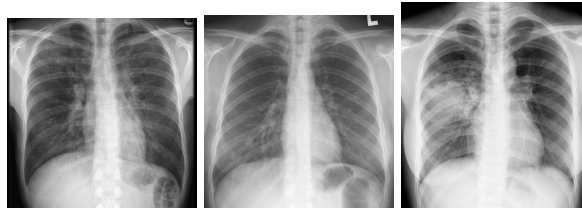
The *Covid-19 X-ray images* [KAGGLE 2020a] dataset, downloaded for use in October 2020, consisted of 329 X-ray images or CT scans of the chest of patients with Covid-19, Middle East respiratory syndrome (MERS), Severe acute respiratory syndrome (SARS), Acute respiratory distress syndrome (ARDS), and Healthy. In addition, information was given related to the patient's health status, whether he survived the disease, gender, age and place where the examination was performed, but this information has been removed for our experiments.

The class distribution consists of 77% of the data to class Covid-19. Given the small number of images of the remaining classes, the dataset was re-divided into Covid-19 and Non-Covid (joining the other classes). Finally, after removing some noise images, we considered 231 images: 67 of the class Non-Covid and 164 of the class Covid-19. Some examples of images from both classes are shown in Figures 1 and 2, respectively.

**Figure 1. Examples of images labeled as Covid-19 [KAGGLE 2020a].**



**Figure 2. Examples of images labeled as Non-Covid [KAGGLE 2020a].**



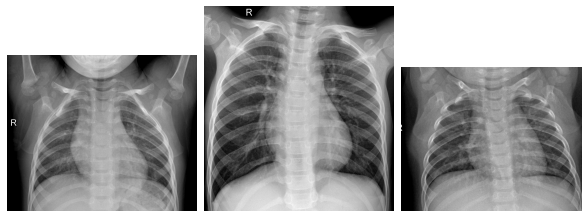
### 3.1.2. Chest X-ray (Covid-19 & Pneumonia)

The *Chest X-ray (Covid-19 & Pneumonia)* [KAGGLE 2020b] dataset is composed of 6,432 X-ray images separated in two folders (test and training) that are subdivided into three others (PNEUMONIA, Covid-19, NORMAL). The test folder consists of 20% of the images. However, for the experiments in this work, 300, 300, and 500 random images of the PNEUMONIA, Covid-19, and NORMAL classes were selected in order to train the GANs in smaller slightly unbalanced datasets. For the evaluation, we considered 300 images of the Pneumonia, 300 images of Normal, and 116 images of the Covid. Some examples of images of the three classes are shown in Figures 3, 4 and 5, respectively.

**Figure 3. Examples of images labeled as Covid-19 [KAGGLE 2020b].**



**Figure 4. Examples of images labeled as NORMAL [KAGGLE 2020b].**

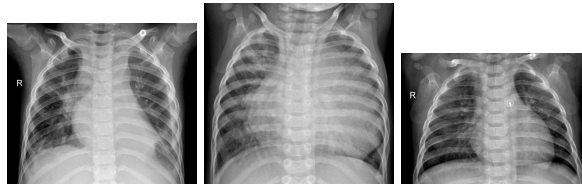


## 3.2. GAN parameters

1. DCGAN: The implementation used was based on the DCGAN demonstrated by Tensorflow<sup>1</sup>, with the following parameters:

<sup>1</sup><https://www.tensorflow.org/tutorials/generative/dcgan>

Figure 5. Examples of images labeled as PNEUMONIA [KAGGLE 2020b].



- Batch: 50
  - Epochs: 8,000 to 15,000
  - Generator:
    - 1 Dense Layer;
    - 5 sequences of Conv2DTranspose (stride 2.2), LeakyReLU and BatchNormalization
    - 1 Conv2DTranspose with TanH activation
  - Discriminator:
    - 2 Sequences of Conv2D, LeakyReLU and Dropout (0.3)
    - 1 Flatten
    - 1 Dense
2. HyperGAN: The default configuration implemented in the library was used, the layers considered are those shown in the HyperGAN github<sup>2</sup>:
- Epochs: 30,000 to 35,000
  - Batch: 16
  - Generator:
    - 1 identity layer (name = z)
    - 1 latent layer
    - 1 split layer Uma28 0
    - 1 layer multi\_head\_attention with ReLu activation and size 256
    - 1 identity layer (name = w)
    - 1 layer Z
    - 1 Latent layer
    - 1 Split layer 128 1
    - 2 layers multi\_head\_attention with Relu activation and size 256 and 25,600 respectively
    - 1 reshape layer for 10\*10\*256 with Relu activation
    - On output 1 layer of resizable\_stack with HardTanH activation
  - Discriminator:
    - 5 Convolution Layers with 64, 128, 256, 512 and 512 filters, with ReLu activation and stride=2
    - 2 Layers “Linear” with ReLu activation and 1024 and 128 outputs, respectively
    - 1 Multi Layer\_Attention\_Head and size 1
3. CycleGAN: A version of the network used and demonstrated in the article [Welander et al. 2018] was used:
- Epochs: 150 to 200
  - Batch: 10
    - 1 Conv2D layer, with kernel 7 and stride 1 and 32 filters with ReLU activation
    - 2 Conv2D Layers with kernel 3 and stride 2, 64 and 128 filters and ReLU activation
    - 9 Blocks composed of two layers Conv2D of kernel three and stride 1 of size 128 and the first with ReLU activation
    - 2 Conv2DTranspose Layers of stride 2, kernel 3, 64 and 32 filters and ReLU activation;

<sup>2</sup><https://github.com/HyperGAN>

- 1 ReflectionPadding2D Layer
  - 1 Conv2D layer of kernel 7 and stride 1 with TanH activation
  - Discriminator:
    - 1 Conv2D layer with input data size and 64 stride=2 and LeakyReLU filters
    - 3 layers of Conv2D with 128, 256 and 512, strides=(2,2,1), normalization and activation LeakyReLU
    - 1 Conv2D layer with 1 filter, kernel 4 stride 1
4. UNIT: A version of the same network presented in the article [Welander et al. 2018] was used:
- Epochs: 150 to 200
  - Batch: 10
    - 3 Blocks composed of Two Conv2D with BatchNormalization, the first with ReLU activation and the second with 50% Dropout, each with 256 filters
    - 2 Layers of Conv2DTranspose with 3 stride 2 kernel and LeakyReLU activation, with 128 and 64 filters
    - 1 Conv2DTranspose layer with kernel and stride 1 and TanH activation
  - Discriminator:
    - 4 Conv2D Layers of kernel 3 and stride 2, of sizes 64, 128, 256 and 512 and LeakyReLU activation
    - 1 Conv2D layer with a single kernel filter and stride 1 with Sigmoid activation

### 3.3. CNNs Parameters

The ResNET50<sup>3</sup>, VGG16, VGG19<sup>4</sup>, and Xception<sup>5</sup> generated by the Keras function were used without considering the predefined Imagenet weights. The following set up was adopted.

- Batch: 64
- Epochs: 50 for Chest X-ray (Covid-19 & Pneumonia) and 35 for Covid-19 X-ray images

### 3.4. Computer environment

All tests were performed on Google Colab<sup>6</sup>, a computational environment for running notebooks in Python through GPUs offered by Google. Every time a run time environment is requested, Google Colab provides access to a random computer.

### 3.5. Experimental set up

In order to train the GANs, it was necessary to adapt the *Chest X-ray (Covid-19 & Pneumonia)* and *Covid-19 X-ray images* datasets, as the size of the images was different. We use the function *resize* from the CV2 library, offered by OpenCV, to normalize the data to the size  $256 \times 256$ , in RGB format. Then, the *hold out* was applied and the images were randomly distributed, using the sklearn library into 70% for training and 30% for testing. The training images were used for training the GANs models HyperGAN, DCGAN, CycleGAN, and UNIT, which were used to generate the images that would be used in the *Data Augmentation*.

The tests were performed varying the CNNs to extract *features* from the images. The ResNet50, VGG16, VGG19, and Xception models were considered. The tests performed to compare the accuracy of the models: i) without using GANs as DA in the data; ii) applying *over-sampling* through GANs only in the minority class; iii) expanding all dataset classes through GANs. Tables 1 and 2

<sup>3</sup><https://keras.io/api/applications/resnet/resnet50-function>

<sup>4</sup><https://keras.io/api/applications/vgg/vgg16-function>

<sup>5</sup><https://keras.io/api/applications/xception>

<sup>6</sup><https://colab.research.google.com/>

show the distribution of examples by class considering the three test cases. The test set has only original images (no DA).

It is noteworthy that when comparing the performance of the same CNN with two different training sets, with or without DA for example, before the model is created the same *seed* is started, so that both have the same initial weights, changing only the training data.

Covid-19 X-ray images			
	Normal	Covid-19	Total
No DA	47	164	211
Balanced	167	164	331
Expanded	267	264	531
Test set	20	71	91

**Table 1. The amount of images used in the experiments performed with the dataset Covid-19 X-ray images.**

Chest X-ray (Covid-19 & Pneumonia)				
	Normal	Covid-19	Pneumonia	Total
No DA	500	300	300	1100
Balanced	500	500	500	1500
Expanded	600	600	600	1800
Test set	300	116	300	716

**Table 2. The amount of images used in the experiments performed with the dataset Chest X-ray (Covid-19 & Pneumonia).**

## 4. Results

### 4.1. Accuracy comparing GANs

Table 3 presents the classification accuracy in the test set comparing the use of HyperGAN, DCGAN, CycleGAN, and UNIT and the four CNNs (ResNet50, VGG16, VGG19, and Xception) for extracting *features* from the images.

For the Covid-19 X-ray dataset applying the GANs, there was improvement in the ResNet50, VGG16 and VGG19 classification and no significant change in the Xception. This is a small dataset with a very unequal data distribution that did not favor the training of the generative network.

In the Chest X-ray (Covid-19 & Pneumonia) dataset, there was an increase in the performance of all CNNs, in VGG16 and VGG19 the accuracy increases around 3%. On the other hand, ResNet and Xception had a lower performance without DA and achieved a significant increase in some of the tests, especially where the DA was performed using DCGAN.

Comparing the CNNs results, VGG16 and VGG19 achieved greater accuracy and consistent results in both Chest X-ray (Covid-19 & Pneumonia) and Covid-19 X-ray datasets for all GANs applied.

**Table 3. Classification accuracy in Covid-19 X-ray images and Chest X-ray (Covid-19 & Pneumonia) datasets, considering HyperGAN, DCGAN, CycleGAN and UNIT for data increase and different CNNs for *features* extraction.**

	ResNet50	VGG16	VGG19	Xception
<i>Covid-19 X-ray images</i>				
No DA	0.7472	0.7582	0.7472	<b>0.7802</b>
Balanced with HyperGAN	<b>0.8131</b>	0.7692	0.7802	0.7802
Expanded with HyperGAN	0.7582	0.7692	0.7582	0.6813
Balanced with DCGAN	0.6483	0.7362	<b>0.7802</b>	<b>0.7802</b>
Expanded with DCGAN	0.6813	0.7582	0.7252	0.7582
Balanced with CycleGAN	0.2197	<b>0.7802</b>	<b>0.7802</b>	0.2197
Expanded with CycleGAN	0.7802	0.6923	<b>0.7802</b>	<b>0.7802</b>
Balanced with UNIT	0.7802	<b>0.7802</b>	<b>0.7802</b>	0.2197
Expanded with UNIT	0.2197	0.7252	<b>0.7802</b>	<b>0.7802</b>
<i>Chest X-ray (Covid-19 &amp; Pneumonia)</i>				
No DA	0.5572	0.9189	0.9287	0.2108
Balanced with HyperGAN	0.7863	0.9357	0.9148	0.8924
Expanded with HyperGAN	0.9050	0.9231	0.9217	0.9078
Balanced with DCGAN	<b>0.9329</b>	0.9329	0.9329	0.7779
Expanded with DCGAN	0.9106	0.9399	<b>0.9427</b>	<b>0.9329</b>
Balanced with CycleGAN	0.5027	0.9203	0.9315	0.2220
Expanded with CycleGAN	0.4483	<b>0.9441</b>	0.9259	0.5279
Balanced with UNIT	0.4189	0.9273	<b>0.9427</b>	0.8840
Expanded with UNIT	0.3868	0.9301	0.9385	0.5642

#### 4.2. F1 score analysis

As accuracy can mask the results having a bias for the majority class, the results of the measures *precision*, *recall* and *F1* of each class are presented below for HyperGAN. Tables 4 and 5 present the measures *precision*, *recall* and *f1* for the dataset *Covid-19 X-ray images*. In Table 4 DA was not applied, in Table 5 HyperGAN was applied to: i) expanding only the minority class and ii) expanding all classes.

From the tested data we observe that in some cases without the use of DA, the CNNs are biased to classify all instances in the majority class, causing a small F1 for Non-Covid, including zero for Xception resulting in a division by 0 in the calculation of some measures statistics, represented as *n/a* in the tables. However, increasing the number of images in all the classes led to better results for all CNNs, increasing the F1 of Non-Covid class.

Tables 6 and 7 present the measures *precision*, *recall* and *F1* for the dataset *Chest X-ray (Covid-19 & Pneumonia)*. In Table 6 DA was not applied, in Table 7 HyperGAN was applied to: i) expanding only the minority class and ii) to expanding all classes.

In this dataset when no DA is used the Normal class are not recognized by some CNNs, while using GANs to expand the number of images for all classes lead to better F1 scores.

Figure 6 presents examples of images generated by HyperGAN, DCGAN, CycleGAN, and UNIT, respectively, in the dataset *Covid-19 X-ray images*. GAN-generated Covid-19 X-rays can have the presence of fuzzy boundaries, mainly because of insufficient data volumes of Covid-19 images used for training. Nevertheless, these fake images still can improve classification accuracy as shown by the previous results.



**Table 4. CNNs results in identifying the 2 classes extracted from the Covid-19 X-ray dataset without the use of DA (using only the 211 original images that are distributed between 47 Non-Covid and 164 Covid).**

Patology	<i>Precision</i>	<i>Recall</i>	<i>F1 measure</i>
ResNet-50			
Non-Covid	0.40	0.30	0.34
Covid-19	0.82	0.87	0.84
VGG16			
Non-Covid	0.43	0.30	<b>0.35</b>
Covid-19	0.82	0.89	<b>0.85</b>
VGG19			
Non-Covid	0.40	0.30	0.34
Covid-19	0.82	0.87	0.84
Xception			
Non-Covid	n/a	0.0	n/a
Covid-19	0.78	1	0.88

**Table 5. CNNs results in identifying the 2 classes extracted from the Covid-19 X-ray dataset using HyperGAN in order to: i) balance the dataset, adding 120 images in the Non-Covid class (totaling 331 images and making the dataset more balanced with 164 Covid and 167 Non-Covid); ii) balance and expand the dataset, adding 220 Non-Covid images and 100 Covid images (totaling 531 images, with 267 Non-Covid and 264 Covid).**

Balanced				Expanded		
Patology	<i>Precision</i>	<i>Recall</i>	<i>F1 measure</i>	<i>Precision</i>	<i>Recall</i>	<i>F1 measure</i>
ResNet-50						
Non-Covid	0.71	0.25	0.37	0.44	0.35	0.39
Covid-19	0.82	0.97	0.89	0.83	0.87	0.85
VGG16						
Non-Covid	0.46	0.30	0.36	0.47	0.35	<b>0.40</b>
Covid-19	0.82	0.90	0.86	0.83	0.89	<b>0.86</b>
VGG19						
Non-Covid	0.50	0.45	<b>0.47</b>	0.44	0.35	0.39
Covid-19	0.85	0.87	<b>0.86</b>	0.83	0.87	0.85
Xception						
Non-Covid	n/a	0.0	n/a	0.20	0.15	0.17
Covid-19	0.78	1	0.88	0.78	0.83	0.80

## 5. Conclusion

This work analyzed techniques of deep neural networks such as Generative Adversarial Networks (GANs) for the generation of artificial data to be used in the problem of unbalanced classes. We tested the models in challenging scenarios such as medical imaging with few labeled data.

The GANs HyperGAN, DCGAN, CycleGAN and UNIT were tested in two datasets related to thoracic images of patients suffering from Covid-19 and pneumonia. We noted that overall GANs have increased the accuracy of the classification. In the tests without DA, the classifier tends to classify all images as being of the majority class, generating a biased classifier. Employing the GANs this problem was reduced. In general, the use of GANs for *data augmentation* brought improvements to the experiments, especially for ResNet-50 and Xcep-

**Table 6. Results of CNNs in identifying the 3 classes of the dataset *Chest X-ray (Covid-19 & Pneumonia)* without the use of DA, thus the Normal, Covid-19, and Pneumonia classes maintained respectively 500, 300, 300 images.**

Patology	<i>Precision</i>	<i>Recall</i>	<i>F1 measure</i>
ResNet-50			
Normal	n/a	0.0	n/a
Covid-19	0.84	0.92	0.88
Pneumonia	0.50	0.97	0.66
VGG16			
Normal	0.88	0.97	0.92
Covid-19	0.91	0.96	0.93
Pneumonia	0.98	0.85	0.91
VGG19			
Normal	0.93	0.92	<b>0.92</b>
Covid-19	0.96	0.96	<b>0.96</b>
Pneumonia	0.92	0.93	<b>0.93</b>
Xception			
Normal	n/a	0.0	n/a
Covid-19	0.17	1	0.29
Pneumonia	0.97	0.12	0.21

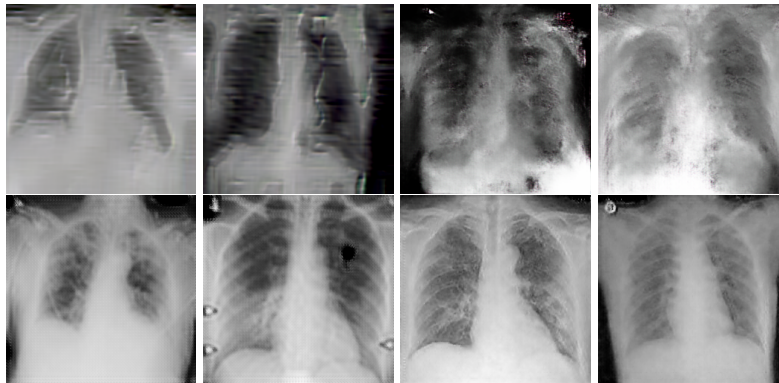
**Table 7. Results of the CNNs in identifying the 3 classes extracted from the *Chest X-ray (Covid-19 & Pneumonia)* datasets with the use of DA by the *HyperGAN* in order to: i) balance the datasets (200 images were added in the Covid-19 and Pneumonia classes, totaling 1500 images); ii) expand and balance the dataset (adding 100, 300 and 300 images in the Normal, Covid-19, and Pneumonia classes respectively, making a total of 1800 images).**

	Balanced			Expanded		
Patology	<i>Precision</i>	<i>Recall</i>	<i>F1 measure</i>	<i>Precision</i>	<i>Recall</i>	<i>F1 measure</i>
ResNet-50						
Normal	0.99	0.52	0.68	0.88	0.91	0.90
Covid-19	0.75	0.96	0.84	0.93	0.95	0.94
Pneumonia	0.72	0.98	0.83	0.92	0.88	0.90
VGG16						
Normal	0.91	0.95	<b>0.93</b>	0.90	0.95	<b>0.92</b>
Covid-19	1	0.94	<b>0.97</b>	0.95	0.91	<b>0.93</b>
Pneumonia	0.94	0.92	<b>0.93</b>	0.94	0.90	<b>0.92</b>
VGG19						
Normal	0.91	0.93	0.92	0.91	0.95	0.93
Covid-19	0.92	0.91	0.91	0.86	0.92	0.89
Pneumonia	0.92	0.91	0.91	0.96	0.89	0.92
Xception						
Normal	0.95	0.80	0.87	0.90	0.89	0.90
Covid-19	0.86	1	0.92	0.86	0.98	0.92
Pneumonia	0.87	0.94	0.90	0.94	0.89	0.91

tion classifiers that obtained a high increase in accuracy/F1 in almost all the tests performed.

One limitation we face using GANs is the high computational cost to train the models and the need for a large set of labeled data. For future work, we can enrich the results by improving these limitations. Moreover, for an intervention in

**Figure 6. Examples of images generated by HyperGAN, DCGAN, CycleGAN and UNIT (two of each model), from the dataset Covid-19 X-ray images.**



public health, further analyses are required.

## **6. Acknowledgements**

Authors thanks Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) grant 2020/03292-6.

## **References**

- Afonso, B., Melo, L., Oliveira, W., Sousa, S., and Berton, L. (2019). Housing prices prediction with a deep learning and random forest ensemble. In *Anais do XVI Encontro Nacional de Inteligência Artificial e Computacional*, pages 389–400. SBC.
- Antoniou, A., Storkey, A., and Edwards, H. (2017). Data augmentation generative adversarial networks. *arXiv preprint arXiv:1711.04340*.
- Fawzi, A., Samulowitz, H., Turaga, D., and Frossard, P. (2016). Adaptive data augmentation for image classification. In *2016 IEEE International Conference on Image Processing (ICIP)*, pages 3688–3692.
- Frid-Adar, M., Klang, E., Amitai, M., Goldberger, J., and Greenspan, H. (2018). Synthetic data augmentation using gan for improved liver lesion classification. In *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, pages 289–293.
- Gandhi, R. T., Lynch, J. B., and del Rio, C. (2020). Mild or moderate covid-19. *The New England Journal of Medicine*.
- Garcia, K. and Berton, L. (2021). Topic detection and sentiment analysis in twitter content related to covid-19 from brazil and the usa. *Applied Soft Computing*, 101:107057.
- Goel, T., Murugan, R., Mirjalili, S., and Chakrabartty, D. K. (2021). Automatic screening of covid-19 using an optimized generative adversarial network. *Cognitive computation*, pages 1–16.
- Goodfellow, I., Bengio, Y., Courville, A., and Bengio, Y. (2016). *Deep learning*, volume 1. MIT press Cambridge.

- Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. (2014). Generative adversarial nets. In *Advances in neural information processing systems*, pages 2672–2680.
- KAGGLE (2020a). Covid-19 chest xray. Disponível em: <https://www.kaggle.com/bachrr/covid-chest-xray>.
- KAGGLE (2020b). Covid-19/pneumonia chest xray). Disponível em: <https://www.kaggle.com/prashant268/chest-xray-covid19-pneumonia>.
- Kazemian, S., Baur, C., Kuijper, A., van Ginneken, B., Navab, N., Albarqouni, S., and Mukhopadhyay, A. (2020). Gans for medical image analysis. *Artificial Intelligence in Medicine*, page 101938.
- LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature*, 521(7553):436–444.
- Loey, M., Smarandache, F., and M Khalifa, N. E. (2020). Within the lack of chest covid-19 x-ray dataset: a novel detection model based on gan and deep transfer learning. *Symmetry*, 12(4):651.
- Menon, S., Galita, J., Chapman, D., Gangopadhyay, A., Mangalagiri, J., Nguyen, P., Yesha, Y., Yesha, Y., Saboury, B., and Morris, M. (2020). Generating realistic covid-19 x-rays with a mean teacher+ transfer learning gan. In *2020 IEEE International Conference on Big Data (Big Data)*, pages 1216–1225. IEEE.
- Nguyen, G., Dlugolinsky, S., Bobák, M., Tran, V., García, Á. L., Heredia, I., Malík, P., and Hluchý, L. (2019). Machine learning and deep learning frameworks and libraries for large-scale data mining: a survey. *Artificial Intelligence Review*, 52(1):77–124.
- Quan, T. M., Thanh, H. M., Huy, T. D., Chanh, N. D. T., Anh, N. T. P., Vu, P. H., Nam, N. H., Tuong, T. Q., Dien, V. M., Van Giang, B., et al. (2021). Xp-gan: X-ray projected generative adversarial network for improving covid-19 image classification. In *2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)*, pages 1509–1513. IEEE.
- Russell, S. and Norvig, P. (2009). *Artificial Intelligence: A Modern Approach*, volume 84. Prentice Hall Press, Upper Saddle River, NJ, USA, 3rd edition.
- Sano, H., Malere, J., and Berton, L. (2019). Single and multiple failures diagnostics of pneumatic valves using machine learning. In *Anais do XVI Encontro Nacional de Inteligência Artificial e Computacional*, pages 202–213. SBC.
- Tanaka, F. H. K. d. S. and Aranha, C. (2019). Data augmentation using gans. *arXiv preprint arXiv:1904.09135*.
- Welander, P., Karlsson, S., and Eklund, A. (2018). Generative adversarial networks for image-to-image translation on multi-contrast mr images—a comparison of cyclegan and unit. *arXiv preprint arXiv:1806.07777*.
- Yi, X., Walia, E., and Babyn, P. (2019). Generative adversarial network in medical imaging: A review. *Medical image analysis*, 58:101552.