Pre-processing and Handling Unbalanced Data in CNN for Improving Automated Detection of COVID-19 Cases: Preliminary Results

Hector Mejia[1 [0000−0002−6635−3189]], Franz Guzman[1 [0000−0003−0079−0367]], Carlos Bustamante-Orellana[4 [0000−0002−7074−1811]], and Lorena Guachi-Guachi[1,2,3 [0000−0002−8951−8150]]

[1] Yachay Tech University, Hacienda San José, Urcuquí 100119, Ecuador
[2] SDAS Research Group (www.sdas-group.com)
[3] Department of Mechatronics, Universidad Internacional del Ecuador, Av. Simon Bolivar 170411, Ecuador
[4] Arizona State University, Tempe, AZ, United States
lguachi@yachaytech.edu.ec

Abstract. In the context of the COVID-19 pandemic, early diagnosis and treatment are crucial to avoid a drastic increase in the number of new infections. Chest imaging plays an important role in the early detection of the disease since it can be used to identify the initial phase lung infection caused by the SARS-CoV-2 virus. Recently, some researchers have begun to explore Convolutional Neural Networks (CNNs) to detect COVID-19 cases from chest X-ray images. CNN is a category of deep artificial neural networks that has demonstrated great success in computer vision applications, such as video and image analysis. However, this type of network is still affected by abnormal real-world factors such as unbalanced data, presence of noise, blurring, or other quality degradation that can diminish their overall performance when they are not properly handled. This work introduces a methodology to explore and compare the overall accuracy achieved by an Xception-based CNN model when it is combined with benchmark techniques such as unsharp masking, batch balance, and data augmentation aiming at enhancing image details (such as the bone structure and lung) and handling unbalanced datasets. Experiments are done referring to the COVIDx dataset. Preliminary results demonstrate that the proposed methodology leads to higher accuracy when implementing both image enhancement and batch balancing.

Keywords: Chest X-ray Classification · Convolutional Neural Networks (CNNs) · Image Processing

1 Introduction

The rapid spread of COVID-19 disease over the world is causing the health system of many places to quickly become overwhelmed. Even major health-care

centers get over-saturated because lower-level centers are not equipped to evaluate the condition of the patients and the medical personnel starts to become scarce. Recently, some works have begun to explore Convolutional Neural Networks (CNNs) approaches for automatic COVID-19 detection from chest X-ray and CT scan images [15], [11], [1], [16]. CNNs, a category of deep learning approaches, have been extensively adopted for image classification purposes due to its ability to learn key features on their own and its successful precision achieved to general purposes analysis of visual imagery.

Although some works, based on deep learning approaches, presented in the literature have demonstrated high sensitivity for COVID-19 detection, they are still affected by abnormal factors such as unbalance data, presence of noise, blurring, or other quality degradations which may diminish the overall performance of a CNN.

In this sense, to evaluate how pre-processing and handling unbalance data influence on the overall accuracy of CNN-based models, in this work, we propose a methodology to explore and compare the performance achieved by an Xception-based CNN [4]. This CNN was combined with unsharp masking, which is a simple and effective method to enhance edges, batch balancing, and data augmentation. Xception CNN is a deep convolutional neural network organized in blocks that implements a modified separable convolutions procedure, to reduce parameters on the neural network. It was chosen due to its successful results achieved in computer vision applications such as breast cancer diagnosis [8], blood cells classification [9], pneumonia classification [3], among others. As a case study, we used the COVIDx dataset [5] since COVID-19 has become a critical viral disease that might be diagnosed from X-ray images. Images in DCM format, exclusive for medical imaging, were transformed into PNG format for the dataset to be homogeneous, as "COVID-19" images were in the later format; thereby losing detail.

Preliminary results show that pre-processing visually causes a slight reduction in the noise of the images, makes the images more focused and highlights the edges of the bone structure and lungs. Besides, when pre-processing is combined with balance correction, it increases the network's confidences on softmax predictions, as well as precision, recall and accuracy metrics.

The remaining of this paper is organized as follows. Section 2 describes the most relevant related works. The proposed methodology is presented in Section 3. The experimental setup is described in Section 4. The experimental results obtained on the processed dataset are illustrated and discussed in Section 5. Finally, Section 6 deals with the concluding remarks and future work.

2 Related work

Among various CNN standard architectures in COVID-19 context: Darknet, MobileNet, VGG16 or Inception-Resnet-based CNNs have been used as the underlying model structure, aiming at exploring the performance to identify COVID-19 cases from chest X-ray images as binary or multi-class classification problem.

The effectiveness of a CNN model to identify COVID-19 cases among two or multiple classes depends mainly on the CNN depth, number of classes, quantity and quality of data. For instance in [11], a framework based on a modified architecture of Darknet [12], with only 17 convolutional layers, formulates COVID-19 detection as a binary classification problem and achieves sensitivity and specificity values of 90.65% and 93.18%, respectively for the COVID-19 class. In order to allow COVID-19 detection on different medical imaging procedures, authors in [1], use 5216 chest X-ray images and 420 Computer Tomography (CT) scans to classify an image as a normal or COVID-19 case using a neural network composed of Inception recurrent residual units. The model reports 84.67% in testing accuracy for X-ray images and 98.78% in CT-images. Aiming at improving the overall performance on a multi-class problem, transfer learning technique has been applied on MobileNet and VGG16 networks [2], from the domain of Imagenet dataset, a large scale set of images of common objects for classification, to chest X-ray images using 1428 samples labeled as Normal, bacterial pneumonia, viral pneumonia or COVID-19 cases. This model reached 98.66%, and 96.46% as sensitivity and specificity metrics, respectively.

In addition, some specialized CNN architectures have been designed to increase sensitivity to COVID-19 cases to limit the number of missed COVID-19 cases as much as possible. For instance, COVID-net [15] achieves precision and recall values of 98% and 91%, respectively, for detecting COVID-19 cases by using projection-expansion-projection-extension design pattern on its deep learning architecture. It also provides enhanced representational capacity while maintaining computational efficiency. This work set the foundations for future models and comparisons, as it was the first successful model built to detect COVID-19 related anomalies in the lungs on a large dataset. On the other hand, a deep neural network called COVID-SegNet is proposed in [16]. This solution is a segmentation-based approach capable of separating the pneumonia-area caused by COVID-19 from lungs in CT images. It follows a sequence of four convolutional encoder layers and three deconvolution-upsampling decoder layers. The work obtained a dice score of 98.7% for lung and 72.6% for COVID-19 segmentation.

3 Proposed Methodology

The proposed methodology consists of three main phases as shown in Fig. 1. Firstly, the X-ray images were collected for training and validation purposes. Then, an Xception-module-based CNN is trained to learn features from COVIDx dataset. Finally, the trained model is tested to determine its ability to correctly classify X-ray images into three classes: Normal, Pneumonia or COVID-19. In order to increase the overall accuracy achieved by the Xception model, two variations are applied over the proposed methodology: a) the X-ray images are pre-processed using benchmark image enhancement techniques to make details more visible, and b) batch balancing is used to cope the unbalance in distribution among class samples.

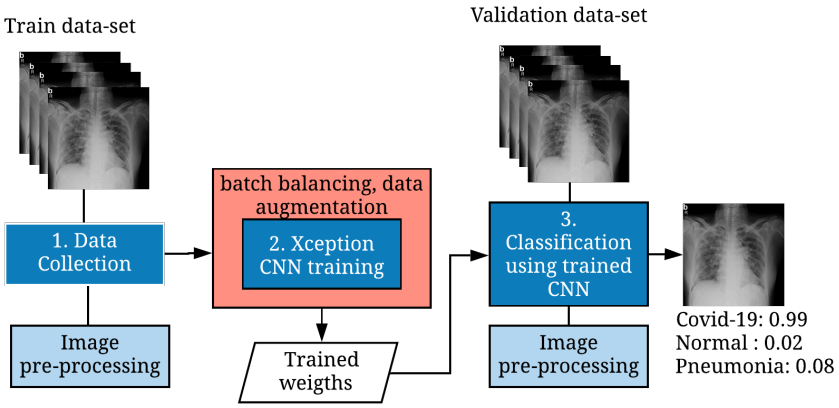


Fig. 1: Proposed methodology using an Xception-based CNN with pre-processing and batch balancing techniques.

1. **Data Collection:** This work uses COVIDx dataset [15]. It is a compilation of chest X-ray images from multiple sources. COVIDx dataset contains a total of 13.8K PNG and RGB images of normal, COVID-19 and pneumonia related cases. Images have different resolutions as 720×1024, 1024×720, 720×720 and 1024×1024. Table 1 shows image samples distribution per cases for training and testing purposes. As it can be seen, COVIDx includes less COVID-19 samples compared to the other ones. This unbalance is due to the novelty of the virus, and the limited amount of chest X-ray images related to this disease that are available.

Type	Normal	Pneumonia	COVID-19 Pneumonia	Total
Train	6049	4163	268	10479
Test	2017	1388	90	3495

Table 1: Number of X-ray samples related to Normal, Pneumonia and COVID-19 cases for training and testing purposes [15].

2. **Xception-block-based CNN:** Xception-module-based CNN [4] illustrated in Fig. 2 is compounded by 14 Xception modules (with linear residual connections between each module, except for the first and last), a global average pooling layer and finally a fully connected layer with softmax activation function for predictions. Xception modules extend the concept of separable convolutions, which reduces the number of parameters by applying a spatial filter per channel, instead of a 3D filter to the whole input.

The difference between original separable convolutions, implemented in networks like Inception [14], is that the point wise convolution is applied to the input first, before applying spatial filters to reduce the number of parameters, while maintaining performance.

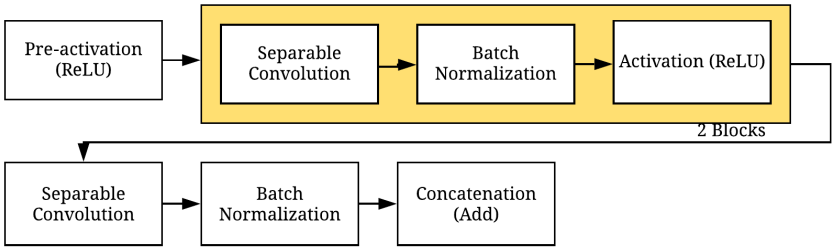


Fig. 2: Core Xception Module.

3. **Classification:** This phase uses the trained CNN, where the probability ($p(c_i|x_i)$) achieved by each output allows to classify each X-ray image (x_i) as Normal, Pneumonia or COVID-19 cases (c_i).

3.1 Methodology Variations

1. **Pre-processing based on unsharp masking technique:** Since COVIDx dataset is a compilation of chest X-rays from various sources, image detail, contrast and noise levels can vary. To mitigate this problem, unsharp masking technique is applied. It enhances the visual quality of an image by processing its edges starting with a Gaussian blur with a kernel of k×k and an standard deviation in the direction X of σ_x. This technique separates the edges, amplifies, and sums them back into the image. Then a weighted sum of the images is performed as shown equation 1:

$$UM = 1.5 * original - 0.5 * blurred \tag{1}$$

2. **Batch balancing and data augmentation:** This step aims to increase the size of a training dataset at each training iteration, generating batches with artificial versions of images. Artificial images are generated by applying horizontal flips, rotations and variations in brightness.

4 Experimental Setup

To evaluate the effectiveness of unsharp masking and batch balancing procedures on the overall accuracy achieved by Xception-module-based CNN, different experiments are conducted. Each one derives from the workflow in 1, where (a)

image enhancement or (b) batch balancing procedures can be present or not; resulting in four experiments as follows.

1. RNB: raw images, no balancing correction.
2. RB: raw images with balance correction.
3. EINB: enhanced images, no balance correction.
4. EIB: enhanced images, balance correction.

Xception-module-based CNN receives normalized input images of $224 \times 224 \times 3$, thus, each X-ray image is resized and multiplied by $1/255$ to normalized them.

For experimental purposes, python routines have been implemented using Google Colab with a GPU accelerator. This environment provides 12 GB of RAM and 100 GB ROM. Tensorflow framework with CUDA support was employed for neural network developing, training and testing. OpenCV was used for image pre-processing, meanwhile scikit-learn was used to compute metrics.

Before training, a grid search was performed to select a suitable set of hyper-parameters. The search space is depicted in table 2:

Hyper-parameter	**Value space**	**Description**
batch size	32, 64, 70	Number of samples per training iteration
epochs	10, 15, 20, 25, 30	Number of passes over the whole dataset
input image	(112,112,3), (224,224,3), (448,448,3)	image dimensions, channels included.
learning rate	0.01, 0.005, 0.001, 0.0005, 0.0001, exponential decay from 0.005 to 0.001	limits the weight update against gradients
optimizer	Adam, SGD	Optimization algorithm to update weights

Table 2: Hyper-parameter values for all experiments.

For high performance the hyper-parameters that achieve minimal loss were chosen: 64 samples batch size, 20 epochs, exponential decay from 0.005 to 0.0001 for learning rate, (224,224,3) for image shape and Adam as optimizer.

5 Experimental Results

Some of the outputs obtained for the enhanced images of the dataset are depicted in Fig.3b. It can be observed that the unsharp mask technique leads to a slight improvement on details, such as the bone structure and lungs, which edges are sharper. Also, these images appear more focused as opposed with the images on the Fig. 3a, which have more diffused edges.

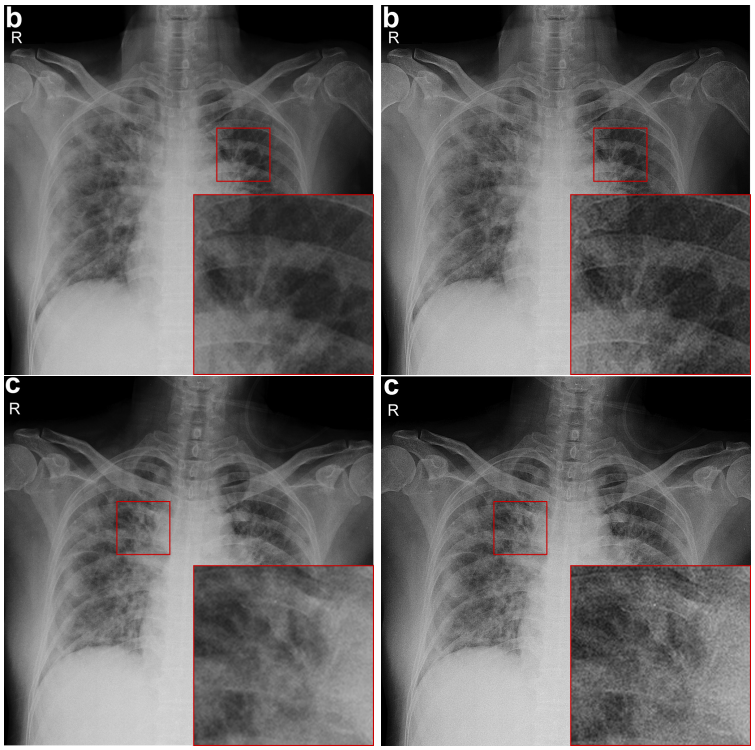
Accuracy (A), precision (P) and recall (R) given by equations 2, 3 and 4, respectively, were computed to measure the ability of the proposed methodology to classify X-ray images into the corresponding case, and to determine how the pre-processing and batch balancing techniques influence on the classification performance achieved.

$$A = \frac{T_P + T_N}{F_P + F_N + T_P + T_N} \tag{2}$$

$$P = \frac{T_P}{T_P + F_P} \tag{3}$$

$$R = \frac{T_P}{T_P + F_N} \tag{4}$$

T_P, T_N, F_P, and F_N refer to the total number of true positives, true negatives, false positives and false negatives, respectively, after comparing the predictions



(a) (b)

Fig. 3: COVID-19 chest X-ray. (a) Samples from COVIDx dataset; (b) Images after enhancement on (a).

of the CNN and the true labels on the test set. In this sense, accuracy quantifies the ratio of correctly classified X-ray images, precision quantifies the ratio of correctly classified positive cases over the total positive samples, and recall describes the proportion of positive cases retrieved by the CNN from the total number of positives samples.

To fulfill the experiments and validate that the models' performances are not due to chance, a 5-fold cross validation was performed by partitioning the dataset into five subsets of equal size, then training the model on four of the subsets and testing on the remaining one for five rounds. Each round had a different training and test set where the model was fitted from scratch. Finally, all metrics retrieved from each fold were averaged. Obtained results are exposed in Table 3. It is important to note that performance metrics from COVID-net [15] (a Deep Convolutional Neural Network designed to detect COVID-19 related anomalies on chest X-rays), are included since both studies use the same dataset.

Class	metric	Experiment				
		RNB	RB	IENB	IEB 4	COVID-net
Normal	Precision	91.8	**93.8**	93.4	93.4	90.5
	Recall	88.0	**95.4**	91.6	95.2	95.0
Pneumonia	Precision	**95.4**	93.2	89.8	92.0	91.3
	Recall	83.4	90.0	87.8	89.8	**94.0**
COVID-19	Precision	55.4	82.2	75.0	88.0	**98.9**
	Recall	85.6	84.4	76.0	80.0	**91.0**
All	Accuracy	85.2	92.8	89.8	92.8	**93.3**

Table 3: Table of results for all experiments and COVID-net as reported in [15].

Table 3 demonstrates that image enhancement and batch balance techniques guarantee higher precision, recall and accuracy on almost all test cases, compared to raw images and no balancing correction. However, the proposed methodology, mainly for COVID-19, reaches (5%) lower overall accuracy with RB and IEB experiments with respect to COVIDNet [15]. This results suggest that more improvements on the data and the CNN model should be explored.

In addition, the receiver operating characteristic (ROC) curve exhibits the performance reached by plotting recall vs false-positive rate at various prediction thresholds [17]. It reveals the capability of a model to distinguish between classes. From Fig 4, it can be seen that when both variations (image enhancement and batch balance) are present, the model achieves the highest area under the curve from all experiments, demonstrating that the best results in class separability are achieved with the proposed methodology.

It was noticed, when utilizing classic augmentation techniques, as no new patterns are introduced into the training [10] and noise from rotations, translations and brightness changes are added, the benefits of the enhancement applied in this work are reduced. Hence, other augmentation techniques such as neural style transfer and generative adversarial networks [6, 7, 13] should be explored.

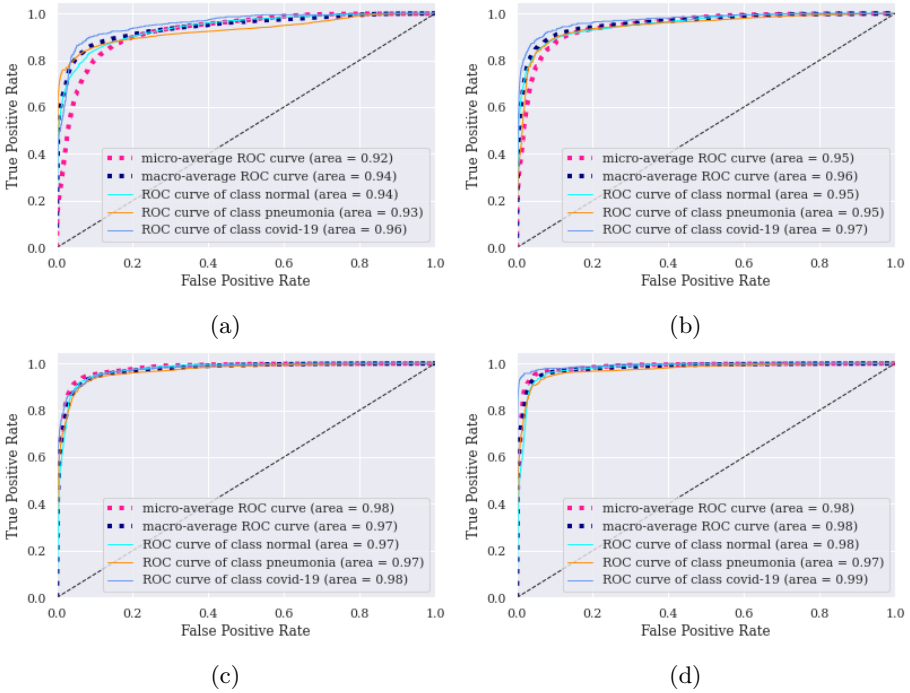


Fig. 4: ROC curves for (a) Raw and imbalanced image samples on training, (b) Raw, balanced image samples, (c) Enhanced, but imbalanced image samples, (d) Enhanced and balanced image samples.

6 Conclusion

The Xception-based models that implemented batch balancing performed better than those that did not, and although combining this procedure with image enhancement seemed not to improve the metrics on Table 3, the ROC curves showed that the models achieve a higher level of class separability when predicting on this setting. It is important to remark that COVID-Net classifier achieved better performance, but this neural network was specifically designed for the COVIDx dataset, while the model used in this work, Xception, is more general. Besides, this work used a considerably bigger test sets in a 5-fold cross validation, giving more fidelity on its results. As future work, new and improved CNN architectures along with other class imbalance optimizations must be used to further increase precision and recall values for COVID-19 class. Besides, a better approach for data augmentation must be explored. Generative adversarial neural networks, and style transfer have shown promising results creating synthetic images for a specific class [7], [13], [6].

Bibliography

[1] Alom, M.Z., Rahman, M.M.S., Nasrin, M.S., Taha, T.M., Asari, V.K.: COVID$_MTNet$: $COVID - 19DetectionwithMulti - TaskDeepLearningApproaches(apr2020), http : //arxiv.org/abs/2004.03747$

[2] Apostolopoulos, I.D., Mpesiana, T.A.: Covid-19: automatic detection from X-ray images utilizing transfer learning with convolutional neural networks. Physical and Engineering Sciences in Medicine **43**(2), 635–640 (jun 2020). https://doi.org/10.1007/s13246-020-00865-4, http://link.springer.com/10.1007/s13246-020-00865-4

[3] Ayan, E., Unver, H.M.: Diagnosis of Pneumonia from Chest X-Ray Images Using Deep Learning. In: 2019 Scientific Meeting on Electrical-Electronics Biomedical Engineering and Computer Science (EBBT). pp. 1–5. IEEE (apr 2019). https://doi.org/10.1109/EBBT.2019.8741582, https://ieeexplore.ieee.org/document/8741582/

[4] Chollet, F.: Xception: Deep Learning with Depthwise Separable Convolutions (oct 2016), http://arxiv.org/abs/1610.02357

[5] Cohen, J.P., Morrison, P., Dao, L.: Covid-19 image data collection. arXiv 2003.11597 (2020), https://github.com/ieee8023/covid-chestxray-dataset

[6] Frid-Adar, M., Diamant, I., Klang, E., Amitai, M., Goldberger, J., Greenspan, H.: GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification. Neurocomputing **321**, 321–331 (dec 2018). https://doi.org/10.1016/j.neucom.2018.09.013, https://linkinghub.elsevier.com/retrieve/pii/S0925231218310749

[7] Han, C., Murao, K., Satoh, S., Nakayama, H.: Learning More with Less: GAN-based Medical Image Augmentation (mar 2019), http://arxiv.org/abs/1904.00838

[8] Kassani, S.H., Kassani, P.H., Wesolowski, M.J., Schneider, K.A., Deters, R.: Breast Cancer Diagnosis with Transfer Learning and Global Pooling (sep 2019), http://arxiv.org/abs/1909.11839

[9] Liang, G., Hong, H., Xie, W., Zheng, L.: Combining Convolutional Neural Network With Recursive Neural Network for Blood Cell Image Classification. IEEE Access **6**, 36188–36197 (2018). https://doi.org/10.1109/ACCESS.2018.2846685, https://ieeexplore.ieee.org/document/8402091/

[10] Mikołajczyk, A., Grochowski, M.: Data augmentation for improving deep learning in image classification problem. pp. 117–122 (05 2018). https://doi.org/10.1109/IIPHDW.2018.8388338

[11] 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. Computers in Biology and Medicine **121**, 103792 (jun 2020). https://doi.org/10.1016/j.compbiomed.2020.103792, https://linkinghub.elsevier.com/retrieve/pii/S0010482520301621

[12] Redmon, J., Farhadi, A.: YOLOv3: An Incremental Improvement (apr 2018), http://arxiv.org/abs/1804.02767
[13] Sandfort, V., Yan, K., Pickhardt, P.J., Summers, R.M.: Data augmentation using generative adversarial networks (CycleGAN) to improve generalizability in CT segmentation tasks. Scientific Reports **9**(1), 16884 (dec 2019). https://doi.org/10.1038/s41598-019-52737-x, http://www.nature.com/articles/s41598-019-52737-x
[14] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z.: Rethinking the Inception Architecture for Computer Vision (dec 2016), http://arxiv.org/abs/1512.00567
[15] Wang, L., Wong, A.: Covid-net: A tailored deep convolutional neural network design for detection of covid-19 cases from chest radiography images (2020)
[16] Yan, Q., Wang, B., Gong, D., Luo, C., Zhao, W., Shen, J., Shi, Q., Jin, S., Zhang, L., You, Z.: COVID-19 Chest CT Image Segmentation – A Deep Convolutional Neural Network Solution (apr 2020), http://arxiv.org/abs/2004.10987
[17] Zou, K.H., O'Malley, A.J., Mauri, L.: Receiver-Operating Characteristic Analysis for Evaluating Diagnostic Tests and Predictive Models. Circulation **115**(5), 654–657 (feb 2007). https://doi.org/10.1161/CIRCULATIONAHA.105.594929, https://www.ahajournals.org/doi/10.1161/CIRCULATIONAHA.105.594929