2021 9th European Workshop on Visual Information Processing (EUVIP) | 978-1-6654-3230-6/21/$31.00 ©2021 IEEE | DOI: 10.1109/EUVIP50544.2021.9484001

MACHINE LEARNING SCREENING OF COVID-19 PATIENTS BASED ON X-RAY IMAGES FOR IMBALANCED CLASSES

ILYES MRAD[1]*, RIDHA HAMILA*[1]*, AIMAN ERBAD*[2]*, TAHIR HAMID*[3]*, RASHID MAZHAR*[3]*, NASSER AL-EMADI*[1]

[1]Qatar University, Doha, Qatar, [2]Hamad Bin Khalifa University, Doha, Qatar, [3]Hamad Medical Corporation, Doha, Qatar

ABSTRACT

COVID-19 is a virus that has infected more than one hundred and fifty million people and caused more than three million deaths by 13th of Mai 2021 and is having a catastrophic effect on the world population's safety. Therefore, early detection of infected people is essential to fight this pandemic and one of the main screening methods is radiological testing. The goal of this study is the usage of chest x-ray images (CXRs) to effectively identify patients with COVID-19 pneumonia. To achieve an efficient model, we combined three methods named: Convolution Neural Network (CNN), transfer learning, and the focal loss function which is used for imbalanced classes to build 3 binary classifiers, namely COVID-19 vs Normal, COVID-19 vs pneumonia and COVID-19 vs Normal Pneumonia (Normal and Pneumonia). A comparative study has been made between our proposed classifiers with well-known classifiers and provided enhanced results in terms of accuracy, specificity, sensitivity and precision. The high performance of this computer-aided diagnostic technique may greatly increase the screening speed and reliability of COVID-19 detection.

***Index Terms*—** COVID-19, chest X-ray images, convolutional neural network, focal loss function.

1. INTRODUCTION

COVID-19 is an outbreak and was proclaimed as a pandemic given the severity of its propagation worldwide [1]. It has been noted that successful diagnosis for the infected persons are a sobbing need to battle the expansion of COVID-19 virus. In fact, the most reliable diagnostic test used to identify COVID-19 cases is Reverse Transcriptase Polymerase Chain Reaction (RT-PCR) examination, which can detect Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) from respiratory specimens [2]. Although PCR screening is attractive due to its high efficiency, it is a manual, complicated and time-consuming process [2]. Therefore, it is causing delays in preventive care efforts. Also, the RT-PCR pack costs approximately 120-130 USD and also necessitates a lab in order to hold PCR equipment, that could cost 15,000-90,000 USD [3]. Such costly and delayed test results are contributing to the transmission of the virus. Although standard CXR images are capable to identify suspicious cases, there is a need to differentiate COVID-19 patients from other pneumonia cases and radiologists have difficulties in this diagnostic [3]. The COVID-19 symptoms are very similar to viral pneumonia which may contribute to an incorrect detection at the crowded emergency services [3, 4].

Several related works reported good results with an accuracy over 85% [5, 6, 7, 8], but there are two important points which they did not focus on which are:

- COVID-19 is a novel disease, its number of chest x-ray images is so far less than other pneumonia and normal cases. Therefore, we have to consider this imbalance between classes which is not taken into account.
- Several works achieved more than 99% accuracy, but this is not the case for sensitivity which is a very important criterion. For example, Shuai *et al.* [5] reported that the external testing dataset shows an accuracy of about 80%, but a sensitivity of 67% which is too low.

In this work, we propose a deep CNN-based transfer learning method for identifying COVID-19 patients utilizing imbalanced data of CXR images. We will investigate three most important problems: 1) How to build a model capable of detecting COVID-19 cases without the need of human intervention. 2) How to deal with the imbalanced database 3) How to enhance the outcomes of a model that has already provided excellent results.
In summary, the main contributions of this paper are presented as follows:

- We propose three classifiers which are COVID-19 vs normal, COVID-19 vs pneumonia and COVID-19 vs Normal Pneumonia. We comprehensively consider the high imbalanced database as our target.
- We have succeeded in obtaining excellent performance in terms of accuracy and especially of perfect sensitivity compared to recent works.

2021 9th European Workshop on Visual Information Processing (EUVIP)
23-25 June, 2021, Paris, France

2. RELATED WORKS

Many deep learning-based artificial intelligence (AI) solutions have been developed and tests have shown that they are very effective in identifying signs of COVID-19 utilizing X-ray images. Recent advances in the field of deep learning, especially CNNs [9], have shown significant improvement in the classification of images. Besides the challenges of computer vision, CNNs have had great results in solving medical problems such as detection of breast cancer, detection of heart disease and diagnosis of Alzheimer's disease[10]. Numerous researchers have recently proposed deep-learning approaches focused on COVID-19 pneumonia detection. Shuai *et al.* [5] utilized 325 CT images of COVID-19 patients and 740 computed tomography (CT) images of viral pneumonia cases to detect COVID-19 cases. Inception [11] was used to build the algorithm. They achieved 89.5 %, 88 % and 87 % of accuracy, specificity and sensitivity, respectively. Ayrton *et al.* [7] used a limited dataset of 339 images with only 39 of positive cases for training, validation and testing using ResNet50 as a pre-trained model and validation accuracy of 96.2 %. Pedro *et al.* Asma *et al.* [8] introduced a new CNN network, called CNN-COVID based on a deep CNN, called Decompose, Transfer, and Compose (DeTraC) to screen COVID-19 cases. Mohamed *et al.* [12] used an amended Manta-Ray Foraging Optimization to choose the most important features and a dataset containing 216 COVID-19 positive images and 1,675 negative COVID-19 images. The proposed model achieved an accuracy rate of 96.09%.

3. SYSTEM MODEL

This section highlights the necessary step to move from Cross-Entropy (CE) loss function to the α-balanced variant of the focal loss (FL) function as shown in Figure 1. The FL considered by Tsung-Yi *et al.* [13] is aimed at addressing the scenario of one-stage object detection where there is a radical class imbalance and very few samples to train and even fewer to validate.

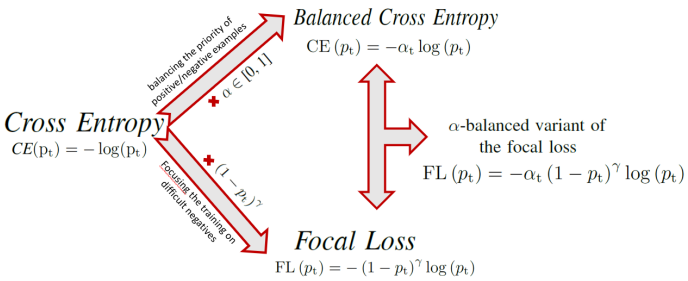

Fig. 1. The necessary steps to obtain the α-balanced variant of the focal loss function.

3.1. Cross Entropy

Before presenting the FL for binary classification, we begin with the CE loss defined by:

$$CE(p, y) = \begin{cases} -\log(p) & \text{if } y=1 \\ -\log(1-p) & otherwise \end{cases} \quad (1)$$

where, $y \in \{1,-1\}$ states the ground-truth class and p $\in$ [0, 1] is the probability of the estimated model for the label $y = 1$ [13]. Also, p_t was introduced as :

$$p_t = \begin{cases} p & \text{if } y = 1 \\ 1-p & \text{otherwise} \end{cases} \quad (2)$$

$$So, CE(\mathrm{p}; \mathrm{y}) = -\log(\mathrm{p_t}). \quad (3)$$

3.2. Balanced Cross Entropy

Introducing of a weighting factor $\alpha \in [0, 1]$ for category 1 and (1-α) for category (-1) is considered a common approach to fix class imbalances [13]. α could be set by the opposite class. The loss of the CE in α-balance could be written as:

$$\mathrm{CE}(p_t) = -\alpha_t \log(p_t). \quad (4)$$

3.3. Focal Loss Definition

Although α balances the priority of positive/negative examples, it does not differentiate between clear/difficult examples. Therefore, it is recommended to reshape the loss function to lighten the easy examples by focusing the training on difficult negatives by multiplying the CE by a modulating factor $(1-p_t)^\gamma$ [13]. Thus, the FL is given by:

$$\mathrm{FL}(p_t) = -(1-p_t)^\gamma \log(p_t). \quad (5)$$

In practice, the α-balanced variant of the FL is used as:

$$\mathrm{FL}(p_t) = -\alpha_t (1-p_t)^\gamma \log(p_t). \quad (6)$$

3.4. Methodology and Database Description

3.4.1. Methodology

Each pixel extracted from the image is a valuable variable. Thus, we got a large number of features due to the image size. Indeed, the resulting large number of neurons will be reduced by the convolution operation helping in decreasing the number of parameters. However, CNN performs better with a larger dataset. So, the model requires a huge dataset to extract significant features and achieve a decent model. Transfer learning can be a remedy for CNN applications when the dataset is not large enough. While there was a significant number of patients diagnosed with COVID-19, the number of freely accessible online images of chest x-rays images were limited and dispersed. Figure 2 shows one image from each

group of the database which are normal, COVID-19 pneumonia, viral pneumonia and bacterial pneumonia group. In fact, in the proposed model part, we used 4 pre-trained models to choose the best and use it in the model improvement part, which results in improvement by image augmentation and improvement by tuning the focal loss hyperparameters as shown in Figure 3.

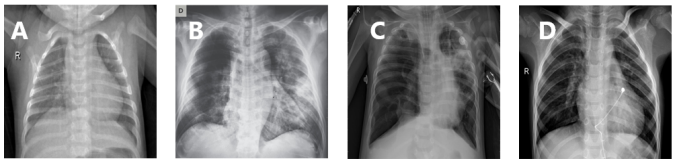

Fig. 2. One chest X-ray image from each class of the dataset: (A) shows a normal case, (B) shows a COVID-19 case, (C) shows Viral Pneumonia case and (D) shows a bacterial pneumonia case.

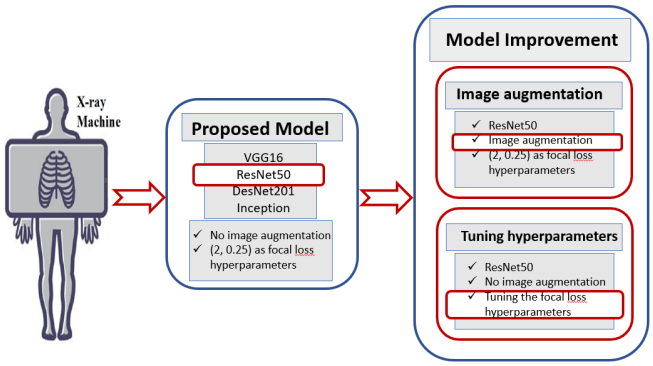

Fig. 3. Proposed COVID-19 detection scheme.

3.4.2. Database Description

The database used in this study was prepared from several different sub-databases. Italian Society of Medical and Interventional Radiology (SIRM) COVID-19 DATABASE [14] contains 71 confirmed COVID-19 cases. Novel Corona Virus 2019 Database in GitHub [15] contains 250 COVID-19 chest X-ray images. Also, from different articles, we collected 60 COVID-19 positive chest X-ray images from 43 new articles [16] and 32 positive CXR images from Radiopaedia [17]. These CXR images were not listed in the GitHub. On the other hand, Kaggle chest X-ray database [18] contains 2561 images of bacterial pneumonia cases, 1345 images of viral pneumonia and 1341 images of normal cases.

4. PROPOSED MODEL

This section elaborates on the implementation steps of the proposed model such as the pre-processing phase and used parameters with discussions of the obtained results.

4.1. Algorithm Selection

In this section, four known pre-trained deep learning CNN models were used, namely VGG16 [11], ResNet50 [11], DenseNet201 [19] and Inception [11], to differentiate between classes using binary classifications.

4.2. Image distribution

Table 1 summarized the distribution of images into training, validation and test set. There are 400 images of COVID-19 cases, 1340 of Normal patients, 1340 of viral pneumonia patients and 2560 of bacterial pneumonia patients.

Table 1. Details of training, validation and test set.

Type of dataset	No. of chest X-ray images	Training set	Validation set	Test set
COVID-19	400	320	40	40
NORMAL	1340	1072	134	134
VIRAL PNEUMONIA	1340			
BACTERIAL PNEUMONIA	2560			
PNEUMONIA	3900	3120	390	390
COVID-19 vs NormalPneumonia	5240	4192	524	524

4.3. Pre-processing

Four ML algorithms were trained and evaluated using the CXR images with and without image augmentation. RMSprop with 0.85 momentum, 10 mini-batches and a learning rate of 0.0001 were used in this study. EarlyStopping with patience equal to 5 and the class weight were set for each classifier. Finally, to evaluate the proposed model, we used 10-fold cross-validation scheme to avoid overfitting. We set 80% of the dataset for training, 10% for validation and 10% for the test. Regarding hyperparameters of the focal loss function, we set (α, γ) at (0.25,2) as in RetinaNet Detector [13]. One of the crucial phases of data pre-processing is image resizing of the image input. For inceptionv3 and VGG16, the images had been resized to 227x227 pixels, while for DenseNet201 and ResNet50, images had been resized to 224x224 pixels.

4.4. Results and Performance

The performance evaluation matrix aims to assess the efficiency of various deep learning algorithms for three classifications. The evaluation of the three models is performed utilizing 5 metrics which are the average accuracy in %, sensitivity, specificity, precision and F1-score.

4.4.1. Results

As we can see from the results shown in Table 2 that VGG16 showed the best performance for COVID-19 vs normal classifier, but ResNet50 gave the best performance while deal-

Table 2. Average of different performance metrics for the three classifiers using four pre-trained models.

Task	Pretrained Model	Accuracy	Specificity	Sensitivity	Precision	F1 score
COVID-19 vs normal	ResNet50	99.48	0.99	0.97	0.97	0.97
	VGG16	99.88	0.99	0.99	0.99	0.99
	Inception	90.18	0.94	0.78	0.83	0.80
	DenseNet201	95.12	0.97	0.88	0.91	0.89
COVID-19 vs Pneumonia	ResNet50	99.18	0.99	0.93	0.98	0.95
	VGG16	97.93	0.98	0.87	0.91	0.89
	Inception	96.88	0.98	0.85	0.88	0.86
	DenseNet201	97.88	0.96	0.87	0.90	0.88
COVID-19 vs Normal & Pneumonia	ResNet50	99.80	0.99	0.98	0.99	0.98
	VGG16	98.97	0.99	0.94	0.91	0.92
	Inception	96.27	0.98	0.72	0.75	0,73
	Densenet201	98,66	0.99	0.91	0.91	0.91

ing with COVID-19 vs pneumonia and COVID-19 vs Normal Pneumonia. VGG16 was producing the highest accuracy and sensitivity of 99.88% and 0.99, respectively, for the first classification schemes, but RestNet50 provided the highest accuracy and sensitivity of 99.18% and 0.93, respectively, for the classifier of COVID-19 vs pneumonia and 99.80% and 0.98, for COVID-19 vs Normal and Pneumonia classifier. However, RestNet50 outperformed other algorithms and Table 2 summarized the performance of the three classifiers using the four pre-trained models.

4.5. Discussion

We notice that it was easy to differentiate between normal and COVID-19 cases, which is illustrated in Table 2. We also notice that several cases of pneumonia are detected by Normal Pneumonia classifier, but COVID-19 vs pneumonia classifier made few wrong classifications. The miss-classification of these few cases could be explained by two reasons. At first, by the genetics and the morphology of patients where the symptoms did not appear clearly in many cases. The second reason is the early diagnostic test made by a few patients.

5. MODEL IMPROVEMENT

This section details the experimental setup and the obtained results of two improved methods by using image augmentation and adjustment of focal loss hyperparameters.

5.1. Image augmentation

5.1.1. Pre-processing & image distribution

Table 3 summarizes the distribution of images without image augmentation into training, validation and test set. The Bacterial pneumonia class containing the maximum number of images was used as it is and only the other classes were augmented to reach about 2500 images. Additionally, the pretrained ResNet50 model that provided the best results in previous experiments is selected with the same parameters.

Table 3. Details of training, validation and test set after using image augmentation.

Type of dataset	No. of images before image aug.	No. of images after image aug.	Without augmentation			After augmentation		
			Training set	Valid. set	Test set	training set	Valid. set	Test set
COVID-19	400	2640	320	40	40	2560	40	40
NORMAL	1340	2412	1072	134	134	2144	134	134
VIRAL PNEUMONIA	1340	2421						
BACT. PNEUMONIA	2560	2560						
PNEUMONIA	3900	4972	3120	390	390	4192	390	390
NORMAL & PNEUMONIA	5240	7384	4192	524	524	6336	524	524

5.1.2. Results

Table 4. Different performance metrics for three classifiers after using image augmentation.

TASK	Pretrained model	Accuracy	Specificty	Sensitivity	Precision	F1 score
COVID-19 vs Normal	ResNet50	99,88	1.0	0.99	1.0	0.99
COVID-19 vs Pneumonia	ResNet50	99.51	0.99	0.98	0.99	0.98
COVID-19 vs NormalPneumonia	ResNet50	99.16	0.98	0.91	0.96	0.93

When image augmentation was used, it was observed from the Table 4 that the model produced an accuracy and a sensitivity of 99.88% and 0.99, respectively, for the classifier of COVID-19 vs normal, 99.51% and 0.98 for COVID-19 vs pneumonia and 99.16% and 0.91 for the classifier of COVID-19 vs Normal Pneumonia. The performance matrix was significantly improved with image augmentation for COVID-19 vs pneumonia classifier but lowered in the third classification which is COVID-19 vs Normal Pneumonia.

5.1.3. Discussion

The performance matrix shown in Table 4 of COVID-19 vs Normal Pneumonia classifier was lowered. That could be

in relationship with hyperparameters of the focal loss function which are fixed. As we know each dataset must have its hyperparameters and in particular for the focal loss function associated to it in addition to the imbalance rate. For example, in pneumonia and COVID-19 dataset, the number of pneumonia samples is 9.71 times higher than COVID-19 samples without image augmentation and only 2 times higher after using image augmentation. In the third classification, the number of Normal Pneumonia samples was 13.08 times higher than COVID-19 patients without image augmentation and 3 times higher after applying image augmentation. Consequently, if data augmentation is produced without changing hyperparameters, not all classifiers will be improved. Additionally, the performance matrix of COVID-19 vs pneumonia classifier was improved.

5.2. Tuning the focal loss hyperparameters

5.2.1. Pre-processing & Distribution of images

In this part, ResNet50 was the only pre-trained deep CNN algorithm used to show the impact of (α, γ) on the accuracy and other performance metrics of the same three binary classifications with respect to α and γ. Also, the second objective of this part is to improve the performance of the proposed model. In RetinaNet Detector [13], the best pair of hyperparameters (α, γ) found is (2, 0.25). First, γ will be varied in $[0, 5]$ for fixed α equal to 0.25. Then, α will be varied in $]0, 1]$ for fixed γ equal to 2. Additionally, the images have been resized to 224×224 pixels.

5.2.2. Result

In this part, the best classifiers' performances in terms of sensitivity as the first criterion and accuracy as the second criterion are presented in Table 5.

Table 5. Different performance metrics for the three classifiers after tuning the focal loss hyperparameters.

Task	Alpha	Sigma	Accuracy	Specificity	Sensitivity	Precision	F1 score
COVID-19 vs Normal	0.25	0.5	99.94	0.99	1.0	0.99	0.99
		1.5					
		3.5					
		4.5					
	0.1	2					
	0.375						
	0.625						
COVID-19 vs Pneumonia	0.25	0.5	99.39	0.99	0.97	0.97	0.97
COVID-19 vs Normal &Pneumonia	0.25	2.5	99.98	0.99	1.0	0.99	0.99
	2	0.75					

Table 5 shows that there were several hyperparameters pairs for COVID-19 vs Normal classifier and two pairs for COVID-19 vs Normal Pneumonia classifier that produced high accuracy with high sensitivity. On the other hand, the performance of COVID-19 vs Pneumonia was improved too and reached 99.39% of accuracy and 0.97 of sensitivity, but its performance was even worse compared to image augmentation method with an accuracy of 99.51% and a sensitivity of 0.98.

5.3. Discussion

Our adopted approach of focal loss hyperparameters tuning was shown to improve the performance of the proposed model and reach accurate results for COVID-19 vs normal and COVID-19 vs normal and pneumonia classifier which is the most important classifier, since it can detect COVID-19 patient among all other cases.

As we mentioned earlier, α balances the priority of positive/negative samples and γ concentrates training on difficult negatives. The obtained results can not be generalized for all pairs of (α, γ). For the first classifier of COVID-19 vs Normal, we noticed from Table 5 that several pairs gave accurate results. Therefore, the priority of the first-class over the second given by α is not important. On the other hand, the existence of several values of γ means that there were no difficult samples, that could be confirmed by the good results obtained with the four pre-trained models. As for COVID-19 vs Normal Pneumonia classifier, two pairs gave accurate results. The first couple was (0.25, 2.5) where COVID-19 group was set as 3 times more important than Normal Pneumonia group and easy negatives down-weighted by 2.5 times. The second couple was (0.75, 2). The value of α which is 0.75, was considered as important as that of the α related to the first couple (α and (1-α) give the same importance).

6. CONCLUSION & PERSPECTIVES

In this work, we choose a CNN-based transfer learning approach for automated identification of COVID-19 utilizing chest x-ray images collected from COVID-19, normal and pneumonia patients.

This work emphasized the utility of using transfer learning approaches when the dataset used was limited to detect COVID-19 images. Also, it was possible to improve the classification efficiency of different CNN models by increasing the number of samples in the dataset. In light of our outcomes, this could help medical professionals in their clinical decision-making.

7. ACKNOWLEDGEMENT

This work was supported by Qatar University Internal Grant IRCC-2020-001. The statements made herein are solely the responsibility of the authors.

8. REFERENCES

[1] "Who director-generals opening remarks at the media briefing on covid-19 - 11 march 2020," *W. H. Organization*, 2020. [Online]. Available: shorturl.at/hmDQX

[2] W. Wang, Y. Xu, R. Gao, R. Lu, K. Han, G. Wu, and W. Tan, "Detection of sars-cov-2 in different types of clinical specimens," *Jama*, vol. 323, no. 18, pp. 1843–1844, 2020.

[3] M. E. Chowdhury, T. Rahman, A. Khandakar, R. Mazhar, M. A. Kadir, Z. B. Mahbub, K. R. Islam, M. S. Khan, A. Iqbal, N. Al-Emadi *et al.*, "Can ai help in screening viral and covid-19 pneumonia?" *arXiv preprint arXiv:2003.13145*, 2020.

[4] M. L. Holshue, C. DeBolt, S. Lindquist, K. H. Lofy, J. Wiesman, H. Bruce, C. Spitters, K. Ericson, S. Wilkerson, A. Tural *et al.*, "First case of 2019 novel coronavirus in the united states," *New England Journal of Medicine*, 2020.

[5] S. Wang, B. Kang, J. Ma, X. Zeng, M. Xiao, J. Guo, M. Cai, J. Yang, Y. Li, X. Meng *et al.*, "A deep learning algorithm using ct images to screen for corona virus disease (covid-19)," *MedRxiv*, 2020.

[6] L. Wang and A. Wong, "Covid-net: A tailored deep convolutional neural network design for detection of covid-19 cases from chest x-ray images," *arXiv preprint arXiv:2003.09871*, 2020.

[7] A. S. Joaquin, "Using deep learning to detect pneumonia caused by ncov-19 from x-ray images." [Online]. Available: shorturl.at/hlCPU

[8] P. M. de Sousa, M. Modesto, G. M. Pereira, C. A. da Costa Junior, L. V. de Moura, C. Mattjie, P. C. Carneiro, A. M. M. da Silva, and A. C. Patrocinio, "Covid-19 classification in x-ray chest images using a new convolutional neural network: Cnn-covid," 2020.

[9] S. Kiranyaz, M. Zabihi, A. B. Rad, A. Tahir, T. Ince, R. Hamila, and M. Gabbouj, "Real-time pcg anomaly detection by adaptive 1d convolutional neural networks," *arXiv preprint arXiv:1902.07238*, 2019.

[10] K. Kallianos, J. Mongan, S. Antani, T. Henry, A. Taylor, J. Abuya, and M. Kohli, "How far have we come? artificial intelligence for chest radiograph interpretation," *Clinical radiology*, vol. 74, no. 5, pp. 338–345, 2019.

[11] "Resnet, alexnet, vggnet, inception: Understanding various architectures of convolutional networks." [Online]. Available: https://cv-tricks.com/cnn/understandresnet-alexnet-vgg-inception/

[12] M. A. Elaziz, K. M. Hosny, A. Salah, M. M. Darwish, S. Lu, and A. T. Sahlol, "New machine learning method for image-based diagnosis of covid-19," *Plos one*, vol. 15, no. 6, p. e0235187, 2020.

[13] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2980–2988.

[14] *Covid-19 Database of Italian Society of Medical and Interventional Radiology.* [Online]. Available: https://www.sirm.org/category/senza-categoria/covid-19/

[15] J. P. Cohen, P. Morrison, L. Dao, K. Roth, T. Q. Duong, and M. Ghassemi, "Covid-19 image data collection: Prospective predictions are the future," *arXiv 2006.11988*, 2020. [Online]. Available: https://github.com/ieee8023/covid-chestxray-dataset

[16] *COVID-19 CXR database.* [Online]. Available: https://www.kaggle.com/tawsifurrahman/covid19-radiography-database

[17] *Chest X-ray images database of Covid-19 Pneumonia cases Radiopedia*. [Online]. Available: shorturl.at/cR125

[18] *(pneumonia and normal cases*. [Online]. Available: https://www.kaggle.com/paultimothymooney/chest-xray-pneumonia

[19] F. Iandola, M. Moskewicz, S. Karayev, R. Girshick, T. Darrell, and K. Keutzer, "Densenet: Implementing efficient convnet descriptor pyramids," *arXiv preprint arXiv:1404.1869*, 2014.