2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)
April 13-16, 2021, Nice, France

2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI) | 978-1-6654-1246-9/20/$31.00 ©2021 IEEE | DOI: 10.1109/ISBI48211.2021.9434159

XPGAN: X-RAY PROJECTED GENERATIVE ADVERSARIAL NETWORK FOR IMPROVING COVID-19 IMAGE CLASSIFICATION

Tran Minh Quan[†‡], *Huynh Minh Thanh*[†], *Ta Duc Huy*[†], *Nguyen Do Trung Chanh*[†‡], *Nguyen Thi Phuong Anh*[†], *Phan Hoan Vu*[†], *Nguyen Hoang Nam*[†], *Tran Quy Tuong*[#], *Vu Minh Dien*[⋆], *Bui Van Giang*[∗], *Bui Huu Trung, and Steven Quoc Hung Truong*[†]

[†]VinBrain, Vietnam [‡]VinUniversity, Vietnam [#]Ministry of Health, Vietnam
[⋆]National Hospital of Tropical Diseases, Vietnam [∗]National Cancer Hospital, Vietnam

ABSTRACT

This work aims to fight against the current outbreak pandemic by developing a method to classify suspected infected COVID-19 cases. Driven by the urgency, due to the vastly increased number of patients and deaths worldwide, we rely on situationally pragmatic chest X-ray scans and state-of-the-art deep learning techniques to build a robust diagnosis for massive screening, early detection, and in-time isolation decision making. The proposed solution, X-ray Projected Generative Adversarial Network (XPGAN), addresses the most fundamental issue in training such a deep neural network on limited human-annotated datasets. By leveraging the generative adversarial network, we can synthesize a large amount of chest X-ray images with prior categories from more accurate 3D Computed Tomography data, including COVID-19, and jointly train a model with a few hundreds of positive samples. As a result, XPGAN outperforms the vanilla DenseNet121 models and other competing baselines trained on the same frontal chest X-ray images.

Index Terms— COVID-19, Classification, Generative Adversarial Network, Chest X-ray, Digitally Reconstructed Radiographs.

1. INTRODUCTION

A standard golden method to diagnose COVID-19 is Reverse Transcription-Polymerase Chain Reaction (RT-PCR). However, due to the sampling collection procedure, this method may not capture well the appearance of COVID-19. Therefore, from filtering, classification, and detection of COVID-19 to examinations and treatments, all suffer from the contagious properties of viruses and pose considerable challenges while being applied on a massive scale. Studies and reports worldwide show that COVID-19 has various clinical manifestations, ranging from asymptomatic infection or just as a common cold to severe illnesses that cause acute respiratory damage, multiple organ failure and can lead to death if not treated promptly. At present, using the RT-PCR molecular biology test to look for specific genes of the virus is a valid test to confirm the diagnosis of infection with a sensitivity of 60% - 70% and a specificity of 95% - 100% [1]. Chest X-ray (CXR) and Computed Tomography (CT) play a particularly important role in screening and diagnosis suggestions. Besides, recent studies also show the essential values of CXR and CT in the diagnosis. CXR diagnosis specificity is 69% [2], and chest CT can be up to 98% [3]. Chest CT is not only valuable in the diagnosis of COVID-19 but also significant in monitoring disease progression and evaluating treatment effects [4, 5, 6].

Medical image-assisted diagnostic tools such as X-ray and CT, alongside RT-PCR, become essential to examine the crowd. Among them, CXR tends to be feasible due to its quick scanning time and sterilization. CXR is one of the most popular diagnostic imaging procedures globally, estimating roughly two billion scans per year. Nevertheless, the image features or indicators of COVID-19 symptoms on CXR can be missed because of various contrasts and scanning angles; or due to the radiologists' reading (mainly noisy from years of experiences, domains of expertise). These drawbacks can be avoided by using deep neural networks that learn statistically from the data and perform consistently as long as enough image samples are trained. However, as the CXR dataset we have collected is highly unbalanced and limited, we propose a novel generative deep learning-based method, called X-ray Projected Generative Adversarial Network (XPGAN), to resolve these challenges. XPGANs addresses the mentioned fundamental issues by projecting 3D CT volumes with golden ground truth labels to generate more realistic X-ray images and use them all for training such a deep learning model to classify COVID-19 chest X-ray images.

In summary, the proposed method aims to produce many more reliable and differentiable X-ray images which have golden labels from CT scans but are still similar to the direct ray-tracing projections (presumably leveraging constraint of per-pixel loss). Additionally, the generated X-ray images

Email:
v.quantran@vinbrain.net, v.thanhhuynh@vinbrain.net, v.huyta@vinbrain.net, v.chanhng@vinbrain.net, v.anhng@vinbrain.net, v.vuphan@vinbrain.net, v.namng@vinbrain.net, tuongtq.cntt@moh.gov.vn, drminhdien.nhtd@gmail.com, buivangiang@hmu.edu.vn, bhtrung@yahoo.com, brain01@vinbrain.net

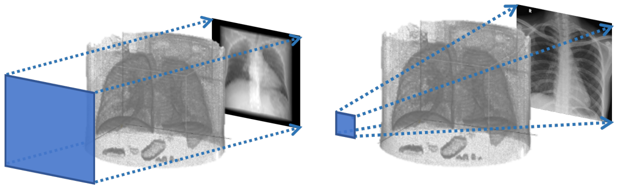


Fig. 1: Orthogonal projection versus perspective projection.

change over time during the training process due to a joint setup and can be regarded as a novel augmentation. To the best of our knowledge, this work is the first attempt that leverages heterogeneous data from both CT and X-ray to train a classification model and improve its performance compared to the baselines of using X-ray data only for COVID-19 diagnosis study.

2. METHOD

2.1. X-ray like image generation

The most straightforward way to convert a typical 3D CT volume to a 2D image is to average the intensity (ray accumulation) along particular directions (view direction) to generate such an X-ray-like image. This approach gives us a sense of how overall the 3D CT data looks like in the context of projection, but its fine-detail of image features appear to be blurred (see Fig. 1a) due to the nature of parallel axes. On the other hand, 2D Digitally Reconstructed Radiographs (DRRs) generated from 3D CT images by more advanced methods such as volume rendering algorithms [7, 8], have better representations of photo-realistic projections. Each pixel in DRRs is composited by the intensities sampled along the ray toward the screen (see Fig. 1b). Learning a specific view direction directly, view position and radiation dose (equivalent to lighting condition) can be done with deep learning [9]. However, in this work, we are interested in how commonly the realistic X-ray images look like, not manipulating the geometric and appearance of projected data. Therefore, we develop a deep neural network that learns to generate a photo-realistic X-ray projection from 3D CT volume data to leverage its label for enriching a handful of annotated X-ray images. Our objective is to produce realistic X-ray images with the difference to direct DRRs are minimal and use the generated X-ray images, which have labels originally from CT data, to improve the COVID-19 classification on a handful X-ray dataset with annotations.

2.2. System Overview

This part explains our model components to the proposed method (XPGAN). As shown in Fig. 2, XPGAN includes three main components: generator **G**, discriminator **D**, and classifier **C**. Generator **G** generates fake images I_f (projection) from a CT volume. Discriminator **D**, on the other hand, attempts to differentiate the real images that have been drawn from the database, i.e., distribution P_x and the fake ones produced by generator **G**. Equivalently, discriminator **D** produces the probability distribution over the CXR images. Last but not least, classifier **C** does its regular job to contrast the types of disease in images, both from manually annotated on CXR images and CT images.

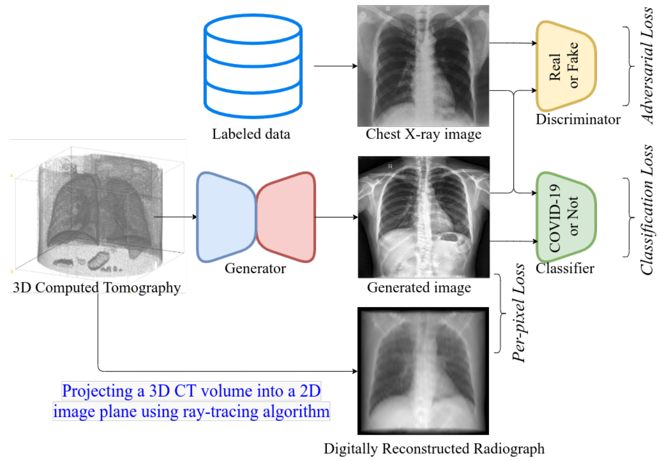


Fig. 2: The proposed method overview: Unlike other classification task, we have an additional projection network (Generator) to transform a CT volume to a CXR image which is close to its DRR.

Unlike the typical approach of COVID GAN [10] (based upon ACGAN [11]), which uses the concatenation of noise and prior annotation to generate an image with a specific class, our method uses labeled COVID-19 diagnosed CT volumes to get less artificial and high-resolution chest X-ray images. We adopt GAN loss for high-resolution image projection to match the output of the projected image to have the same distribution as normal chest X-ray image distribution. We add a per-pixel matching loss between the generated image and the ray tracing rendered image to constraint the projected images to have the same anatomical structure as of its original CT volumes. Finally, we train a classifier to assign the corresponding label to the image projected from CT volume by the Generator.

Training XPGAN is equivalently attempting to find the solution (the weights θ_G,θ_C and θ_D) of this minimax problem:

$$\min_{\theta_C,\theta_G} \max_{\theta_D} L(C) + V(G,D) + \lambda R(G) \quad (1)$$

where the classification loss $L(C)$, the adversarial loss $V(G,D)$ and DRR reconstructed loss $R(G)$ are defined as follows:

$$L(C) = \mathop{\mathbb{E}}_{x_2 \sim P_{XR}} \left[\sum_c -p(c|x_2) \log C(c|x_2) \right] + \mathop{\mathbb{E}}_{x_3 \sim P_{CT}} \left[\sum_c -p(c|x_3) \log C(c|G(x_3)) \right] \quad (2)$$

$$V(G, D) = \mathop{\mathbb{E}}_{x_2 \sim P_{XR}} [\log D(x_2)] + \mathop{\mathbb{E}}_{x_3 \sim P_{CT}} [1 - \log D(G(x_3))] \quad (3)$$

$$R(G) = \mathop{\mathbb{E}}_{x_3 \sim P_{CT}} [\|G(x_3) - DRR(x_3)\|_1] \quad (4)$$

where P_{XR} and P_{CT} are the data distribution of chest X-ray and CT data, respectively. We use a 3D CNN encoder to extract feature vector from a 3D CT volume. We then use the output feature vector as input for the Generator. For the 3D encoder, we adopt the skip connection architecture of the network in [12]. We further follows the design pattern in [13] by replacing all convolution-batch normalization block with equalized learning rate convolution layer and ReLU activation [14] with leaky ReLU [15] and $\alpha = 0.2$.

2.3. Implementation details

To leverage the power of current state-of-art models, we utilize either pre-trained networks or architectures of the current latest methods. Specifically, for Generator, which has inclusively two parts of Encoder and Decoder, we adopt the encoding network from [12] to get the high-dimensional representation of the input. Furthermore, the decoding network from StyleGAN2 [16] to transform such a latent vector from the previous stage to an X-ray image. Our Generator accepts the 256×256×64 3D CT volumes and produces a result of X-ray at 256×256 resolution. The architecture of the Discriminator is also copied from StyleGAN2. For the classification part, we utilize the pretrained DenseNet121 [17] on large-scale natural images (ImageNet) and transfer to our COVID-19 task. All the networks are jointly trained with learning rates of 0.002 and Cosine Annealing schedulers. Input to the Generator G is normalized to $[-1, 1]$, input to the Discriminator D, and classifier C are normalized to $[0, 1]$ to stabilize the training. We turn off the augmentation to observe the clear effects of having additional projected images for all experiments. The entire system is trained for 300 epochs with 4 NVIDIA V100 GPUs for 2 weeks. We empirically set $\lambda = 2$ to favor the reconstruction of CXR images.

3. DATA

3.1. X-ray images

We use a retrospective method to collect the data from the National Hospital of Tropical Diseases (NHTD) in Vietnam.

Table 1: Number of CXR images and CT volumes that have been used in the experiment.

Modalities		Source	# Neg.	# Pos.	Sum
CXR	Train	Vietnam - VM	9386	0	9386
		Vietnam - NLH	7600	0	7600
		Vietnam - NHTD	381	369	750
		Public - [18]	50	368	418
		Public - [19]	1	37	38
		Total	17418	**774**	**18192**
	Test	Vietnam - VM	2130	0	2130
		Vietnam - NLH	2400	0	2400
		Vietnam - NHTD	74	79	153
		Public - [18]	5	68	73
		Public - [19]	0	4	4
		Total	4609	**151**	**4760**
CT	Train	Public - [20]	254	**856**	**1110**

The collected data include personal information (genders and ages); pandemic declaration; clinical information (symptoms and temperatures); Reverse-Transcription Polymerase Chain Reaction (RT-PCR) tests; and the CXR images from positive COVID-19 cases during the treatment process. The CXR images are all frontal. For mild cases, while patients can stand, we perform AP scans. Otherwise, for severe cases, images were taken at the patients' beds are performed as PA scans. RT-PCR tests are used to confirm the status of COVID-19 and are treated as golden ground truth. In total, we have **903** images from NHTD, which have relatively equal distributions within the positive and negative samples of infection. For most non-COVID-19 CXR images, we also crowd-source the data from VinMec Hospital (VM), National Lung Hospital (NLH) in Vietnam and retrieve **21,516** images. We also grasp the CXR images with comprehensive annotations and medical notes from two sources: one from Github Repository of Cohen *et al.* [18] and one from [19] which have **491** and **42** positive COVID-19 images, respectively. Labels are extracted from confirmed metadata and notes. Table 1 summarizes the number of CXR images and their distributions. We separate this entire set of **22,952** images into the training and testing sets at ratio 80:20 using stratified sampling while *keeping images of the same patients in either train or test set* to ensure the similar distributions between two partitions and complete splits of train and test set.

3.2. Computed Tomography images

We further collect the 3D CT data from public domain [20] (**1110** volumes) which has **254** negative volumes and **856** positive COVID-19 volumes spread from mild to severe grading. These volumes are then pre-processed to produce corresponding projected DRR images and use their golden ground truth as the corresponding image-level COVID-19 labels. All pairs of volumes and DRR images are resized to 256×256×64 and 256×256, respectively.

Table 2: Comparison of XPGAN over other works

Model	Precision	Recall	F1
DenseNet121	0.839	0.762	0.799
DenseNet121 + DRR	0.849	0.742	0.792
COVID GAN [10] (Densenet121)	0.821	0.780	0.800
DeepCOVID [21] (ResNet50)	0.649	0.726	0.685
CVD-Net [22]	**0.897**	0.695	0.784
CovidAID [23]	0.692	**0.920**	0.790
CoroNet [24]	0.811	0.770	0.790
XPGAN (Proposed method)	0.831	0.815	**0.823**

4. RESULTS

4.1. Evaluation Metrics

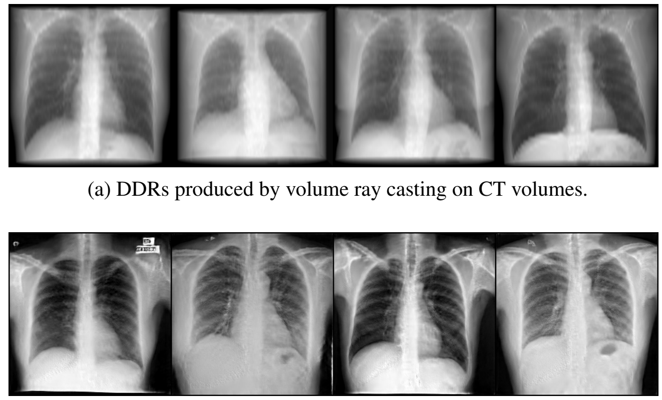
We use the standard evaluation for statistical classification in machine learning, such as its confusion matrix's derivations: Precision, Recall, F1 Score, instead of the Accuracy to measure the effectiveness of the comparing method. Since the data distribution is highly imbalanced, the F1 score becomes an essential metric while harmonizing the high Precision (or Positive Predictive Value) and the Sensitivity (or Recall) as we do not want to miss the positive cases but still want to classify them accurately.

4.2. Baselines of classification with DenseNet121 and X-ray-only data and with additional DRRs data

A straight forward baseline is to train a classifier (backboned with DenseNet121) on the X-ray-only data. Table 2 shows that this baseline can achieve the F1 score at 0.7986 with Precision and Recall are 0.8394 and 0.7616, respectively. In case DRRs are added into the training set, although we have more training samples, it reaches similar performance (Precision 0.8485, Recall 0.7417, F1 score 0.7915) compared to the previous naive approach. It clearly shows that leveraging direct ray casting projection data without further processing does not improve the classification metrics. This can be visually explained in Fig. 3a that even though DDRs have better representation compared to averaging intensity projection in terms of the view frustum, bone sharpness, and pseudo-lightning condition, they can not reflect fully the non-linearity of how X-rays penetrate through the object and burn the captured film. On the other hand, Fig. 3b illustrates our X-ray images generated by XPGAN from the same CT volumes, which look more realistic and more like standard X-ray images compared to DDRs.

4.3. Comparison with other related work

We roughly make a comparison with other concurrent work which are: COVID GAN [10], Deep-COVID [21], CVD-Net [22], CovidAID [23], and CoroNet [24]. These models have been fine-tuning on our training set for 300 epochs and evaluated on the test set. Deep-COVID [21] shipped with two backbones: ResNet18 and ResNet50 [25], both also pre-trained from ImageNet and finetuned on our training set. We empirically observe that the result from Deep-COVID



(a) DDRs produced by volume ray casting on CT volumes.

(b) X-ray-like images produced by XPGAN on the same CT volumes.

Fig. 3: Visual comparison between DDRs using Siddon algorithm [7] (a) and the generated samples by XPGAN (b).

(ResNet50) outperforms ResNet18. Intuitively, yet because ResNet50 is deeper than ResNet18 and hence results in better classification performance. CoroNet [24] also presents an exciting approach that uses pre-trained Xception as its backbone on ImageNet. CVD-Net [22] shows yet another good approach that designs a new network architecture that extracts multi-scale information, which makes the network better at capturing large structures. This matter explains its high Precision (0.8974). As also shown in Table 2, CovidAID [23] achieved the best Recall (0.92) compared to other and ours (0.78). It can be explained by the reason that CovidAID [23] makes use of large number of natural images and X-ray images in its pre-trained weight while we leverage the checkpoint of DenseNet121 on ImageNet only. However, in terms of Precision, CVD-Net [22] outperforms the baselines and the other concurrent works but suffers from great reduction in Recall. On the other hand, our method XPGAN has relatively high in both Precision and Recall. It is understandable because XPGANs avoid false-positive predictions by using the generated positive samples from confirmed statuses of CT volumes. Perhaps, COVID GAN [10] is the most closely related method compared to ours. Their result also achieves not too much difference between Precision (0.8210) and Recall (0.7800). Consequently, in terms of the F1 score, a harmonic mean of the Precision and Recall, our XPGAN model obtains the highest value (0.8227) compared to the others.

5. CONCLUSION

We present XPGAN in training a deep neural network to classify CXR images, targets to COVID-19 detection, with limited labeled data. The results show that the F1 score from the XPGAN improves up to $\sim$2% over the baselines, which has the same architecture of classifier (DenseNet121). Thanks to the nature of the generative model, we can synthesize the CXR images from confirmed cases of CT data. In future work, we plan to have an in-depth study of jointly training both 2D and 3D model to improve XPGAN.

6. COMPLIANCE WITH ETHICAL STANDARDS

This is a retrospective study of the COVID-19 pandemic using data in Vietnam and the public domain. The study was performed in line with the principles of the Declaration of Helsinki. Approval was granted by the Ministry of Health, Vietnam (Date: April 15 2020, No. 1724/QD-BYT).

7. CONFLICTS OF INTEREST

This work was supported by VinBrain, a company funded by VinGroup, Vietnam. Authors Tran Quy Tuong, Vu Minh Dien, Bui Van Giang and Bui Huu Trung have served on advisory boards for VinBrain. All authors have no conflicts of interest to disclose.

8. REFERENCES

[1] Yicheng Fang et al., "Sensitivity of chest CT for COVID-19: Comparison to RT-PCR," *Radiology*, p. 200432, Feb. 2020.

[2] Ho Yuen Frank Wong et al., "Frequency and distribution of chest radiographic findings in COVID-19 positive patients," *Radiology*, p. 201160, Mar. 2019.

[3] Jian-Long He et al., "Diagnostic performance between CT and initial real-time RT-PCR for clinically suspected 2019 coronavirus disease (COVID-19) patients outside wuhan, China," *Respiratory Medicine*, vol. 168, pp. 105980, July 2020.

[4] Zheng Ye, Yun Zhang, Yi Wang, Zixiang Huang, and Bin Song, "Chest CT manifestations of new coronavirus disease 2019 (COVID-19): a pictorial review," *European Radiology*, pp. 4381—4389, Mar. 2020.

[5] J.C.L. Rodrigues et al., "An update on COVID-19 for the radiologist - a british society of thoracic imaging statement," *Clinical Radiology*, vol. 75, no. 5, pp. 323–325, May 2020.

[6] Pascal Lomoro, Francesco Verde, Filippo Zerboni, Igino Simonetti, Claudia Borghi, Camilla Fachinetti, Anna Natalizi, and Alberto Martegani, "COVID-19 pneumonia manifestations at the admission on chest ultrasound, radiographs, and CT: single-center study and comprehensive radiologic literature review," *European Journal of Radiology Open*, vol. 7, pp. 100231, 2020.

[7] Robert L. Siddon, "Fast calculation of the exact radiological path for a three-dimensional CT array," *Medical Physics*, vol. 12, no. 2, pp. 252–255, Mar. 1985.

[8] Filip Jacobs, Erik Sundermann, Bjorn De Sutter, Mark Christiaens, and Ignace Lemahieu, "A fast algorithm to calculate the exact radiological path through a pixel or voxel space," *CIT. Journal of computing and information technology*, vol. 6, no. 1, pp. 89–94, 1998.

[9] Konstantinos Rematas and Vittorio Ferrari, "Neural voxel renderer: Learning an accurate and controllable rendering tool," in *IEEE/CVF CVPR*, June 2020, pp. 5416–5426.

[10] Abdul Waheed, Muskan Goyal, Deepak Gupta, Ashish Khanna, Fadi Al-Turjman, and Placido Rogerio Pinheiro, "CovidGAN: Data augmentation using auxiliary classifier GAN for improved COVID-19 detection," *IEEE Access*, vol. 8, pp. 91916–91923, 2020.

[11] Augustus Odena, Christopher Olah, and Jonathon Shlens, "Conditional image synthesis with auxiliary classifier GANs," in *ICML*, 2017, pp. 2642–2651.

[12] Liyue Shen, Wei Zhao, and Lei Xing, "Patient-specific reconstruction of volumetric computed tomography images from a single projection view via deep learning," *Nature Biomedical Engineering*, vol. 3, no. 11, pp. 880–888, Oct. 2019.

[13] Tero Karras, Timo Aila, Samuli Laine, and Jaakko Lehtinen, "Progressive growing of GANs for improved quality, stability, and variation," in *ICLR*, 2018, pp. online 1–26.

[14] Vinod Nair and Geoffrey E. Hinton, "Rectified linear units improve restricted boltzmann machines.," in *ICML*, 2010, pp. 807–814.

[15] Bing Xu, Naiyan Wang, Tianqi Chen, and Mu Li, "Empirical evaluation of rectified activations in convolutional network," *CoRR*, vol. abs/1505.00853, 2015.

[16] Tero Karras, Samuli Laine, Miika Aittala, Janne Hellsten, Jaakko Lehtinen, and Timo Aila, "Analyzing and improving the image quality of StyleGAN," in *IEEE CVPR*, 2020, pp. 4401–4410.

[17] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q. Weinberger, "Densely connected convolutional networks," in *IEEE CVPR*, July 2017, pp. 4700–4708.

[18] Joseph Paul Cohen, Paul Morrison, and Lan Dao, "COVID-19 image data collection," *arXiv 2003.11597*, 2020.

[19] Kor. Soc. of Thoracic Radiology, "Weekly COVID-19 cases," https://kstr.radiology.or.kr/weekly/corona/.

[20] Sergey Morozov, Anna Andreychenko, Nikolay Pavlov, Anton Vladzymyrskyy, Natalya Ledikhova, Victor Gombolevskiy, Ivan Blokhin, Pavel Gelezhe, Anna Gonchar, Valeria Chernina, and Vladimir Babkin, "MosMedData: Chest CT scans with COVID-19 related findings," *arXiv 2005.06465*, May 2020.

[21] Shervin Minaee, Rahele Kafieh, Milan Sonka, Shakib Yazdani, and Ghazaleh Jamalipour Soufi, "Deep-COVID: Predicting COVID-19 from chest x-ray images using deep transfer learning," *Medical Image Analysis*, vol. 65, pp. 101794, Oct. 2020.

[22] Chaimae Ouchicha, O. Ammor, and M. Meknassi, "CVDNet: A novel deep learning architecture for detection of coronavirus (COVID-19) from chest x-ray images," *Chaos, Solitons, and Fractals*, vol. 140, pp. 110245 – 110245, 2020.

[23] Arpan Mangal, Surya Kalia, Harish Rajgopal, Krithika Rangarajan, Vinay Namboodiri, Subhashis Banerjee, and Chetan Arora, "CovidAID: COVID-19 detection using chestx-ray," *arXiv 2004.09803*, 2020.

[24] Asif Iqbal Khan, Junaid Latief Shah, and Mohammad Mudasir Bhat, "CoroNet: A deep neural network for detection and diagnosis of COVID-19 from chest x-ray images," *Computer Methods and Programs in Biomedicine*, vol. 196, pp. 105581, Nov. 2020.

[25] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, "Deep residual learning for image recognition," in *IEEE CVPR*, 2016, pp. 770–778.