2023 IEEE Symposium Series on Computational Intelligence (SSCI)
Mexico City, Mexico. December 5-8, 2023

Synthetic Generation of Pneumonia Images using CycleGAN Model

Gerardo Lugo-Torres
Center for Computing Research (CIC)
Instituto Politécnico Nacional
Mexico City, Mexico
glugot2022@cic.ipn.mx

Diego A. Peralta-Rodríguez
Computational Cognitive Sciences Lab. - CIC
Instituto Politécnico Nacional
Mexico City, Mexico
dperaltar1100@alumno.ipn.mx

José E. Valdez-Rodríguez
Computational Cognitive Sciences Lab. - CIC
Instituto Politécnico Nacional
Mexico City, Mexico
jvaldezr2018@cic.ipn.mx

Hiram Calvo
Computational Cognitive Sciences Lab. - CIC
Instituto Politécnico Nacional
Mexico City, Mexico
hcalvo@cic.ipn.mx

***Abstract*—The utilization of generative models in image synthesis has become increasingly prevalent. Synthetic medical imaging data is of paramount importance, primarily because authentic medical imaging data is scarce, costly, and encumbered by legal considerations pertaining to patient confidentiality. Consent forms are typically required from patients in order to utilize their data for publication in medical journals or educational purposes. Consequently, the accessibility of medical data for general public research is limited. Synthetic medical images offer a potential resolution to these issues. The predominant approaches primarily assess the quality of images and the degree of resemblance between these images and the original ones employed for their generation. In this study, we employ a CycleGAN model to produce artificial images depicting several types of pneumonia, including general, bacterial, and viral pneumonia. We then evaluate the performance of these synthetic images by comparing them with ratings made by three respiratory care professionals. Consequently, a range of pneumonia pictures were acquired, exhibiting diverse levels of performance, ranging from being easily identified as false to being correctly identified as real in over 80% of cases.**

***Index Terms*—Synthetic, Chest x-ray, Cyclic, Generative Adversarial Network, Pneumonia,Image-to-image,Translation**

Recent research has emphasized generative models for image synthesis [1]. AI's growing role in medical research, especially in processing images, text, and sound, demands significant data. Collecting this data involves time, resources, and obtaining enough annotated data for effective CNN training.

This study explores using a CycleGAN [2] to produce chest X-ray images of various pneumonia types with quality comparable to "real" data for AI medical models. Although previous research has utilized generative models for medical image creation [1], few [3] delve into interpreting metrics determining when generated images can be deemed natural for AI medical use.

Our aim is to examine the correlation between the Frechet Inception Score (FID) and Inception Score (IS) [4] with respiratory care expert evaluations, gauging the "realness" of the generated images. This work contributes by: 1) Comparing FID and IS metrics with pneumonia images from a CycleGAN model over 100 training epochs; and 2) presenting evidence that these metrics don't align with expert evaluations on image authenticity.

The CycleGAN architecture is different from other GANs because it contains 2 mapping function (G and F) that acts as generators and their Discriminators (D_x and D_y): The generator mapping functions are: $G : X \rightarrow Y$ $F : Y \rightarrow X$ where X is the input image distribution and Y is the desired output distribution. The cost function is the sum of adversarial loss and cyclic consistent loss:

$$L(G, F, D_x, D_y) = L_{\text{advers}}(G, D_y, X, Y) + L_{\text{advers}}(F, D_x, Y, X)$$
$$+\lambda L_{\text{cycl}}(G, F, X, Y)$$

with an objective function with the form of:

$$\min_{G,F} \max_{D_x,D_y} L(G, F, D_x, D_y)$$

We trained a CycleGAN, whose architecture is described in -B, for 100 epochs with 3 datasets: 1341 normal X-ray images, 2531 bacterial pneumonia images and 1345 viral pneumonia images. A total of 1,341 synthetic images of general pneumonia (GP), bacterial pneumonia (BP), and viral pneumonia (VP) were generated at a different number of training:25, 50, 75, and 100 epochs. For each set of images, its corresponding FID and IS values were calculated.

Three questionnaires, each containing 100 images, were sent to three respiratory care experts. The images were divided into four sections for evaluation. Out of the 100 images, 80 were generated, and 20 were real. In the questionnaires, they were given the option to choose if the image they saw was real or fake and if the image corresponded to general pneumonia, bacterial pneumonia, viral pneumonia, or did not correspond to pneumonia.

978-0-7381-4408-5/23/$31.00 ©2023 IEEE

Table I
QUANTITATIVE RESULTS.

Dataset	Epochs	FID	IS
General Pneumonia	25	64.629	1.8011
	50	93.1049	1.8217
	75	106.5786	1.8568
	100	86.5787	2.3461
Bacterial Pneumonia	25	68.9245	2.1217
	50	64.5383	1.9673
	75	78.4502	1.824
	100	67.4041	1.296
Viral Pneumonia	25	54.4719	2.2069
	50	63.9624	2.2485
	75	77.9493	2.489
	100	60.9601	2.5063

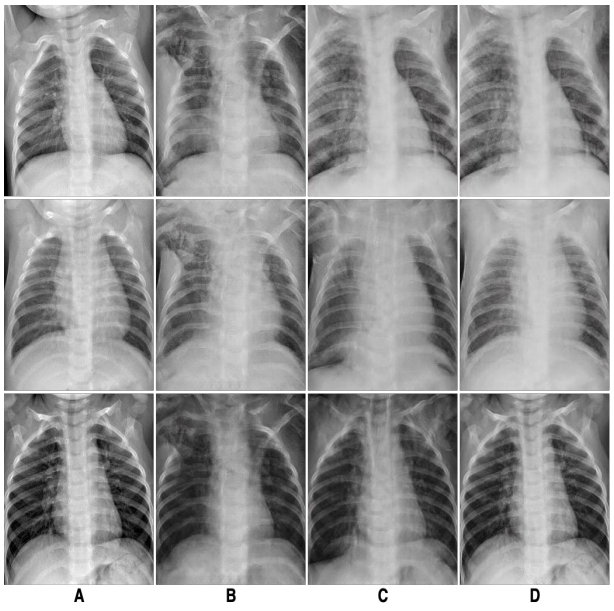

Figure 1. Results obtained from our CyclicGAN model, A)Original Image, B)General pneumonia, C)Bacterial pneumonia, D)Viral pneumnonia

For the GP generated images the best expert assessment average of 0.58 was with 25 epochs of training and match also with the best FID but not IS score of its group. The best BP generated images was with 75 epochs of training and got an average expert assessment of 0.9 which which does not match the best FID or IS score of its group, overall this group produce the best score assessment images by the experts of all. And for the VP generated images the best expert assessment average of 0.53 was with 100 epochs of training which does not match the best FID or IS of its group.

Table II
PERCENTAGE OF SAMPLES IDENTIFICATION.

Dataset	Epochs	Expert 1		Expert 2		Expert 3	
		SR*	RR*	SR	RR	SR	RR
GP	25	0.55	0.8	0.6	0.8	0.6	0.6
	50	0.35	0.8	0	0.8	0.35	0.8
	75	0.45	1	0.4	0.8	0.35	0.8
	100	0.45	0.6	0.3	0.6	0.45	0.8
BP	25	0.7	0.6	0.6	1	0.7	0.6
	50	0.7	0.8	0.45	0.6	0.65	0.4
	75	0.8	0.2	0.95	1	0.95	0.8
	100	0	0.6	0.2	2	0.15	0.8
VP	25	0.55	1	0.6	1	0.35	1
	50	0.35	1	0.8	0.6	0.35	0.6
	75	0.68	0.8	0.55	1	0.25	1
	100	0.55	0.6	0.7	1	0.35	1

*SR:synthetic image and class identified as real , RR: real image and class identified as real.

The results obtained give us evidence that the images generated score either FID or IS do not translate into their ability to pass as images of real pneumonia.

A. Conclusions

In this work, we performed a comparison between scores and the subjective assessment of experts. We didn´t see a relationship between the best FID, IS, and the best assess expert score images. This will suggest the need for a proposal of a better score that will allow the use of synthetic images more reliably and helpfully. A DCNN model has been proposed for evaluating the quality of the images generated, an Augmenteed Cycle GAN to improve image quality, and sought radiology experts to include more and better assessments.

B. Dataset & Code

The code used in this paper and the dataset is available at: https://github.com/Lugo1025/PneumoCGAN.

ACKNOWLEDGMENTS

We gratefully acknowledge the help with the proffesional assesment of the images to Dr. Yazmin Guillen Dolores from the National Institute of Cardiology, Dr. Gustavo Lugo Goytia and Dr. Sergio Gustavo Monasterios López from the National Institute of Respiratory Diseases and the National Council of Humanities, Science and Technology (CONAHCYT). Thanks for its support to the Instituto Politécnico Nacional (COFAA, SIP-IPN, Grant SIP 20230140).

REFERENCES

[1] D. I. Morís, J. de Moura, J. Novo, and M. Ortega, "Cycle generative adversarial network approaches to produce novel portable chest x-rays images for covid-19 diagnosis," in *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2021, pp. 1060–1064.

[2] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2223–2232.

[3] B. Segal, D. M. Rubin, G. Rubin, and A. Pantanowitz, "Evaluating the clinical realism of synthetic chest x-rays generated using progressively growing gans," *SN Computer Science*, vol. 2, no. 4, p. 321, 2021.

[4] A. Borji, "Pros and cons of gan evaluation measures: New developments," *Computer Vision and Image Understanding*, vol. 215, p. 103329, 2022.