Class-Specific GANs to Improve Synthesis of Bacterial and Viral Pneumonia Chest X-Ray Images

Daksh Kalra[1]*, Vijay Khare[1], Ramit Kumar[2]

[1] Department of Electronics and Communication Engineering, Jaypee Institute of Information and Technology, Noida 201309, India
[2] Department of Information Technology, Netaji Subhas University of Technology University, Delhi 110078, India

Corresponding Author Email: vijay.khare@mail.jiit.ac.in

ABSTRACT

In the last couple of years, deep learning models have been extensively used in the field of pneumonia detection using image processing. Making a real world chest X-Ray dataset takes time and comes with a finite number of images. But in order to train an AI model from scratch, the need for readily-available, limitless and accurate data is of utmost importance. So, it is more prudent to use generative adversarial network (GAN). This technology converts a noise matrix to the desired image matrix. Undoubtedly the technique has been used by researchers in this field but our work tries to make this generated image more accurate using a specific GAN for each class of data. For each of the three classes of chest X-Ray images, namely bacterial, viral and normal, a GAN has been trained to provide images of utmost accuracy. We believe that specific GAN for each subgroup is capable of producing more accurate images in desired quantities, as the GAN itself is trained on focussed images and thereby not only helping to limit the problem of model overfitting but also perfecting the discriminators of each GAN. In our case we were able to reach similar levels of clarity in the high resolution images as compared to single chest X-Ray GAN in a lesser amount of time and using a lesser amount of training data.

1. INTRODUCTION

The recent pandemic caused by Coronavirus also known as COVID-19 emphasized better diagnostic techniques than the ones that had been employed. This led to extensive work in the field of image processing. The X-Rays are capable of passing through the air and tissues but not through the Infiltrates (the accumulated infection in lungs which is visible as white spots in chest X-Ray images) which makes it an excellent choice for diagnostic tests. CT scans also play a vital role in the area but X-Rays tend to be more widespread owing to their affordability and simplicity. Using X-Ray imaging we can determine any abnormalities in the lungs, as they are images to which the techniques of machine learning can be applied. This provides a solution to diagnose diseases with a non-invasive technique and without using a lot of time. Machine learning can help doctors and radiologists in making right decisions and reducing their workload by filtering out the data beforehand. This is one of the major reasons why hundreds of artificial neural networks and Deep Learning models based on CNN and GAN have been designed in the last couple of years.

But it usually takes about 10 to 15 minutes to get an X-Ray report and the person is exposed to about 0.1 mSv of radiation in a single chest X-Ray. Even after X-Ray films are ready, they still need to be digitized. These factors negatively affect the collection of data and machine learning requires a lot of data to train and perfect itself. This would mean exposing the human body to more radiation and gigantic time consumption, which isn't good. Moreover, the data has to be accumulated, processed and stored which is time-consuming and arduous. Moreover, artificially generated data using a single GAN model tends to lack clarity. Therefore, we need a solution that can solve both of these problems and still be viable and pragmatic enough for use in day-to-day life and on a large scale.

We came up with the solution of class-specific GANs. A GAN to synthesize the images related to a class instead of a single GAN for all the images in the dataset. This would potentially eliminate the above mentioned problems without drawbacks in the long run. The only couple of loopholes that this idea might suffer from is that it takes more time initially to develop and is more complex to implement than the traditional GAN and CNN model combination.

2. RELATED WORK

The paper [1] discussed the course and severity of infection in the lungs from the day of diagnosis till the day the patient recovers fully. Researchers have also emphasized the role of using X-Rays instead of CT-scan since they are portable, available and can be used extensively. In this paper, each lung was given a rating of 0-four depending on the extent of lung involvement (rating 0 = no involvement, $1 \leq 25\%$, 2 = 25%–

50%, 3 = 50%–75%, 4 ≥ 75% lung affection). A complete severity rating was calculated by summing each lung score (general severity scores ranged from zero to eight).

The studies [2, 3] noted following deep learning-based approaches: deep feature extraction, fine-tuning of pre-trained convolutional neural networks (CNNs), and end-to-end training of the developed CNN models along with pre-trained deep CNN models (ResNet18, ResNet50, ResNet101, VGG16, and VGG19) were used for deep feature extraction. Deep feature classification was done by support vector machine (SVM) classifiers with different kernel functions (linear, quadratic, cubic, Gaussian). The pre-trained deep CNN model above was also used for the fine-tuning procedure. A new CNN model with end-to-end training is proposed. A dataset of 180 COVID-19 and 200 normal (healthy) chest radiographs were used for the experiments in this study. Classification accuracy was used as a study performance indicator. Experimental studies show that deep learning has the potential to detect COVID-19 based on chest X-Ray images. Deep features extracted from the ResNet50 model and SVM classifier using a linear kernel function had the highest accuracy rating of 94.7% among all results obtained. The performance of the fine-tuned ResNet50 model was 92.6%, but the end-to-end training of the developed CNN model was 91%.

The research [4] revolved around 1000 images evenly distributed between normal and chest X-Ray. Then the images were utilized in the training of the DenseNet-161 network, with a training-to-test ratio of 80:20 and were able to segregate pneumonia from COVID-infected chest X-Rays with 99% accuracy. The researcher also attempted to classify chest X-Rays into three subclasses based on the severity of the infection. light, medium and heavy. That resulted in only 80 labeled images for each subclass. ResNet18 was better at classifying between his three subclasses with 76% accuracy. This model can be improved further for more accuracy.

The work [5] suggested a new method to detect COVID-19 and pneumonia using chest X-Rays. The three-step process described in the paper performs the following task. In the first step, a conditional generative adversarial network (C-GAN) segments tube radiographs to obtain lung images. Then the segmented lung images are put into the new pipeline. This pipeline then combines a key point extraction method with a trained deep neural network (DNN) for discriminating feature extraction. Many machine learning (ML) models are used to classify COVID-19, pneumonia, and normal lung images in the final step. Proposed architectures which are a combination of DNNs, keypoint extraction methods, and ML models have been analyzed. The highest test classification accuracy achieved was of 96.6% using the VGG-19 model along with the Binary Robust Invariant Scalable Keypoint (BRISK) algorithm.

Further, the study [6] discussed the synthetic chest X-Ray images generated by CovidGAN with the CNN architecture can prove to be more fruitful. This research has the following contributions: Proposed an auxiliary classifier generative adversarial network (ACGAN)-based GAN called CovidGAN for synthetic chest X-Ray image generation. Augment the training dataset with a CNN model using CovidGAN to improve COVID-19 detection. The method used to generate synthetic chest X-Ray (CXR) images is the ACGAN based model called CovidGAN. Moreover, they were able to present that CNN alone yielded 85% accuracy but adding synthetic images produced by CovidGAN increased the accuracy to 95%.

The above cited papers have done a good job describing ways to segregate the lungs based on the percentage of lung affected and training ResNet50 architecture CNN model to classify chest X Ray images and CT scans. We set out to prove that even with a smaller model architecture as discussed in the next section can be used to attain similar, if not better results with the help of specific data. Moreover, in this paper, a single GAN is used to increase the amount of data by 30 times and the generator has to generate data for a number of lung infections and diseases caused by different pathogens which leads to more confusion and the GAN model has a hard time converging.

3. DEEP LEARNING ALGORITHMS USED

3.1 CNN

The convolutional neural networks are a subgroup or subclass of artificial neural networks which are mainly used for processing images and visual data as shown in Figure 1. This machine-learning algorithm can also be applied to audio signals. It is able to recognize patterns and differences based on similarities. There are three layers, namely the convolutional layer, pooling layer and fully connected layer placed in a sequence [7, 8].
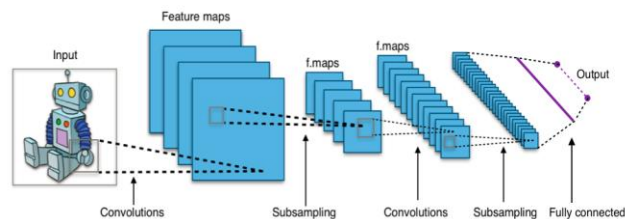

Figure 1. CNN architecture

```
Model: "model"

Layer (type)                Output Shape            Param #
==================================================================
input_1 (InputLayer)        [(None, 64, 64, 1)]     0

conv2d (Conv2D)             (None, 64, 64, 16)      160

conv2d_1 (Conv2D)           (None, 64, 64, 32)      4640

max_pooling2d (MaxPooling2D (None, 32, 32, 32)      0
)

conv2d_2 (Conv2D)           (None, 32, 32, 16)      2064

conv2d_3 (Conv2D)           (None, 32, 32, 32)      2080

max_pooling2d_1 (MaxPooling (None, 16, 16, 32)      0
2D)

flatten (Flatten)           (None, 8192)            0

dense (Dense)               (None, 100)             819300

dense_1 (Dense)             (None, 50)              5050

dense_2 (Dense)             (None, 3)               153
==================================================================
Total params: 833,447
Trainable params: 833,447
Non-trainable params: 0
```

Figure 2. Model parameters

But this is not so straightforward in practical life since there are a lot of different factors like amount of data, the number of parameters your model requires and different resolution images. All of these factors affect the amounts of layers the model will need. Generally, when there is a large amount of training data available it is prudent to limit the number of layers to prevent overfitting of the model, as shown in Figure 2. In our experiment we used 11 layers in total (4 convolutional layers and 3 dense layers) since we were able to

generate large amounts of data [9-11].

3.2 GAN

Developed back in 2014 by Ian Goodfellow, this novel type of machine learning concept generates images from noise matrix. To perform this trick, the GAN trains two individual agents namely generator and discriminators, as shown in Figure 3. After each iteration in the training loop, these two agents play a zero-sum game, i.e., the generator tries to fool the discriminator into believing that the images generated by it are real and the discriminator tries to differentiate between the real and synthetic generator's image by focusing on each and every detail [12-15].

The GANs break down if the generator starts focusing on the one type of images that the discriminator tends to classify as real image which will result in bad synthetic images. But since these class-specific GAN would have to focus only on one type of image, it would be easier for them to understand the area in which or the nature of infection a pathogen generally exhibits.
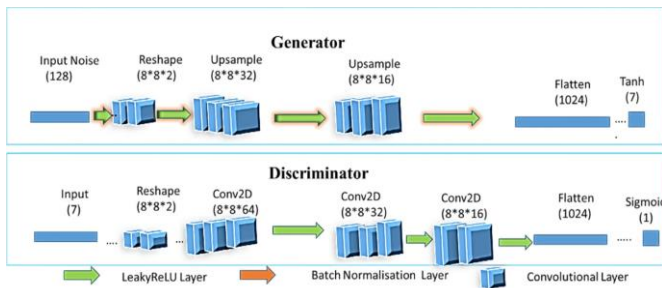

Figure 3. Generative adversarial network

4. PROPOSED NEW MODEL

When a single GAN is trained to generate the synthetic images of chest X-Ray pneumonia causes images to be blurry and inaccurate. As shown in the Figure 4, the synthetic image set generated is in a matrix of 16×16 which allows for decrease in the size of a single image and hence reduces the computation yet there are many images like the one pointed by red arrow, which are not up to the mark.
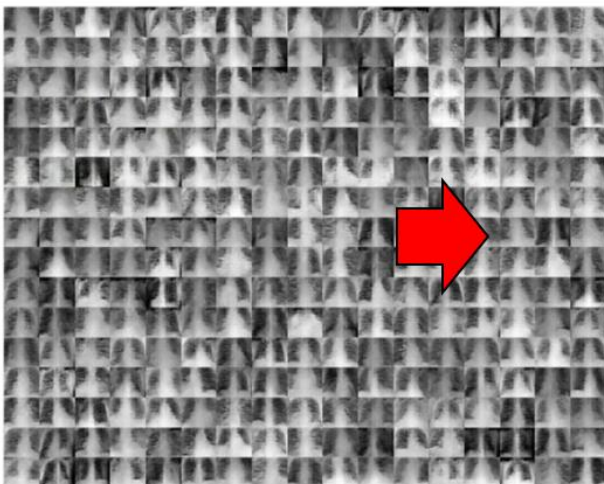

Figure 4. Depiction of error

So, in order to get at least the same level, if not more, of image clarity for the image set of 8×8 matrix, but a class-specific dedicated GAN for each of the three classes is as shown in Figure 5. These would generate image sets tailored to the specifications and completely eliminate the need for image preprocessing. To implement this technology, the unsorted data is first passed through a trained CNN network and then sent to different dedicated GANs for each of the particular classes which in our case are namely viral, bacterial and normal. This way these GANs can be trained to generate more accurate images and we would get more than 4 times the data than the researchers in the above-mentioned paper were able to get. Moreover, in this model, we will then train a CNN which will limit problems of overfitting, resulting in better accuracy.
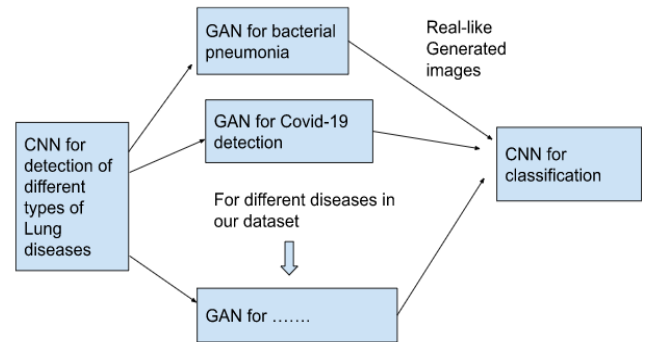

Figure 5. Model architecture

For the sake of simplicity and practicality, we limit ourselves to three GANs for bacterial pneumonia chest X-Ray images, viral pneumonia chest X-Ray images and normal person chest X-Ray images. The synthesized image from the GANs can be then sent to the CNN model for further training or be stored for future use [16-20].

The dataset we chose was very limited in images (containing only a thousand of each type), so as to check the performance of the model with constraints. Furthermore, some of the images lack quality and are a bit hazy which would provide a more challenging environment for the model [21-26].

5. RESULTS AND DISCUSSIONS

The CNN classifier is based on ResNet18 and is able to segregate the images into three subgroups with a training accuracy of 94%.

The testing accuracy as shown in Figure 6 clearly depicts the accuracy to be a maximum of 70%. This is not up to the mark for a field of study that directly curtains with people's health and lives. The precision, recall and F1 score also had poor performance (below 0.6). Figure 7 shows the confusion matrix that was obtained.

This is the loss function of the Normal chest X-Ray images from GAN after training. Plotting the generator and discriminator on a single graph provides valuable insight into the functioning of GAN and their intersection (Figure 8). This could be useful later in improving the performance further. This convergence clearly shows that the GAN model has found optimum from where it cannot improve more. To verify that it was not a mode collapse, we examined the generated images in Figure 9 alongside the real images.
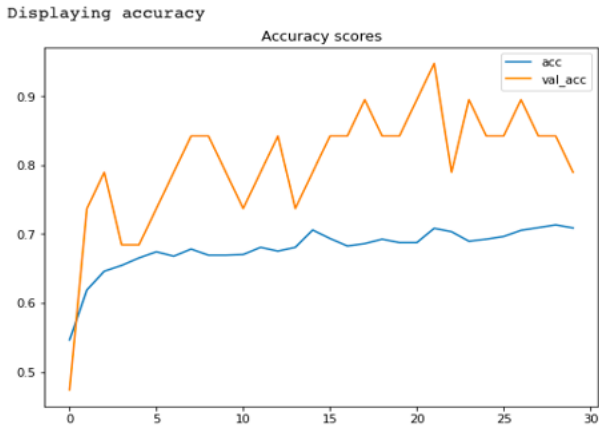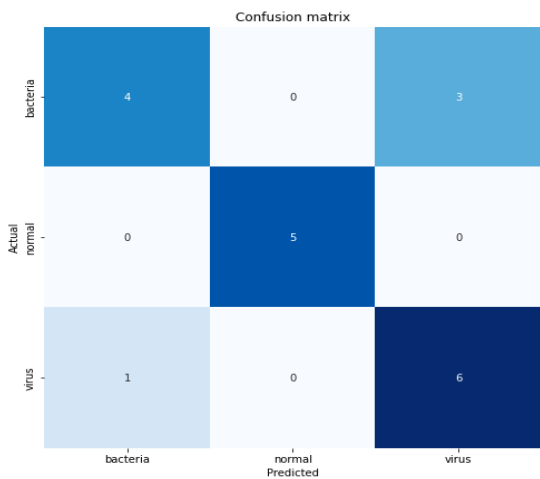
Figure 6. Accuracy plot
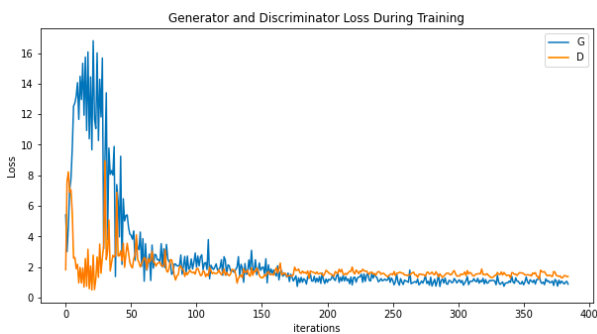

Figure 7. Confusion matrix

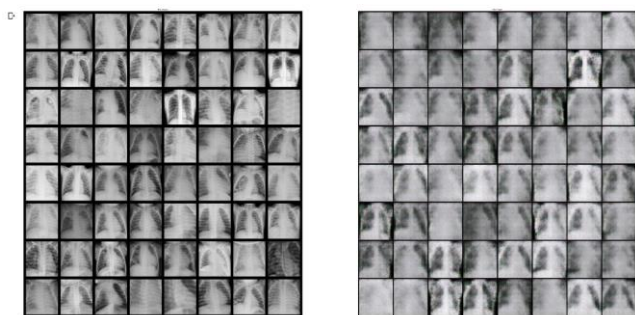**Figure 8.** Loss plot for normal chest X-Ray images

Figure 9. Side by side real and generated images respectively for normal chest X-Ray

Evidently, the loss between the discriminant and generator for normal CXR decreases with the increase in iterations. Hence one can believe that a minimum has been found at around the 200th iteration.

Figure 8 proves that in this case the model has not suffered mode collapse since the synthetic image set on the left has images that have great resemblance to real images in the image set on the right. Similarly, the trained model for bacterial chest X-Ray images has the loss graph as shown in Figure 10.

Although the loss graph just imitates the loss graph of the normal chest X-Ray GAN model yet images made by bacterial chest X-Ray image GAN are not as clear as the normal chest X-Ray synthetic images, as presented in Figure 11.
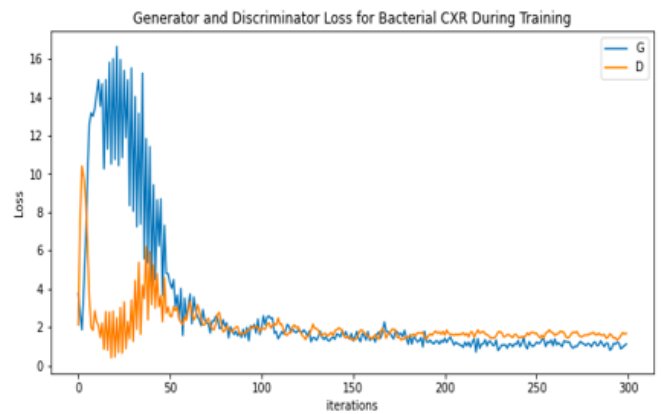

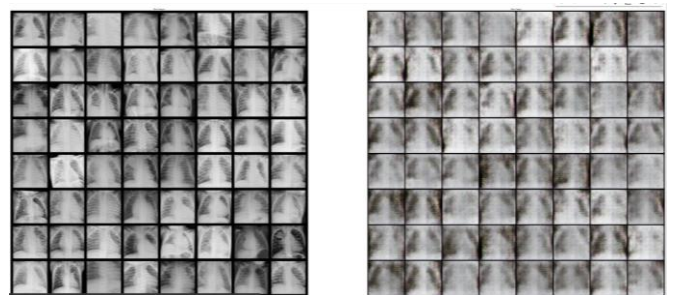**Figure 10.** Loss plot of GAN for bacterial chest X-Ray images

Figure 11. Side by side real and synthetic images respectively for bacterial chest X-Ray
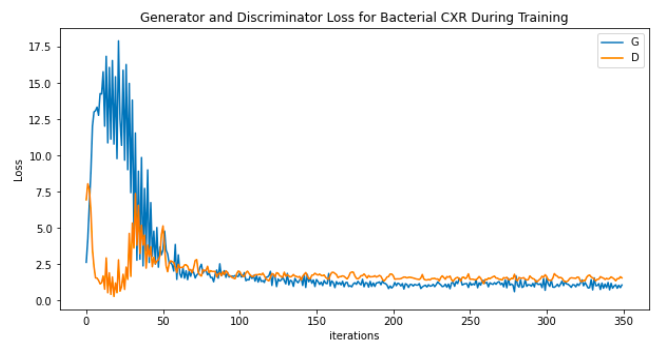

Figure 12. Loss plot for viral chest X-Ray images

As illustrated in Figure 12, the plot for viral CXR images appears similar to the one we used for training with normal CXR images.

Figure 13 shows a distinction between generated and real

images obtained from viral CXR image GAN.

Although score and accuracy showed only a slight improvement (Figure 14), we were able to shift the CNN bias from high 'False Virals' to 'True Bacterials'. This might not be a massive jump nor does this accuracy make the CNN model ready for real world application but the idea to try to balance out the dataset is the strength. It can be clearly inferred from the confusion matrix below that the use of class-specific GANs confusion matrix obtained after adding the fake images to test data.
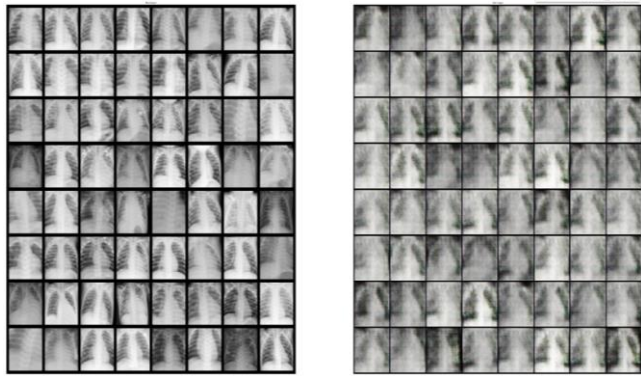

Figure 13. Side by side real and generated images respectively for viral chest X-Ray GAN
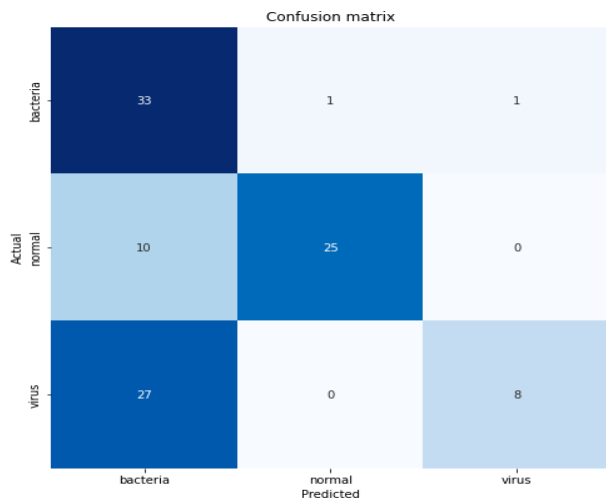

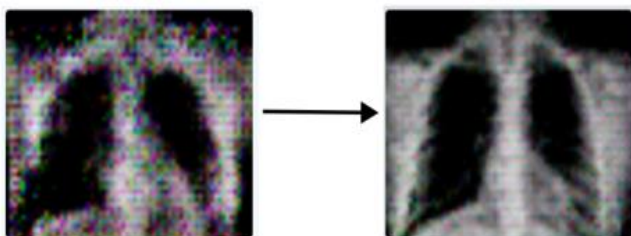**Figure 14.** Confusion matrix after addition of synthetic images

Figure 15. Improvement by changes in hyperparameters

As shown in Figure 15, most of the hyper parameters of the optimizer (Adam) were set to default values like beta1=0.9, beta2=0.999, epsilon=1e-8, decay=0.0. When the number of epochs and Learning Rate (LR) for GAN models were changed, the observations were as shown in Table 1.

Table 1. Hyperparameter changes

Learning Rate (LR)\Epochs	Low Number of Epochs (in 100s)	High Number of Epochs (in 1000s)
Low LR (0.00002)	Images were blurry and not with distinct features	Images bore too much resemblance to each other (Overfitting)
High LR (0.01)	Images were far better than the others	Images were just bunch of pixels without much data

To get a perspective of the improvement by hyperparameter changes look at the image below:

After that we decided to make a user program which would help doctors or technicians to train the GAN at their end with the live and updated dataset for future use. The user program shows a synthetic image along with a question "what do you think about this image?" and accepts Boolean values. This not only allows us to figure out whether the GANs are getting trained correctly but also helps us determine the ratio of True to False which we coin as the improvement ratio. This allows for continuous training of the GANs and keeps a balance for overfitting.

Following are some of the features present in the user program:

User can define a dataset folder of his/her choice to train the GANs.

The program stores the improvement ratio in a list and checks if the number of false's have increased. In such a scenario it pops up a notification and suggests the user to change the training dataset.

Otherwise, it starts the GANs training again if the improvement ratio is more than the 0.9.

The limitations of this idea are if the virus goes on evolving, the variations in the images might lead to GANs breaking down and as the CNN model is dependent on the synthetic image to train further, even one.

6. CONCLUSIONS AND FUTURE WORK

Therefore, one can conclude the below mentioned points from this experiment on the idea of class-specific GANs:

That there are appreciable improvements in accuracy and performance of CNN model.

That the dataset can be increased massively.

The CNN model can be made more robust.

This idea based model can be deployed in hospitals and diagnosis centers while still improving in its performance with the help of the above mentioned user interface. The technicians and doctors would get at least a hint or a heads up before directly reviewing the chest X Ray. This would greatly help them in making a prudent decision.

The GANs designed specifically for each type of chest X-Ray evidently helped in balancing data and were able to train and converge quickly, even with less amount of data. Although the results haven't been able to improve the model's figure of merit to acceptable levels but is able to provide clarity about its true performance as well as the CNN model's behavior when faced with challenges. For future work, we will train the CNN with even more precise synthetic images.

For the future, it is possible to generate even better images with further training of the model on different loss functions

and optimizer parameters. Moreover, the basic architecture can be changed with varying number of layers and other pre-trained models like GoogleNet, ImageNet etc. There is also a scope to include Progressive GANs and adaptive instance normalization to further improve the image quality.

ACKNOWLEDGMENT

REFERENCES

[1] Loey, M., Smarandache, F., Khalifa, N.E.M. (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. https://doi.org/10.3390/sym12040651

[2] Yasin, R., Gouda, W. (2020). Chest X-ray findings monitoring COVID-19 disease course and severity. Egyptian Journal of Radiology and Nuclear Medicine, 51(193): 1-18. https://doi.org/10.1186/s43055-020-00296-x

[3] Ismael, A.M., Şengür, A. (2021). Deep learning approaches for COVID-19 detection based on chest X-ray images. Expert Systems with Applications, 164: 114054. https://doi.org/10.1016/j.eswa.2020.114054

[4] Shelke, A., Inamdar, M., Shah, V., Tiwari, A., Hussain, A., Chafekar, T., Mehendale, N. (2021). Chest X-ray classification using deep learning for automated COVID-19 screening. SN Computer Science, 2(4): 300. https://doi.org/10.1007/s42979-021-00695-5

[5] Bhattacharyya, A., Bhaik, D., Kumar, S., Thakur, P., Sharma, R., Pachori, R.B. (2021). A deep learning-based approach for automatic detection of COVID-19 cases using chest X-ray images. Biomedical Signal Processing and Control, 68: 102588. https://doi.org/10.1016/j.bspc.2021.102588

[6] Waheed, A., Goyal, M., Gupta, D., Khanna, A., Al-Turjman, F., Pinheiro, P.R. (2020). Covidgan: Data augmentation using auxiliary classifier gan for improved covid-19 detection. IEEE Access, 8: 91916-91923. https://doi.org/10.1109/ACCESS.2020.2976556

[7] El-Sawy, A., El-Bakry, H., Loey, M. (2017). CNN for handwritten Arabic digits recognition based on LeNet-5. In Proceedings of the International Conference on Advanced Intelligent Systems and Informatics 2016 2, Springer International Publishing, pp. 566-575. https://doi.org/10.1007/978-3-319-48308-5_56

[8] El-Sawy, A., Loey, M., EL-Bakry, H. (2017). Arabic handwritten characters recognition using convolutional neural network. WSEAS Transactions on Computer Research, 5(1): 11-19.

[9] LeCun, Y., Huang, F.J., Bottou, L. (2004). Learning methods for generic object recognition with invariance to pose and lighting. In Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Washington, DC, USA, pp. II-104-II-109. https://doi.org/10.1109/CVPR.2004.1315150

[10] Stallkamp, J., Schlipsing, M., Salmen, J., Igel, C. (2011). The German traffic sign recognition benchmark: A multi-class classification competition. In Proceedings of the 2011 International Joint Conference on Neural Networks, San Jose, CA, USA, pp. 1453-1460. https://doi.org/10.1109/IJCNN.2011.6033395

[11] Deng, J., Dong, W., Socher, R., Li, L., Kai, L., Li, F.F. (2009). ImageNet: A large-scale hierarchical image database. In Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, pp. 248-255. https://doi.org/10.1109/CVPR.2009.5206848

[12] Liu, S., Deng, W. (2015). Very deep convolutional neural network-based image classification using small training sample size. In Proceedings of the 2015 3rd IAPR Asian Conference on Pattern Recognition, Kuala Lumpur, Malaysia, pp. 730-734. https://doi.org/10.1109/ACPR.2015.7486599

[13] Szegedy, C., Wei, L., Yangqing, J., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A. (2015). Going deeper with convolutions. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, pp. 1-9. https://doi.org/10.1109/CVPR.2015.7298594

[14] He, K., Zhang, X., Ren, S., Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition, Cairo, Egypt, pp. 770-778. https://doi.org/10.1109/CVPR.2016.90

[15] Chollet, F. (2017). Xception: Deep learning with depthwise separable convolutions. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 1800-1807. https://doi.org/10.1109/CVPR.2017.195

[16] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z. (2016). Rethinking the inception architecture for computer vision. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, pp. 2818-2826. https://doi.org/10.1109/CVPR.2016.308

[17] Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q. (2017). Densely connected convolutional networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, pp. 4700-4708. https://doi.org/10.1109/CVPR.2017.243

[18] Stephen, O., Sain, M., Maduh, U.J., Jeong, D.U. (2019). An efficient deep learning approach to pneumonia classification in healthcare. Journal of Healthcare Engineering, 2019(1): 4180949. https://doi.org/10.1155/2019/4180949

[19] Kermany, D.S., Goldbaum, M., Cai, W., Valentim, C.C.S., Liang, H., Baxter, S.L., McKeown, A., Yang, G., Wu, X., Yan, F. (2018). Identifying medical diagnoses and treatable diseases by image-based deep learning. Cell, 172(5): 1122-1131. https://doi.org/10.1016/j.cell.2018.02.010

[20] Ayan, E., Ünver, H.M. (2019). Diagnosis of pneumonia from chest X-ray images using deep learning. In 2019 Scientific Meeting on Electrical-Electronics & Biomedical Engineering and Computer Science (EBBT), Istanbul, Turkey, pp. 1-5. https://doi.org/10.1109/EBBT.2019.8741582

[21] Varshni, D., Thakral, K., Agarwal, L., Nijhawan, R., Mittal, A. (2019). Pneumonia detection using CNN-based feature extraction. In Proceedings of the 2019

IEEE International Conference on Electrical, Computer and Communication Technologies (ICECCT), Coimbatore, India, pp. 1-7. https://doi.org/10.1109/ICECCT.2019.8869364

[22] Wang, X., Peng, Y., Lu, L., Lu, Z., Bagheri, M., Summers, R.M. (2017). ChestX-Ray8: Hospital-scale chest X-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, pp. 3462-3471. https://doi.org/10.1109/CVPR.2017.369

[23] Chouhan, V., Singh, S.K., Khamparia, A., Gupta, D., Tiwari, P., Moreira, C., Damaševičius, R., de Albuquerque, V.H.C. (2020). A novel transfer learning-based approach for pneumonia detection in chest X-ray

images. Applied Sciences, 10(2): 559. https://doi.org/10.3390/app10020559

[24] Islam, S.R., Maity, S.P., Ray, A.K., Mandal, M. (2019). Automatic detection of pneumonia on compressed sensing images using deep learning. In 2019 IEEE Canadian Conference of Electrical and Computer Engineering (CCECE), Edmonton, AB, Canada, pp. 1-4. https://doi.org/10.1109/CCECE.2019.8861760

[25] Cohen, J.P., Morrison, P., Dao, L. (2020). COVID-19 image data collection. arXiv, arXiv:2003.11597. https://doi.org/10.48550/arXiv.2003.11597

[26] Cohen, J.P., Morrison, P., Dao, L., Roth, K., Duong, T.Q., Ghassemi, M. (2020). Covid-19 image data collection: Prospective predictions are the future. arXiv preprint arXiv:2006.11988.