International Journal of Computing and Digital Systems
ISSN (2210-142X)
Int. J. Com. Dig. Sys. 12, No.1 (Oct-2022)

https://dx.doi.org/10.12785/ijcds/120192

Medical Image Augmentation Framework for Resolving Chest X-Ray Data Imbalance

Rachna Sethi[1,2], Monica Mehrotra[1], Dhaarna Sethi and Gaurang Mehrotra

[1]*Department of Computer Science, Jamia Millia Islamia, Delhi, India*
[2]*Sri Guru Gobind Singh College of Commerce, University of Delhi, Delhi, India*

Received 22 Jan. 2021, Revised 15 Jul. 2022, Accepted 23 Jul. 2022, Published 31 Oct. 2022

Abstract: Deep learning techniques, particularly convolutional neural networks (CNNs), have led to an enormous breakthrough in the field of medical imaging. Since the onset of the COVID-19 pandemic, studies based on deep learning systems have shown excellent results for diagnosis through the use of Chest X-rays. However, these methods are data sensitive, and their effectiveness depends on the availability and reliability of data. Models trained on a class-imbalanced dataset tend to be biased towards the majority class. The class-imbalanced datasets can be balanced by augmenting them with synthetically generated images. This paper proposes a method for generating synthetic COVID-19 Chest X-Rays images using Generative Adversarial Networks (GANs). The images generated using the proposed GAN were augmented to three imbalanced datasets of real images. It was observed that the performance of the CNN model for COVID-19 classification improved with the augmented images. Significant improvement was seen in the sensitivity or recall, which is a very critical metric. The sensitivity achieved by adding GAN-generated synthetic images to each of the imbalanced datasets matched the sensitivity levels of the balanced dataset. Hence, the proposed solution can be used to generate images that boost the sensitivity of COVID-19 diagnosis to the level of a balanced dataset. Furthermore, this approach of synthetic data augmentation can be used in other medical classification applications for improved diagnosis recommendations.

Keywords: CNN, GAN, Deep Learning, Synthetic Images, Data Augmentation, Data Imbalance

1. Introduction

The COVID-19 pandemic that began in December 2019 is still affecting public health globally. Real-time PCR is a standard diagnostic tool being used for pathological testing, but the sensitivity data of these tests are not available. Chest X-Rays are significantly being used as an alternative tool for COVID-19 diagnosis due to their high sensitivity [1] [2]. A variety of deep learning convolutional neural network (CNN) based models have been proposed in the recent past that accurately diagnose COVID-19 patients through the use of chest X-rays [3] [4] [5]. However, the amount of COVID-19 Chest X-Rays images available on public domains is scant. The datasets are highly imbalanced, i.e. the amount of data belonging to each class differs significantly in number. The performance of deep learning CNN models degrades if the input data is imbalanced as models trained on a class-imbalanced dataset tend to be biased towards the majority class [6]. One solution to this problem is the generation of synthetic images that can be used to augment the dataset to make it balanced. Research in this area is still in its early stages, but it is getting positive attention [7].

Generative Adversarial Networks (GAN) is a powerful tool to generate images without supervision [8]. Since its inception in 2014 [9], GANs have been used to generate artificial data in multiple studies. The Vanilla GAN framework by Goodfellow et al. consists of two CNNs trained in an adversarial process where one network generates fake images and the other network repeatedly discriminates between real and fake images. This study aims to design a GAN framework for synthesising images from the set of COVID-19 Chest X-Rays for the augmentation of the real dataset to make it a balanced one. In addition, the framework for the CNN model for COVID-19 classification will be designed to test the effectiveness of the GAN framework. This CNN model will be trained with a real dataset as well as with an augmented dataset, and both performances will be compared.

The organisation of this paper is as follows: section 2 outlines related work in the generation of synthetic images using the GAN framework. The architecture of the proposed GAN model for the generation of synthetic images has been presented in section 3. Also, the architecture of the CNN model that will be used for performance comparison has been discussed in this section. Section 4 describes the dataset used for the experiment, setup and flow of the experiment process. Results obtained have been discussed

E-mail address: rachna.sethi@sggscc.du.ac.in, mmehrotra@jmi.ac.in, dhaarna29@gmail.com, gaurang.dtu@gmail.com

in section 5. The last section discusses the conclusions and future work on data augmentation using synthetic images.

2. Related Work

Many variations of GAN have been proposed in the recent past to overcome the limitations of Vanilla GAN proposed by Goodfellow et al. In Deep Convolutional GAN (DCGAN), the training of the model is stabilised with a set of constraints [10]. Conditional GAN (CGAN) trained the model and generated output conditioned to some auxiliary information [11]. Auxiliary Classifier GAN or ACGAN [12] generated more coherent global images. LAPGAN [13] uses cascaded CNN within the Laplacian pyramid framework for image synthesis. WGAN [14] improved the training stability and removed the mode collapse problem. InfoGAN [15] learns disentangled representations in an unsupervised manner. Progressive GAN or PCGAN can speed up as well as stabilize training [8]. In PCGAN, both generator and discriminator grow progressively, which results in the generation of high-quality images.

In recent years, GAN has been used for image synthesis in various domains such as text, music, video, etc.. Of late, it has revolutionalised the field of medical imaging. Talha Iqbal and Hazrat Ali proposed MI-GAN - a GAN framework for Medical Imaging and tested it successfully on retinal images [16]. There have been studies where artificial synthesis of medical images using GAN has significantly improved the accuracy of diagnosis. The most commonly adopted GANs for image synthesis in these studies are DCGAN, WGAN, and PGGAN, as they have better training stability [17]. G. Wang et al. proposed a DCGAN-based framework for augmentation of palm print dataset that improved the recognition ability [18]. Frangi et al. presented a 3D-Multi conditional GAN for augmentation of brain MRI images for efficient tumor detection [7]. Frid-Adar et al. modified DCGAN for augmentation of Liver CT-scans that resulted in remarkable improvement in the classification of liver lesions [19]. A PCGAN was trained by Beers et al. for the synthesis of fundus images having premature retinopathic vascular pathology [20].

Specific to GAN-based data augmentation of chest X-Rays datasets, models proposed by Zulkifley et al., Salehinejad et al. and Waheed et al. have demonstrated that COVID-19 detection can be improved by augmenting the dataset with synthetic images generated with their proposed frameworks. Waheed et al. proposed COVIDGAN - the GAN framework based on ACGAN for image synthesis [21] that achieved a classification accuracy of 95%. The GAN framework of Zulkifley et al. is based on DCGAN and had an accuracy of 97% [22]. Their model did not perform hyper-parameters tuning but applied separable convolution and simplified spatial pyramid pooling module to produce a lightweight network. Salehinejad et al. used a DCGAN-based framework to augment real datasets of chest X-Rays that increased the volume of data for training and balanced the dataset. It resulted in a significant improvement in classification performance [23]. In all these models, along with the GAN framework, the architectures of the classification models were also discussed. These classification models were used for the performance evaluation of the proposed GAN frameworks. But Loey et al. presented a GAN model with deep transfer learning for coronavirus detection in chest X-ray images [24]. They used three deep transfer models - Alexnet, Googlenet, and Restnet18. These models contain a small number of layers that help in reducing the complexity, the consumed memory and the execution time.

3. Proposed Architecture

This section describes two architecture –

1) GAN for generating chest X-Ray images and
2) CNN for COVID-19 detection.

A. GAN Framework

The proposed framework is loosely based on DCGAN architecture that has been strengthened by incorporating the features of WGAN to provide more stability. The base architecture of GAN follows the DCGAN architecture of Radford et al. [10] that has two constituents - a discriminator D and a generator G as shown in Figure 1. Both the G and D networks are deep CNNs and are trained concurrently. The task of generator G is to learn the mapping of random input noise to given 2D input data. In contrast, the discriminator D maps these 2D images to an output representing the probabilities of the image being 'real' or being 'fake'. Both these components work in a minimax fashion where generator G tries to minimise the loss by producing more realistic images, and discriminator D tries to maximise this loss by learning how to differentiate between real and fake images more accurately. Two concurrent feedback loops run through the model that pit against each other to produce optimal results.

Mathematically, Let

Pz = simple noise distribution of input z

Pr = original data distribution over real sample x

Pg = generated image distribution over data x

Generator G gets input from input samples and maps G(z) to Pg. The goal of G is to achieve Pg = Pr. The image x generated by G becomes the input for D. Output of D is D(x), probability of x being a real sample. During the training, discriminator D tries to maximize D(x) for images with x $\simeq$ Pr and minimize D(x) for images with x Pr. G generates images G(z) to fool D during training such that D(G(z)) $\sim$ Pr. Therefore, G is trained to maximize D(G(z)), or equivalently minimize 1 - D(G(z)) [19]. Thus, the loss function can be represented as

$$E_x[log(D(x))] + E_z[1 - log(D(G(z))] \quad (1)$$

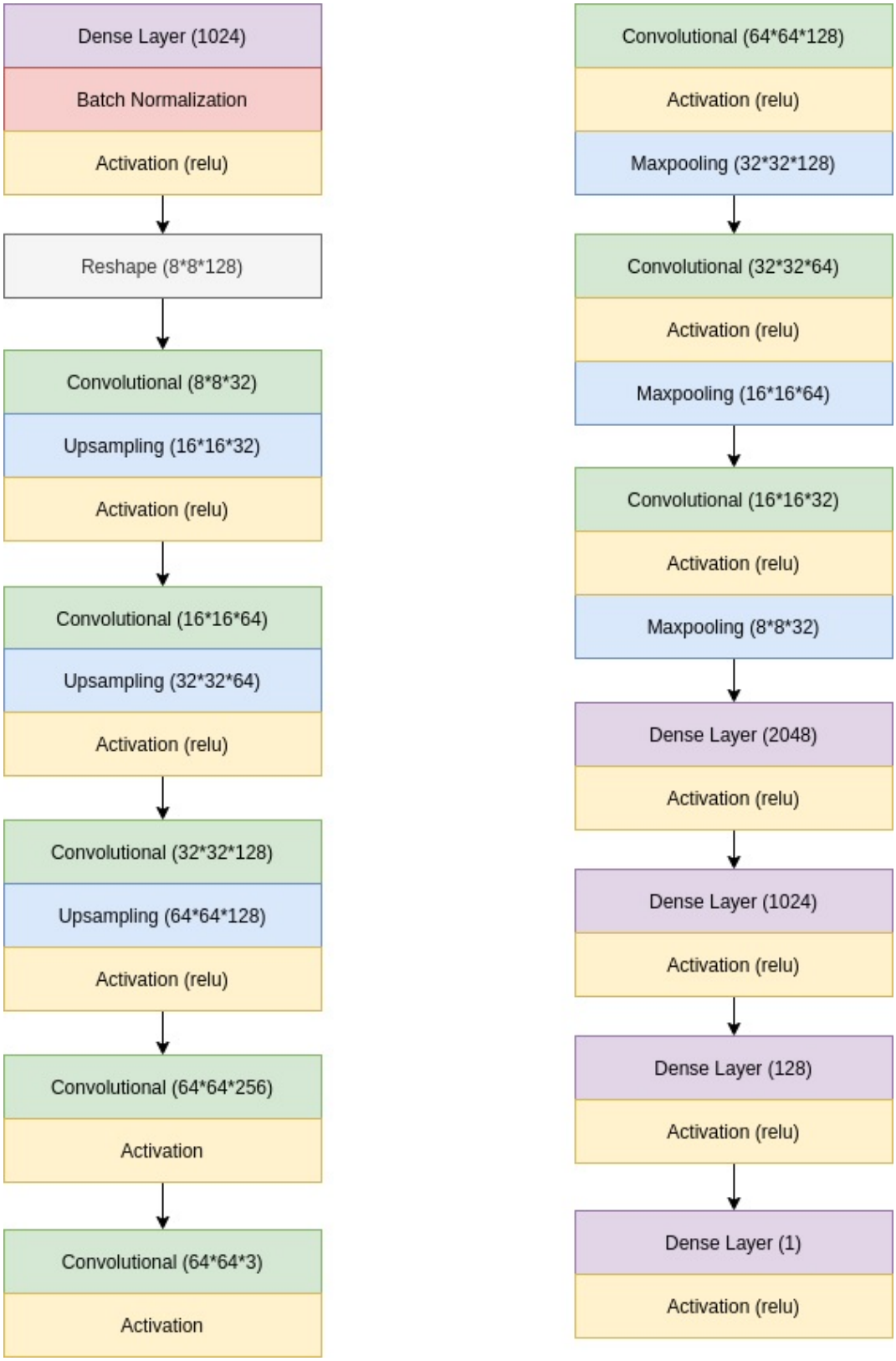


Figure 1. Visualisation of GAN architecture - Generator and Discriminator

Training of these two adversarial networks is the optimization of this loss function. The traditional loss functions are based on Kullback-Leibler (KL) and Jensen-Shannon (JS) functions. But here, Wasserstein loss was applied to both networks for stabilising the training. Wasserstein loss measures the distance between two probability distributions, Pg and Pr [25]. Also known as Earth Mover (EM) Distance, it is essentially the effort required to move one distribution to the other. It is represented as

$$L(p_r, p_g) = W(p_r, p_g) = \max_{wW} E_{xp_r}[f_w(x)] - E_{zp_r}[f_w(g_\theta(z))] \quad (2)$$

The function f belongs to the family of K-Lipschitz continuous functions parameterised by w. Discriminator D, during training, learns a K-Lipschitz continuous function and calculates Wasserstein distance. As the loss function decreases in training, the Wasserstein distance gets smaller and the generator model's output grows closer to the real data distribution.

For Vanilla GANs, the discriminator gives a binary classification on the input as real or fake, whereas, in WGAN the discriminator measures the realness or fakeness of the image. Thus, the discriminator is no longer a binary classifier. Instead, it is used as a regression model that gives the EM distance between the input and the real data distribution. The benefit of using this loss function over the traditional functions is that the training process becomes much more stable and produces better quality images.

Proposed GAN architecture is different from other DCGAN-based architectures in the following ways:

1) DCGAN architecture of Radford et al. had a set of constraints for stabilized training. In the proposed model, the constraints were not strictly applied. The differences are as follows:
   a) batch normalization was not applied to convolutional layers,
   b) instead of LeakyRelu activation, only Relu activation was used in both generator G and discriminator D.
2) Instead of the traditional loss functions, Wasserstein loss is applied to stabilise the training.

The architectures of generator and discriminator are discussed below:

Generator Architecture: A diagrammatic representation of the generator model is given in Figure 1. It takes a noise vector as input, derived from a random uniform distribution spread between the values -1.0 and 1.0. This vector is then reshaped using a dense layer to the dimensions of 128x8x8 and fed to convolution blocks. Four sequential convolutional blocks are used in the model, where each block consists of a convolution layer, an upsampling layer, and an activation layer. The convolution layers apply the filter maps to the input using a kernel size of 5x5 and the output is fed to the upsampling layer. The image is then upsampled by 2x2 i.e the rows and columns are repeated two times each. Finally, the activation layer applies the tanh activation function to the image and passes it to the next block. The final output of the model is of size 64x64x3.

Discriminator Architecture: The discriminator model takes an input image of size 64x64x3 and gives a single output giving the EM distance. The model has a series of convolutional blocks consisting of a convolutional layer, a max-pooling layer, and an activation layer. Similar to the generator, the convolution layer applies the filter maps on the input using a kernel size of 5x5. Then, the output is fed to the max-pooling layer which downsamples the image or halves it by using a pool size of 5x5. The activation function applied to the image is tanh. The output is then fed to the next convolutional block, which performs similar functions. After three such processes, the image is flattened and fed to a dense layer consisting of 1024 neurons. The final output layer has one neuron and sigmoid activation is applied to it. The architecture of the discriminator is visualised in Figure 1.

B. CNN Architecture

Many CNN architectures have been proposed by various researchers for the detection of chest diseases using chest X-Rays [26] [27] [19]. The CNN model, that will be used to compare the performance parameters, has been derived from these studies. The proposed CNN model consists of four convolutional blocks each comprising of a convolutional layer, activation layer, and a max-pooling layer. The first two convolutional layers have a filter size of 3x3 and the next two have a filter size of 5x5. The pool size used for max-pooling is 2x2 in all four blocks, and the activation function used is Relu. The output of these convolutional blocks is flattened and fed to a fully connected dense layer having 1000 neurons. Finally, sigmoid activation is applied in the softmax layer to get the binary classification output. The model is compiled using binary cross-entropy loss and Adam optimizer. A diagrammatic representation of the model and its working is given in Figure 2.

4. Experiment

A. Dataset

To generate the datasets for experiment purposes, images were collected from publicly available datasets of COVID-19 X-Rays on Github [28] and Kaggle [29]. Figures 3(a) and 3(b) illustrate some of the images collected from these domains. It was observed that only a scarce number of COVID-19 positive chest X-Rays were available on these public domains. Also, the dataset was a mix of X-ray images and CT-Scan images. CT-Scans, X-Rays of a single lung, and side X-Ray images were manually removed to enhance the quality of the dataset. Finally, 352 COVID-19 positive X-Rays were selected to be used for the generation of synthetic images. Another set of 353 Normal chest X-Rays

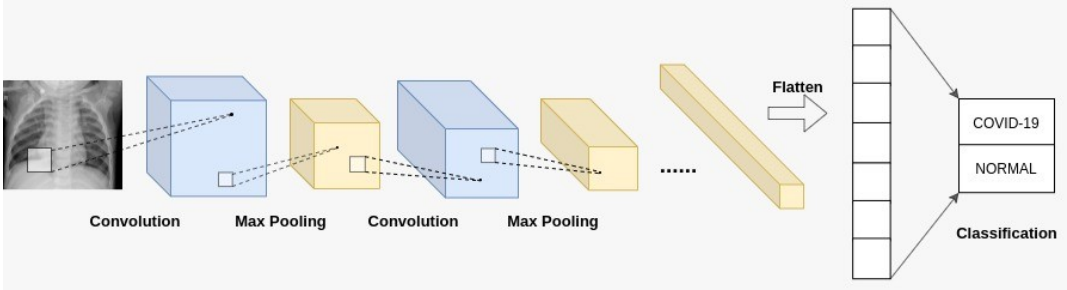
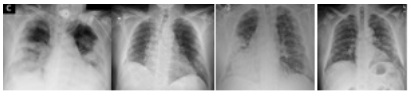
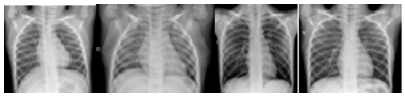


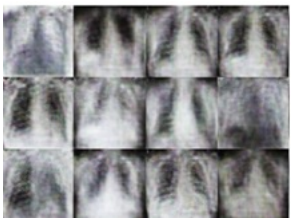
Figure 2. Performance evaluation of augmented dataset using CNN model



(a)



(b)



(c)

Figure 3. (a) COVID-19 infected real chest X-Rays (b) Normal Chest X-Rays (c) COVID-19 infected generated chest X-Rays

were also selected and a balanced dataset of 705 images was created having two class labels – Normal and COVID-19. This dataset was called DS4-B. Further, three datasets of imbalanced class data were created. These datasets were named DS1–I, DS2–I, and DS3-I. The ratio of COVID-19 X-Rays to Normal X-Rays varied from 2:15, 1:5, and 2:5 in these datasets respectively. The composition of these datasets is shown in Figure 4.

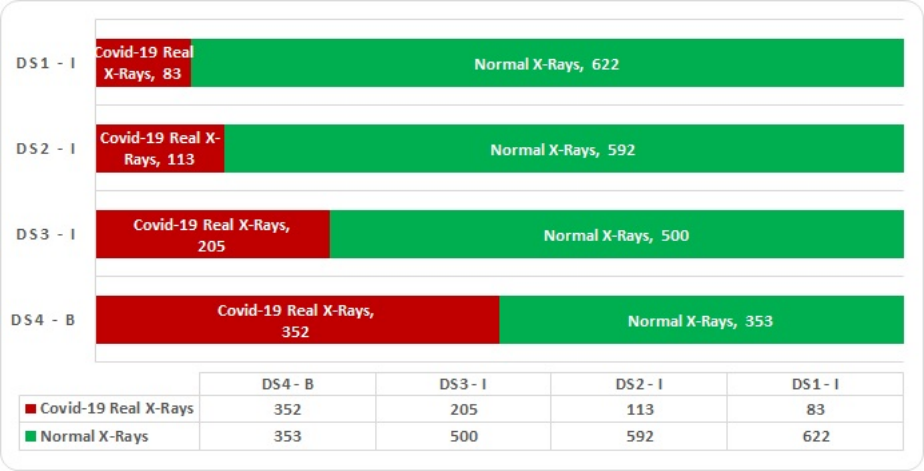
B. Setup and Process Flow

CNN and GAN architecture have been implemented using Keras deep learning libraries with Tensorflow framework in the back-end. All training and testing processes were executed on an Intel Broadwell 6 vCPU with 15GB RAM. The first step was the generation of synthetic images. As depicted in Figure 5, the process started with simultaneous training of Generator and Discriminator using 352 real images of COVID-19 positive chest X-Rays. In the first step, images were resized and normalised from [0, 255] to [1, 1]. Normalisation was applied to change the range of pixel values. The input images were converted into a range of more familiar pixel values. For optimisation, Adam optimizer was used due to its known efficiency. The Generator had around 9 million parameters and the Discriminator had 2.5 million parameters. The GAN model took around 5 hours to train. The training process was run for 400 epochs with a batch size of 64 and a learning rate of 0.002.

The trained generator was used to generate synthetic images for balancing the class difference in the three imbalanced datasets - DS1-I and DS2-I and DS3-I. Three new balanced datasets were created as DS1-B, DS2-B, and DS3-B from DS1-I, DS2-I, and DS3-I respectively by augmenting them with generated synthetic images. Some samples of images generated have been illustrated in Figure 3(c). It was ensured that the size of all datasets will be the same and balanced datasets have the same class composition as of DS4-B. The class composition of these four balanced datasets is shown in Figure 6. The next step required testing the performance of the proposed CNN model for COVID-19 diagnosis on these datasets. The training set and test set were split before image augmentation to ensure that the accuracy received on the testing set was unbiased. The CNN model has 2.3 million parameters and was iterated on each of the seven datasets for 100 epochs with batch size 64 and a learning rate of 0.0001. The process flow of this performance evaluation is depicted in Figure 2.

5. Results and Discussion

The performance of the CNN model is evaluated based on the following metrics:



	DS4 - B	DS3 - I	DS2 - I	DS1 - I
Covid-19 Real X-Rays	352	205	113	83
Normal X-Rays	353	500	592	622

Figure 4. Class Composition of unbalanced datasets created for the experiment

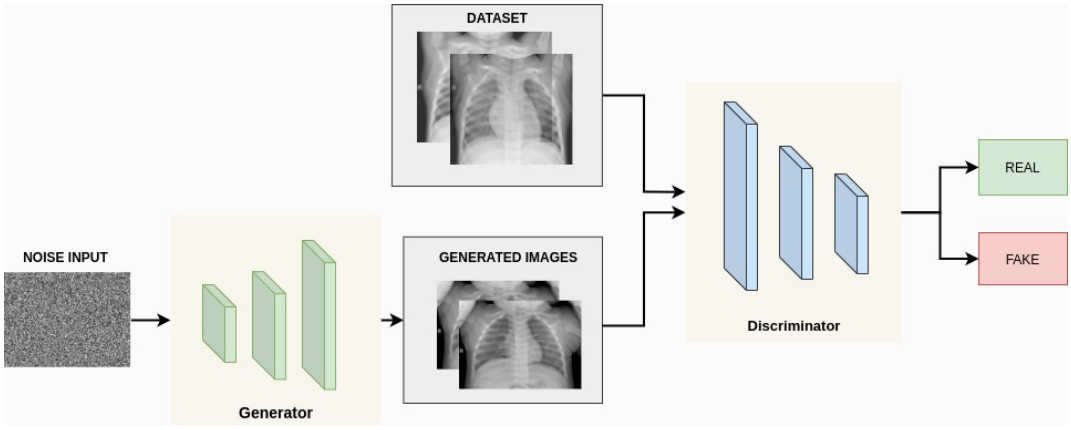


Figure 5. Process Flow of generating synthetic Images

- **Accuracy:** Ratio of correctly predicted outcomes to all the prediction

$$Accuracy = \frac{(TN + TP)}{(TN + TP + FN + FP)} \tag{3}$$

- **Precision:** Number of correct predictions

$$Precision = \frac{TP}{(TP + FP)} \tag{4}$$

- **Recall/Sensitivity:** Also known as True Positive rate, it is the number of correct positive predictions out of all the positive predictions.

$$Recall/Sensitivity = \frac{TP}{(TP + FN)} \tag{5}$$

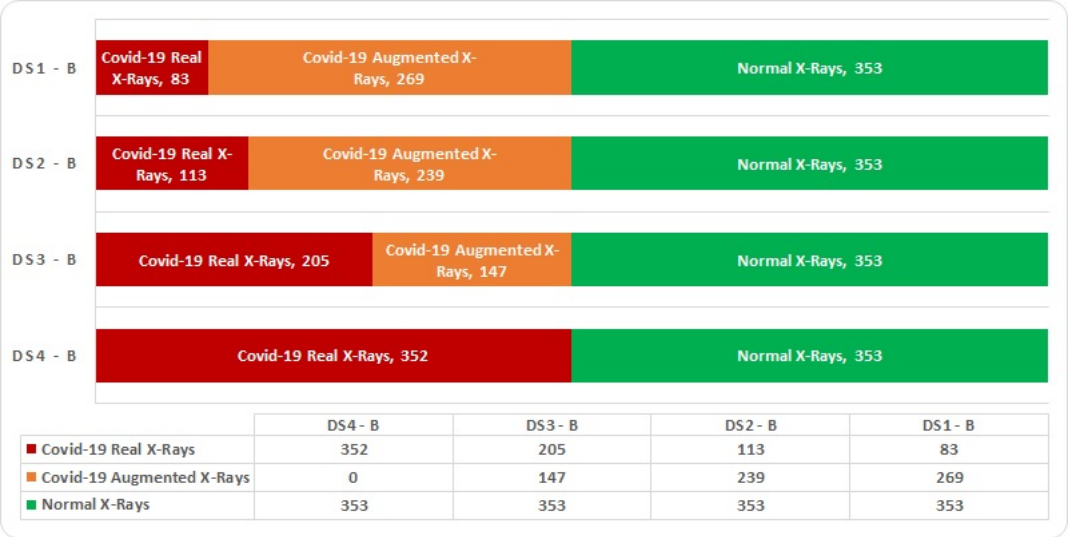
- **Specificity:** Also known as True Negative rate, it is the proportion of actual negatives that were correctly predicted.

$$Specificity = \frac{TN}{(TN + FP)} \tag{6}$$

- **F1 Score:** Harmonic mean of precision statistic and the recall statistic. In comparison to Accuracy, the F1 Score is a better measure for estimating incorrectly classified cases

$$F1\ Score = 2 * \frac{(Precision * Recall)}{(Precision + Recall)} \tag{7}$$

Here TP is true positive, TN is true negative, FP is false positive and FN is false negative. The values observed for the above-discussed parameters for all the seven datasets for COVID-19 diagnosis using the proposed CNN model have been presented in Table I and Table II. DS1-I, DS2-I, and DS3-I are imbalanced datasets of actual images. DS1-B, DS2-B, and DS3-B are their balanced counterparts consisting of synthetic images. DS4-B is a balanced dataset of real images. Performance of all experiments was measured using test sets that were isolated from the augmented images and the training set to ensure real and fair results. The confusion matrices of augmented balanced datasets DS1-B, DS2-B, and DS3-B are shown in Figure 7, Figure 8, and Figure 9 respectively. Figure 10 depicts the confusion matrix of



	DS4 - B	DS3 - B	DS2 - B	DS1 - B
Covid-19 Real X-Rays	352	205	113	83
Covid-19 Augmented X-Rays	0	147	239	269
Normal X-Rays	353	353	353	353

Figure 6. Class Composition of balanced datasets created from unbalanced datasets by augmenting with generated images

DS4-B, the balanced dataset of real images. Table I is a comparison of performance parameters of the balanced and unbalanced dataset for each of DS1, DS2, and DS3. The performances of all the balanced datasets DS1-B, DS2-B, DS3-B and DS4-B have been compared in Table II.

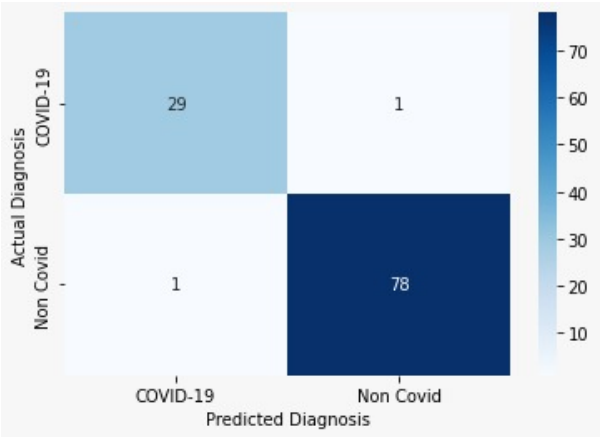


Figure 7. Confusion Matrix for Dataset DS1-B

In medical classification domains, Sensitivity or Recall is a very critical metric as classifying true positives correctly holds higher importance from the context of a patient. The sensitivity of real-time PCR, the standard diagnostic tool being used for COVID-19 detection, is largely unknown and therefore the results obtained cannot be assessed or ascertained with confidence. Thus, the main aim of the project was to improve the sensitivity of the model. It was observed that the impact of data augmentation with synthetic images on sensitivity metrics was quite discernible. By comparing the F1 score and sensitivity values for all datasets given in Table I it was observed that by adding augmented GAN-generated synthetic images to balance the imbalanced dataset, the model's performance showed an increase and matched the results obtained using the balanced set. The sensitivity obtained by the model significantly increased from

- 85% to 96% for Dataset DS1
- 88% to 97% for Dataset DS2
- 92% to 98% for Dataset DS3

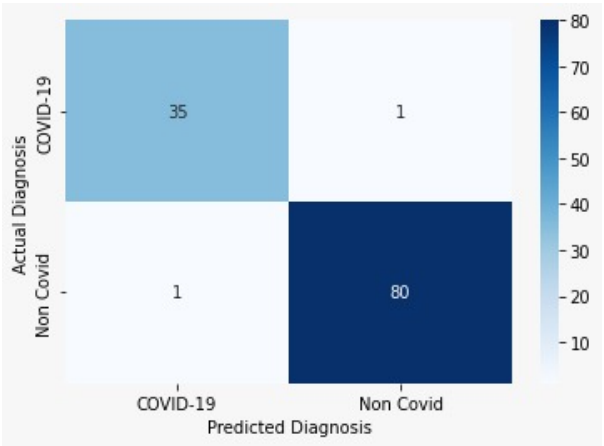


Figure 8. Confusion Matrix for Dataset DS2-B

A visual depiction of this performance is illustrated in Figure 11. It can also be seen that Dataset DS1 which had the least quantity of minority class images exhibited a maximum increase in performance. Since all models were trained on an eclectic dataset and tested against isolated test sets it can be ensured that GAN-generated images do not overfit the model and produce equally accurate results

TABLE I. Comparative of performance of balanced and imbalanced datasets

Dataset	Accuracy	Precision	Specificity	Sensitivity	F1 Score
DS1-I (Imbalanced)	0.977	0.973	0.99	0.857	0.986
DS1-B (Balanced)	0.981	0.987	0.987	0.966	0.987
DS2-I (Imbalanced)	0.966	0.972	0.985	0.888	0.978
DS2-B (Balanced)	0.982	0.987	0.987	0.972	0.987
DS3-I (Imbalanced)	0.977	0.968	0.99	0.928	0.983
DS3-B (Balanced)	0.985	0.987	0.987	0.983	0.987

TABLE II. Comparative of performance of balanced datasets of synthetic images with the balanced dataset of real images

Dataset	Accuracy	Precision	Specificity	Sensitivity	F1 Score
DS1-B	0.981	0.987	0.987	0.966	0.987
DS2-B	0.982	0.987	0.987	0.972	0.987
DS3-B	0.985	0.987	0.987	0.983	0.987
DS4-B	0.988	0.986	0.986	0.990	0.986

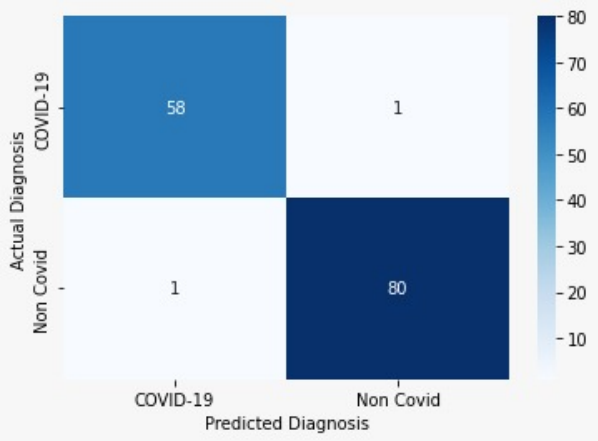


Figure 9. Confusion Matrix for Dataset DS3-B

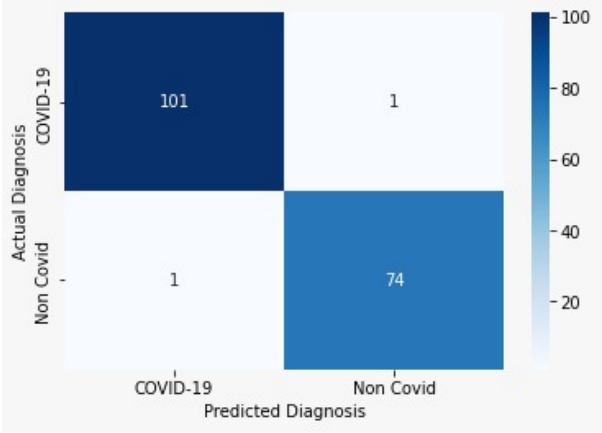


Figure 10. Confusion Matrix for Dataset DS4-B

for all sets. Here it was observed that the proposed GAN model is capable of improving existing performance results even with the minimal number of COVID-19 Chest X-Ray images.

Another significant observation from Table II is that the performance parameters achieved with the augmented images were at par with the balanced real images dataset. The datasets DS1-B, DS2-B, and DS3-B that were augmented with GAN-generated synthetic images achieved the same performance parameters as that of the DS4-B dataset of actual images. In other words, if the results obtained by the classification model on the DS4-B dataset are set as benchmark results, then it was observed that the images generated by the proposed GAN model boosted the performance obtained by all three imbalanced datasets to this benchmark level. Although all three datasets contain different ratios of imbalanced classes, after balancing the sets by adding augmented images the results are at par with the balanced dataset.This shows that the proposed solution can be used to create images that boost the sensitivity of the model in every case to the level of a balanced dataset.

TABLE III. Comparison of average accuracy of the proposed model with average accuracy from other GAN-based data augmentation models for Chest X-Rays.

Model	Type of GAN	Accuracy
Proposed Model	**DCGAN & WGAN**	**98**
[21]	ACGAN	95
[22]	DCGAN	97
[24]	GAN with AlexNet	80.6
[24]	GAN with GoogleNet	85.2
[24]	GAN with Resnet18	100
[23]	DCGAN	92

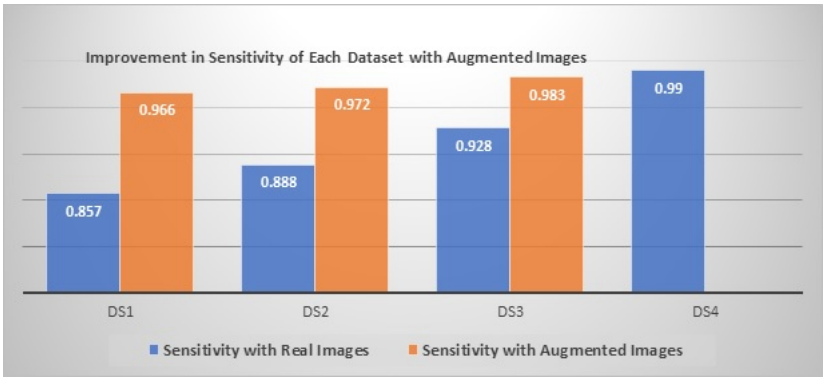


Figure 11. Improvement in sensitivity with augmented images

Table III makes a comparison between the average accuracy of the proposed model with those from other GAN-based data augmentation models for Chest X-Rays. Waheed et al. [21] were able to achieve the sensitivity of 90% and accuracy of 95% with their ACGAN-based model for generating synthetic COVID-19 infected Chest X-Rays on a similar dataset. DCGAN-based model of Zulkifley et al. had an accuracy of 97%. Although DCGAN-based model of Salehinejad et al. using Resnet18 for classification [23], could achieve 100% accuracy but all other models had accuracy and sensitivity less than the proposed model. Hence, the proposed architecture for GAN can be considered efficient and effective.

6. Conclusions

In this paper, a model has been proposed for generating synthetic COVID-19 positive chest X-Rays. The proposed framework is loosely based on DCGAN architecture but strengthened by WGAN to provide more stability. The images generated with the proposed GAN were used to resolve the data imbalance problem in COVID-19 datasets available in the public domain. The synthetic images were generated using the proposed GAN and then added to three imbalanced datasets of real images. The ratio of COVID-19 X-Rays to Normal X-Rays varied from 2:15, 1:5, and 2:5 in the three imbalanced datasets. Augmenting these imbalanced datasets with generated images not only brought variability in the training dataset but proved quite effective in improving the sensitivity of the CNN model for COVID-19 diagnosis. The sensitivity improvement observed in these three augmented datasets was 11%, 9% and 6%, respectively. And most significantly, the sensitivity achieved in all these datasets was almost the same and matched the sensitivity score of the balanced real images dataset. Thus, if the performance parameters of the balanced real images dataset are taken as benchmark results, then it can be concluded that the images generated by the proposed GAN model boosted the performance obtained by all three imbalanced datasets to this benchmark level. These results prove that the proposed model is effective on all kinds of data imbalance and can be used effectively in all cases of COVID-19 data scarcity.

The future scope of this research includes extending it to cover a wider range of pulmonary diseases. Furthermore, similar synthetic data augmentation framework design will be worked upon for other medical classification applications.

References

[1] C. Huang, Y. Wang, X. Li, L. Ren, J. Zhao, Y. Hu, L. Zhang, G. Fan, J. Xu, X. Gu, Z. Cheng, T. Yu, J. Xia, Y. Wei, W. Wu, X. Xie, W. Yin, H. Li, M. Liu, Y. Xiao, H. Gao, L. Guo, J. Xie, G. Wang, R. Jiang, Z. Gao, Q. Jin, J. Wang, and B. Cao, "Clinical features of patients infected with 2019 novel coronavirus in wuhan, china," vol. 395, no. 10223, pp. 497–506. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0140673620301835

[2] D. Wang, B. Hu, C. Hu, F. Zhu, X. Liu, J. Zhang, B. Wang, H. Xiang, Z. Cheng, Y. Xiong, Y. Zhao, Y. Li, X. Wang, and Z. Peng, "Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus–infected pneumonia in wuhan, china," vol. 323, no. 11, p. 1061. [Online]. Available: https://jamanetwork.com/journals/jama/fullarticle/2761044

[3] R. M. Pereira, D. Bertolini, L. O. Teixeira, C. N. Silla, and Y. M. Costa, "COVID-19 identification in chest x-ray images on flat and hierarchical classification scenarios," vol. 194, p. 105532. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0169260720309664

[4] M. Z. Islam, M. M. Islam, and A. Asraf, "A combined deep CNN-LSTM network for the detection of novel coronavirus (COVID-19) using x-ray images," vol. 20, p. 100412. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S2352914820305621

[5] R. Sethi, M. Mehrotra, and D. Sethi, "Deep learning based diagnosis recommendation for COVID-19 using chest x-rays images," in *2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA)*. IEEE, pp. 1–4.

[6] M. A. Mazurowski, P. A. Habas, J. M. Zurada, J. Y. Lo, J. A. Baker, and G. D. Tourassi, "Training neural network classifiers for medical decision making: The effects of imbalanced datasets on classification performance," vol. 21, no. 2, pp. 427–436. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0893608007002407

[7] A. F. Frangi, S. A. Tsaftaris, and J. L. Prince, "Simulation and synthesis in medical imaging," vol. 37, no. 3, pp. 673–679. [Online]. Available: http://ieeexplore.ieee.org/document/8305584/

[8] T. Karras, T. Aila, S. Laine, and J. Lehtinen, "Progressive growing of GANs for improved quality, stability, and variation." [Online]. Available: http://arxiv.org/abs/1710.10196

[9] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks." [Online]. Available: http://arxiv.org/abs/1406.2661

[10] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks." [Online]. Available: http://arxiv.org/abs/1511.06434

[11] M. Mirza and S. Osindero, "Conditional generative adversarial nets." [Online]. Available: http://arxiv.org/abs/1411.1784

[12] A. Odena, C. Olah, and J. Shlens, "Conditional image synthesis with auxiliary classifier GANs." [Online]. Available: http://arxiv.org/abs/1610.09585

[13] E. Denton, S. Chintala, A. Szlam, and R. Fergus, "Deep generative image models using a laplacian pyramid of adversarial networks." [Online]. Available: http://arxiv.org/abs/1506.05751

[14] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein GAN." [Online]. Available: http://arxiv.org/abs/1701.07875

[15] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. Courville, "Improved training of wasserstein GANs." [Online]. Available: http://arxiv.org/abs/1704.00028

[16] T. Iqbal and H. Ali, "Generative adversarial network for medical images (MI-GAN)," vol. 42, no. 11, p. 231. [Online]. Available: http://link.springer.com/10.1007/s10916-018-1072-9

[17] X. Yi, E. Walia, and P. Babyn, "Generative adversarial network in medical imaging: A review," vol. 58, p. 101552. [Online]. Available: http://arxiv.org/abs/1809.07294

[18] G. Wang, W. Kang, Q. Wu, Z. Wang, and J. Gao, "Generative adversarial network (GAN) based data augmentation for palmprint recognition," in *2018 Digital Image Computing: Techniques and Applications (DICTA)*. IEEE, pp. 1–7.

[19] M. Frid-Adar, I. Diamant, E. Klang, M. Amitai, J. Goldberger, and H. Greenspan, "GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification," vol. 321, pp. 321–331. [Online]. Available: http://arxiv.org/abs/1803.01229

[20] A. Beers, J. Brown, K. Chang, J. P. Campbell, S. Ostmo, M. F. Chiang, and J. Kalpathy-Cramer, "High-resolution medical image synthesis using progressively grown generative adversarial networks." [Online]. Available: http://arxiv.org/abs/1805.03144

[21] A. Waheed, M. Goyal, D. Gupta, A. Khanna, F. Al-Turjman, and P. R. Pinheiro, "CovidGAN: Data augmentation using auxiliary classifier GAN for improved covid-19 detection," vol. 8, pp. 91 916–91 923. [Online]. Available: https://ieeexplore.ieee.org/document/9093842/

[22] M. A. Zulkifley, S. R. Abdani, and N. H. Zulkifley, "COVID-19 screening using a lightweight convolutional neural network with generative adversarial network data augmentation," vol. 12, no. 9, p. 1530. [Online]. Available: https://www.mdpi.com/2073-8994/12/9/1530

[23] H. Salehinejad, S. Valaee, T. Dowdell, E. Colak, and J. Barfett, "Generalization of deep neural networks for chest pathology classification in x-rays using generative adversarial networks," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, pp. 990–994.

[24] M. Loey, F. Smarandache, and N. E. M. Khalifa, "Within the lack of chest COVID-19 x-ray dataset: A novel detection model based on GAN and deep transfer learning," vol. 12, no. 4, p. 651. [Online]. Available: https://www.mdpi.com/2073-8994/12/4/651

[25] M. Arjovsky and L. Bottou, "Towards principled methods for training generative adversarial networks." [Online]. Available: http://arxiv.org/abs/1701.04862

[26] R. H. Abiyev and M. K. S. Ma'aitah, "Deep convolutional neural networks for chest diseases detection," vol. 2018, pp. 1–11. [Online]. Available: https://www.hindawi.com/journals/jhe/2018/4168538/

[27] J. Ker, L. Wang, J. Rao, and T. Lim, "Deep learning applications in medical image analysis," vol. 6, p. 15.

[28] J. P. Cohen, "ieee8023/covid-chestxray-dataset," original-date: 2020-02-14T23:22:23Z. [Online]. Available: https://github.com/ieee8023/covid-chestxray-dataset

[29] CoronaHack -chest x-ray-dataset. [Online]. Available: https://kaggle.com/praveengovi/coronahack-chest-xraydataset

Rachna Sethi is a Ph.D. student under Professor Monica Mehrotra in Department of Computer Science at Jamia Millia Islamia in New Delhi, India. Her research area include Machine learning-based healthcare applications. In 1993, she earned a master's degree in computer applications from the Institute of Engineering and Technology in Lucknow, India. She has sixteen years of teaching experience and over twelve years of industry experience. She has been teaching at the University of Delhi for past sixteen years.

Monica Mehrotra is presently working as Professor in the Department of Computer Science, Jamia Millia Islamia (Central University). She completed her Ph.D. in August 2007 from Jamia Millia Islamia. She has over twenty- five years of teaching experience. Her research interest areas include Data Mining, Information Retrieval and Social network analysis. She has published over 70 papers in International conferences and Journals of repute. She has won 'Excellent Researcher Award (female)' in 2nd International academic and research excellence awards (IARE – 2020). She was presented with 'International Distinguished Education Excellence Award 2021' in Global Annual Education and Research Excellence Awards 2021 organised by Centre for Professional Advancement, India.

Dhaarna Sethi is currently employed as a Software Engineer at Visa Inc. In May 2020, she graduated with a Bachelor's degree in Information Technology from Delhi Technological University, Delhi, India (previously Delhi College of Engineering) . As an undergraduate, she has worked on many research projects in the field of Computer Vision and Natural Language Processing.

Gaurang Mehrotra is currently working at Amazon as a Software Development Engineer. He completed his Computer Engineering in June of 2020 from Delhi Technological University (formerly Delhi College of Engineering), Delhi, India. He graduated in First Division with Distinction. His research interests include image processing and artificial intelligence. He also won the second prize in Global Hackathon that was conducted by Optum.