Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 218 (2023) 283–290

Procedia Computer Science

www.elsevier.com/locate/procedia

International Conference on Machine Learning and Data Engineering

Comparison of Affine and DCGAN-based Data Augmentation Techniques for Chest X-Ray Classification

Mayank Bali[a],Tripti Mahara[b]

[a,b]*SBM, Christ University, Bangalore,560029, India*

Abstract

Data augmentation, also called implicit regularization, is one of the popular strategies to improve the generalization capability of deep neural networks. It is crucial in situations where there is a scarcity of high-quality ground-truth data. Also getting new samples is expensive and time consuming. This is a typical issue in the medical domain. Therefore, this study compares the performance of Affine and Generative Adversarial Networks (GAN)- based data augmentation techniques on the chest image X-Ray dataset. The Pneumonia dataset contains 5863 chest X-Ray images. The traditional Affine data augmentation technique is applied as a pre-processing technique to various deep learning-based CNN models like VGG16, Inception V3, InceptionResNetV2, DenseNet-169 and DenseNet-202 to compare their performance. On the other hand, DCGAN architecture is applied to the dataset for augmentation. Evaluation measures like accuracy, recall and AUC depict that DCGAN outperforms other traditional models. The most important advantage of DCGAN is that it is able to identify fake images with 100% accuracy. This is especially relevant for the medical domain as it deals with the life of individuals. Thus, it can be concluded that DCGAN has better performance as compared to affine transformations applied to traditional CNN models .

Keywords: Data Augmentation; Generative adversarial networks (GANs); Affine; Deep Learning; Medical Domain

1. Introduction

Deep Learning is becoming one of the fastest growing Machine Learning areas. It is used in a variety of domains to accomplish classification and forecasting and works well with different types of data. Convolutional Neural Networks (CNN), one of the types of Deep Learning models, are predominantly used for image analysis and classification. Although CNN models work very well for images, there are some challenges that need to be overcome. Large amount of data is needed to develop CNN models. Obtaining large amounts of ground-truth data is a challenge. This becomes more difficult especially in the medical domain as the process is time-consuming and costly. Also, there are privacy issues related to sharing of the images. Also furthermore, manually annotated image sets are unbalanced, with instances from certain classes under-represented. To overcome this challenge, data augmentation techniques are deployed [1]. Data augmentation is a set of techniques that are used to increase the number and quality of images to generate a large training dataset [2]. It also tackles the overfitting problem by artificially inflating the training dataset size by either

10.1016/j.procs.2023.01.010

data warping or oversampling. Transformation of existing images is accomplished using data warping augmentation methods. The labels of the existing images are retained [3].
To understand the relevance of data augmentation in the medical domain, the study comprises analyzing Affine and DCGAN data augmentation techniques. To achieve this, the X-Ray images of patients suffering from Pneumonia along with normal individual scans are used as the input. This dataset is predominantly selected as Pneumonia is one of the diseases that affects around 150 million people each year. Children under the age of five, particularly in developing countries are majorly affected by Pneumonia and hence proper diagnosis is inevitable. With COVID outbreak, correct diagnosis of pneumonia became more relevant as people affected with COVID get pneumonia that leads to untimely death. Therefore, proper identification of the disease from the images will provide help to the doctors to plan the treatment along with saving time and money for individuals [4].

In this research, Generative Adversarial Network (GAN) based data augmentation technique combined with DCGAN architecture is compared with traditional Affine data augmentation techniques when applied to various Convolutional Neural Network (CNN) models to compare the performance of the models. This can potentially be used in the medical industry to handle image classification that will lead to timely and accurate diagnosis of the disease. The contributions of the study are summarized as follows:

a. The traditional Affine data augmentation techniques are applied to the input data set and the performance of various CNN models like VGG16, Inception V3, InceptionResNetV2, DenseNet-169 and Densenet-202 is evaluated.
b. DCGAN based data augmentation is done and the performance is evaluated.
c. The performance of traditional CNN models is compared with DCGAN to find the best technique for classification.

The structure of the paper is as follows. Literature survey is presented in section 2 followed by Methodology in section 3. Model analysis and conclusion are presented in section 4 and 5 respectively.

2. **Background Study**

The primary goal of any data augmentation technique is to generate more data. Combination of Affine transformation and color modifications are the most widely used traditional augmentation techniques [5]. One of the first applications of CNN using data warper was developed for handwritten digit classification [6]. AlexNet CNN architecture revolutionized data augmentation using ImageNet dataset [7]. Advanced data augmentation techniques include GAN [8] and Neural Architecture Search [9] that diversifies data sets by generating new samples. Constructing artificial instances with similar features from an original dataset is generative modelling.

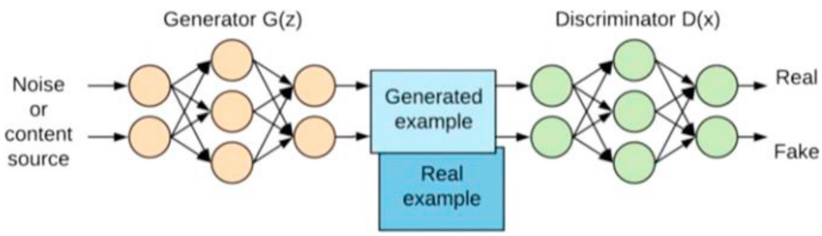

Fig. 1. Illustration of GAN concept provided by Mikolajczyk and Grochowski (M. & G, 2018)

GANs are one of the best generative modelling tools available because of the computation speed and performance quality. Ian Goodfellow proposed the first GAN architecture [8]. It can be well understood through the example of a cop and an imposter. The imposter which is the generator network takes input in the form of either vector, text, image etc. The cop is the discriminator network and it cannot predict whether the content produced by the imposter is real or fake. The real vs fake difference corresponds to whether images were generated by the generator network or it came from the training set as in Fig.1. This helps in distinguishing the real images from the fake ones.Various extensions of GANs namely Deep Convolutional Generative Adversarial Network (DCGAN) [10], CycleGAN [11] and

Progressively Growing GANs [12] are published. Although GAN based techniques are very powerful data generating techniques [13], how to generate high-quality data and evaluate them still remains a challenging problem.

Many studies have shown use of both techniques in the medical domain. For instance, Nalepa et.al [14] presented a review on use of Affine techniques for data augmentation in brain tumor images and concluded that the images produced are highly correlated to the original images. Also, studies have exhibited that Affine transformation does not provide any benefit in terms of generalization [13,15]. Liu et.al [16] presented literature review on application of GAN in the medical domain. [17] evaluated the performance of DCGAN to generate a large number of synthetic images to improve the performance of the classifier. A comparison of Affine and GAN based data augmentation has been done depicting better performance of GAN for data augmentation [18]. Similar work has been found for protein classification by comparing performance of traditional CNN architectures with GAN based augmentation techniques]and the results depict that the Inception V4 model yielded better results [19]. Zang et.al [20] compared performance of AlexNet, VGG16 and Inception V3 after using traditional affine transformation techniques to evaluate their performance.
None of the studies as known to the authors are found where the traditional CNN models are used after applying affine data transformation to the input and the results are compared with the DCGAN based model. Thus, there is a need for research in this area.

3. Methodology

3.1 Dataset Used

The data for the research came from the Kaggle repository [21] which includes images of Chest X-Rays of normal, and pneumonia patients. The dataset is imbalanced with training data having 1314 normal X ray images and 3875 X Rays of patients suffering from pneumonia. Similarly, the test dataset has 234 normal X-Ray images and 390 images of infected patients. The 5863 chest X-Ray images are classed as ‘Normal’ and ‘Pneumonia’.

3.2 Affine Image Transformation

Various Affine based transformations as presented in Table 1 are applied to the existing image dataset to increase the number of training images. The transformations are applied on the input image data set for the various CNN architectures like VGG16, Inception V3, InceptionResNetV2, DenseNet-169 and Densenet-202 to compare the performance of the models. The VGG16 architecture is a CNN architecture that won the 2014 Imagenet competition. VGG16 has 3x3 filter convolution layers with a stride one. It uses the padding and max pool layer of 2x2 filter stride instead of having a large number of hyper-parameters. It has two fully connected layers and softmax layer for output. It has 16 layers with different weights.

The Inception V3 was built taking input from the ImageNet database. It was trained for more than a million photos from this database. This network can identify images into 1000 different categories. These categories include keyboard, pencil, mouse and a variety of animals. As a result, this network has a library encompassing rich feature representations for a wide range of images. The input for the network is a 299X299 image and the output is a list of predicted class probabilities. Inception ResNetV2 is a hybrid of the Inception and ResNet model with more layers. DenseNet (Dense Convolutional Network) is a collection of dense blocks with different numbers of layers. It is a feed-forward network. A standard network with ‘P’ layers has ‘P’ connections. For the research, P(P+1)/2 direct conventions are used. Feature maps of all previous layers are used as inputs to each layer and successive layers. It is primarily used to deal with the vanishing gradient problem. It also improves feature propagation and promotes feature reuse. It also reduces the number of parameters to half. DenseNet-169 is a 169-layer deep convolutional neural network. DenseNet-201 was chosen for this research, as it has 201 layers and has low parameter count as compared to other models.

Table 1. Affine Transformations

Technique Used	Description
Flipping	This technique flips the image either horizontally, vertically or both. It is common to flip the horizontal axis rather than the vertical axis. It is one of the simplest augmentation techniques.
Space for Colour	It is used to change the image color channels. Simple color augmentation involves isolating a single color channel to R,G or B. RGB values can also be modified to enhance or reduce the brightness of the image. [8](Goodfellow, et al., 2014)
Cropping	It involves trimming a focal region of each picture. Editing a photo can be a helpful handling venture for picture information with blended stature and width aspects.
Rotation	It is used to rotate the image at an angle to either left or right on an axis lying between 1° and 359°.
Translation	It is used to move the image horizontally, vertically or both.

3.3 DCGAN based Data Augmentation

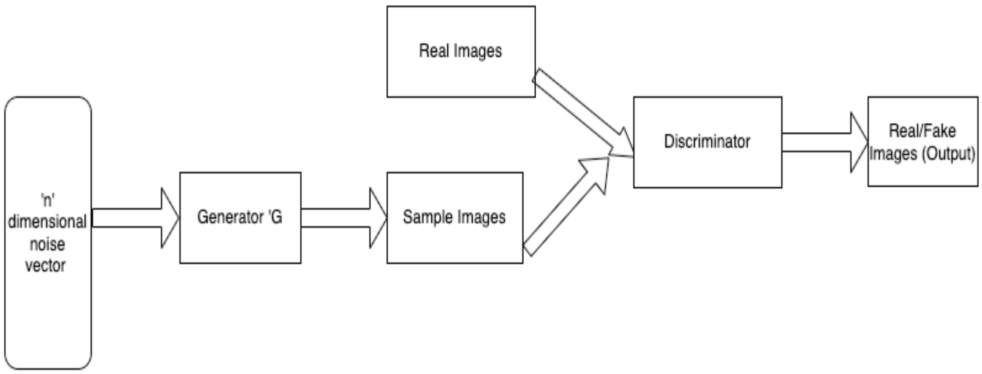

Figure 2: DCGAN Architecture

DCGAN combines the GAN with CNN [10] to obtain better and stable training results by increasing the generator and discriminator networks' internal complexity. Instead of Multilayer perceptron it uses CNN for generator and discriminator. It has convolutional layers in Generator and Convolutional-transpose layers in the discriminator. The architecture of DCGAN is presented in Fig 2. The input to the generator is a 'n' dimensional noise vector to generate sample images. It starts with a fully connected layer followed by transpose convolution, batch normalization where LeakyRelu is the activation function. At the end there is a convolution layer with filters and tanh activation function that generates images. On the other hand, the discriminator takes both the sample and real images and outputs the probability of the image being fake or real. To achieve this, strided-convolution is used with LeakyRelu as activation function and drop out included. At the end, a fully connected layer with sigmoid activation function is implemented to flatten the feature map and get the output as a real/ fake image.

4. Model Analysis

4.1 Data Preparation

The example size of normal and Pneumonia images was 885,1292. It is essential to standardize the images so that the pixel density for the image is distributed as it is critical for checking for Pneumonia as the Pneumonia looks for opacity but in comparison to standard x-rays, the images appear opaque. However, it's crucial to realize that chest x-rays don't always tell the whole story, and the visual outcome can be deceiving. Fig 3 shows the pixel density distribution in the dataset images. The blue color shows the range of pixel distribution whereas the red area shows the area where the pixel density is maximum.

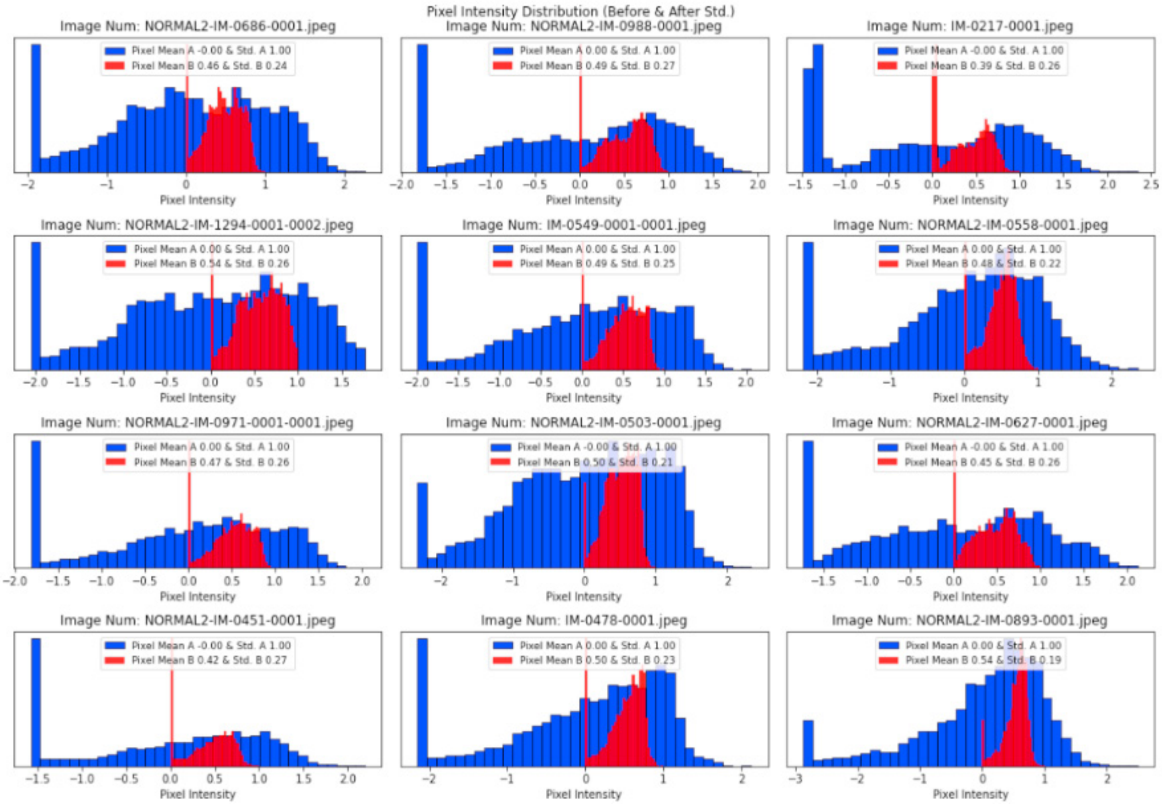

Fig. 3. Pixel distribution of images

The Affine based augmentation techniques were the input to various CNN architectures. The size distribution of the images can vary; hence we took 300x300 pixels for InceptionResNet V2, 224x 224 pixels for VGG16 models, 224x224 pixels for Densenet-169 and Densenet-201.

4.2 Selection of Hyper Parameters

Table 2 summarizes the various hyper parameters that were selected for the experiment.

Table 2. Hyper Parameters Selection

Hyper Parameters	Description
Weighted Binary Cross-Entropy	Weighted binary cross-entropy loss is used to deal with many false positives.
Sequential	A sequential model is used, which means the layers will be placed in a linear stack (sequence). The model layers are passed to this function Object () as parameters.
Dropout Rate	Neural networks trained on smaller datasets have a tendency to overfit, making them less reliable on new data. Theoretically, the optimal strategy to train a model is to try out as many distinct combinations of different parameter values as possible, then average the results to get a generalized result. However, training the model several times with varying combinations of these parameters would take a lot of time and computer resources. To deal with this, the dropout rate was created. Some units/nodes in a layer are 'dropped,' making their incoming and outgoing connections disappear (from input or hidden layers, but not from the output layer).
Optimizer	Adam optimizer is used. Categorical cross-entropy is used as it is a multi-class classification problem.
Epochs	During training, the number of epochs refers to the number of times the complete dataset is run

	through the neural network. There is no perfect number, and it is dependent on the data. In this example, the training starts with 20 epochs. As the number of epochs increases the training time also rises.
Activation	The ReLU activation function is used because of its computational simplicity (among other advantages). For negative inputs, this function returns zero; for positive information, it returns the value itself.

4.3 *Model Analysis using Affine Data Augmentation Techniques*

After applying Flipping, Space for colour, Cropping, Rotation and Translation for data augmentation, the input is utilized for classification of images using VGG16, Inception V3, InceptonResNetV2, DenseNet-201 and DenseNet-169 to evaluate the performance. The results after applying Affine transformations are presented in Table 3.

Table 3: Results after applying Affine Transformations

Technique	Accuracy	Precision	Recall	F1 Score	AUC
VGG16	82.05	89.36	85.61	87.44	90
Inception V3	90.38	91.5	94.4	93.41	96.7
Inception ResNetV2	90.2	92.38	97.5	94.8	97
DenseNet169	91.19	88.7	92.2	90.41	95.8
DenseNet 201	90.7	87.8	96.9	92.12	97

Table 3 depicts the accuracy, precision, recall, F1 score and AUC for the various CNN models after applying Affine transformation. Accuracy of DenseNet169 is highest (91.19%), but as it is the classification of the images in the medical domain and is related to the patient's well -being, Recall also needs to be considered to decide the performance of the technique. As observed in Table 3, Recall of Inception ResNetV2 is highest (97.5%) along with high AUC. Thus, it can be concluded that Inception ResNetV2 outperforms other CNN architectures after applying Affine transformations.

4.4 Model Analysis using DCGAN based Data Augmentation Techniques

Both the discriminator and the generator network are trained in the training process of DCGAN. The discriminator is trained to differentiate between the real and the fake images whereas the generator is trained to produce better fake images. The performance accuracy of DCGAN for various epochs ranging from 20 to 200 with a batch size of 128, LeakyRelu as the optimizer and sigmoid as an activation function is evaluated. The results depict that the algorithm obtains the real image classification accuracy of 98% in 200 epochs. Figure 4 depicts the generator and discriminator loss which is around 0.5 for both real and fake images during the training. This indicates that the generator has a 50% probability of fooling the discriminator.

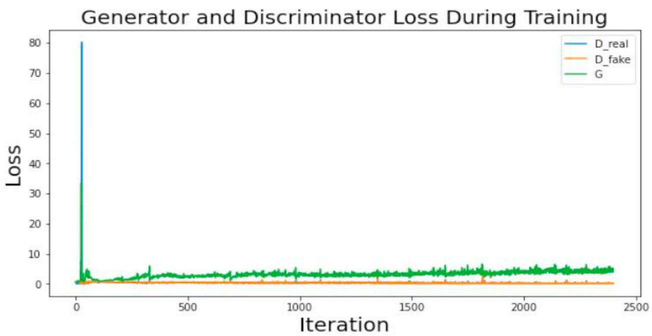

Figure 4: Generator and Discriminator Loss During Training

Table 4: Results after applying DCGAN transformation

Technique	Accuracy	Precision	Recall	F1 Score	AUC
DCGAN	98	99.14	100	99.2	99

As observed in Table 4, DCGAN is able to predict with 98% accuracy and 100% recall. Also, the AUC obtained is 99%. The results are better than that of the most suitable CNN model (Inception ResNetV2) with Affine methods applied. DCGAN is able to detect fake X-Ray images with 100% accuracy, which the traditional models cannot accomplish. Thus, DCGAN as a data augmentation technique should be preferred over traditional augmentation methods.

5. Conclusion

Traditional Affine and DCGAN based data augmentation techniques were compared in the context of Pneumonia detection in this study. Although affine transformations are one of the widely used techniques in practice, they are not as accurate as the GAN-based model. This is validated by the results obtained. The results show that DCGAN outperforms other CNN models in terms of accuracy and recall. Recall becomes the most important evaluation measure as we do not want to miss out on any person who has disease. This can be achieved if there are minimum false positives in the results. This will result in a high recall value. The study depicts that the recall for DCGAN is 100% and hence it is the best data augmentation method. The traditional affine based techniques can no way distinguish between the fake and real images, but DCGAN can do this with 100% accuracy. This becomes vital in the medical domain, where the image determines the fate of the patient in terms of the diagnosis and the further treatment plan.

The future work will involve comparing more GAN based techniques for evaluation of the best model.

References

[1] Ting, M. G., Karthikesalingam, D., D, A. K., Darzai, A. H., Aggarwal, R., Sounderajah, V. (2021). Diagnostic accuracy of deep learning in medical imaging: a. NPJ Digit Med 4, 4.

[2] Shorten, C., Khoshgoftaar, T.M. (2019), A survey on Image Data Augmentation for Deep Learning. *Journal of Big Data* 6, 60 https://doi.org/10.1186/s40537-019-0197-0.

[3] Rice, L., Wong, E., & Kolter, Z. (2020). Overfitting in adversarially robust deep learning. .Proceedings of the 37th International Conference on Machine Learning,119, pp. 8093–8104.

[4] Stutz, D., Hein, M., & Schiele, B. (2019). Disentangling adversarial robustness and generalization. IEEE Conference on Computer Vision and Pattern Recognition, 6976–6987.

[5] Mikołajczyk, A., Grochowski, M., (2019), Data augmentation for improving deep learning in image classification problem, 2018 International Interdisciplinary PhD Workshop (IIPhDW), 2018, pp. 117-122

[6] LeCun Y, Bottou L, Bengio Y, Haffner P.1998). Gradient-based learning applied to document recognition. Proc IEEE. 86(11):2278–324.

[7] Krizhevsky , A., Sutskever , I., Hinton , G. E.,(2012). ImageNet classification with deep convolutional neural networks. NIPS'12, Proceedings of the 25th International Conference on Neural Information Processing Systems,1:1097-1105.

[8] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde. F. D., Ozair, S., Courville, A., Bengio, Y. (2014). Generative Adversarial Networks. ArXiv, abs/1406.2661.

[9] Zoph., B.,Le, Q.V. (2017). Neural architecture search with reinforcement learning. International conference on learning representations.

[10] Radford, A.; Metz, L.; Chintala, S. (2016). Unsupervised Representation learning with Deep Convolutional GANs. In Proceedings of the International Conference on Learning Representations, 1–16.

[11] Zhu, J., Park, T., Isola, P., Efros, A.A., (2017). Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks. *IEEE International Conference on Computer Vision (ICCV)*, 2242-2251, doi: 10.1109/ICCV.2017.244.

[12] Collier, E., (2018) *et al*., "Progressively Growing Generative Adversarial Networks for High Resolution Semantic Segmentation of Satellite Images," *2018 IEEE International Conference on Data Mining Workshops (ICDMW)*, 2018, pp. 763-769, doi: 10.1109/ICDMW.2018.00115.

[13] Parez, L., Wang, J., (2017). The Effectiveness of Data Augmentation in Image Classification using Deep Learning. arXiv.

[14] Nalepa, J., Marcinkieqicz, M., Kawulok, M., (2019). Data Augmentation for Brain-Tumor Segmentation: A Review.Frontiers of Computer Neuroscience.

[15] Castro, E., Cardoso, J., & Pereira, J. (2018). Elastic deformations for data augmentation in breast cancer mass detection. IEEE EMBS International Conference on Biomedical Health Informatics (BHI), (pp. 230–234).

[16] Liu, X., Song, L., Liu, S., & Zhang, Y. (2021). A Review of Deep-Learning-Based Medical Image Segmentation Methods. Sustainability, 13(3).

[17] Venu, S., Ravula, S. (2021), Evaluation of Deep Convolutional Generative Adversarial Networks for Data Augmentation of Chest X-ray Images. Future Internet, 13, 8. https://doi.org/10.3390/fi13010008

[18] Adar, M., Diamant, I., Klang, E., Amitai, M., Goldberger, J., Greenspan, H., (2018), GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification, Neurocomputing, pp: 321-331.

[19] Verma R, Mehrotra R, Rane C, Tiwari R, Agariya AK. (2020). Synthetic image augmentation with generative adversarial network for enhanced performance in protein classification. Biomedical Engineering Letters. 10(3):443-452. doi: 10.1007/s13534-020-00162-9.

[20] Zhang, R, Zhou, B., Lu, C., Ma, M., (2022). The performance research of the data augmentation method for image Classification, Mathematical Problems in Engineering,

[21] Mooney SJ, Pejaver V. (2018). Big Data in Public Health: Terminology, Machine Learning, and Privacy. Annu Rev Public Health. 39, 95-112.