Medical Image Synthesis for Imbalanced Datasets Using CycleGan Under Label Noise

Md. Gazuruddin
Electrical and Computer Engineering
North South University
Dhaka, Bangladesh
gazur.uddin@northsouth. edu

HM Adnan Zami
Electrical and Computer Engineering
North South University
Dhaka, Bangladesh
hm.zami@northsouth.edu

Sajib Arafat Siddiqui
Electrical and Computer Engineering
North South University
Dhaka, Bangladesh
sajib.siddiqui@northsouth.edu

***Abstract*— Class imbalance issue in medical image datasets is a common issue that makes it difficult to train machine learning models for image classification with great performance. Recent developments in computer vision have proven to be effective in classification models, however, imbalanced datasets and noisy labels are a concern. To address this issue, we have looked into models to generate synthetic images to balance medical datasets. We used the ISIC-2016 dataset which is a binary classification dataset that suffers from class imbalance with benign being the dominant class. In our approach, we applied CycleGAN which has a Unet generator and PatchGAN discriminator and used a binary cross-entropy loss function. AlexNet was used for classification. Initially, the experiment was carried out on the base dataset after which we injected noise to create datasets with NAR and NCAR label noises. Afterwards, the classifier was used on these datasets and the results were compared with previous models. Classification results showed that our AlexNet model performed better than previous models such as VGG-GAP and VGG-GAP + Augment10x. Moreover, classification under label noise showed little changes in the accuracy which justifies the claim that neural networks are robust at handling label noise. Futhermore, we have shown that CycleGAN is an effective approach method for generating synthetic images from unpaired dataset. This approach can be applied to other datasets concerning different diseases to resolve the issue of imbalanced datasets.**

Keywords—CycleGAN, medical image synthesis, imbalanced datasets, label noise

I. INTRODUCTION

Medical image classification using machine learning models has been proven to be an effective method for detecting deadly disease in early stages to increase survival rate of patients. Nigel Hawkes reported that the Office for National Statistics in the UK stated prostate cancer, colorectal cancer and lung cancer has a survival rate of around 100%, 97.7% and 87.3% respectively if detected during the early stages [15]. Advancement in computer vision and deep learning technology has led to several machine learning models that can successfully classify images to detect such anomalies. However, due to the limitations in collecting quality data, much of the datasets available are imbalanced, resulting in poor prediction models for minority classes. To address this issue, conditional medical image synthesis using types of GAN models has been adopted.

GAN was first introduced by Ian Goodfellow et al. in 2014 which proved to be an effective unsupervised generative model for image synthesis. He introduced an adversarial network to a generative model to determine if generated images are real or fake and implemented a minimax loss function. However, several challenges such as vanishing gradient, mode collapse and convergence failure occurred. Researchers focused on finding new types of GAN models for medical image synthesis that overcame these issues.

This research focuses on a specific type of GAN called CycleGAN which is used to train a deep convolutional neural network for image-to-image translation. Addressing the class imbalance issues in the medical dataset, we will develop a model that will address the limitations of GAN in synthesizing data for minority classes in a dataset. Previous work has focused on classifying specific diseases from a niche of medical imaging fields such as dermoscopy, endoscopy, histology and others. Some of the popular datasets available for this are ISIC Challenge Datasets, MoNuSAC dataset, Kvasir-Capsule Dataset, ImageCLEFtuberculosis and such. Our project specifically focuses on developing a CycleGAN to perform image-to-image translation of melanoma images to balance the ISIC 2016 dataset. We have added label noise to further investigate the properties of images that affect classification. Contributions from this research can help improve datasets which will ultimately assist in training better models since image classification performs well when trained with large, balanced datasets.

II. LITERATURE REVIEW

A. Class Imbalance Issue in Medical Datasets

Medical datasets are often not balanced in their class labels. This leads to poorly performing classification models because models are optimized based on overall accuracy rather than considering the relative distribution of each class [1]. Rahman et al. addressed this issue in 2009 and aimed to overcome this by oversampling and under-sampling techniques. They implemented SMOT oversampling and cluster-based under-sampling on two cardiovascular datasets from Hull and Dundee. Five datasets were created using under-sampling by clustering and random under-sampling and one dataset was created using SMOTE Over-sampling. Classification was done using FURIA and decision tree the performance of which was evaluated by determining the

accuracy, sensitivity, specificity rate and positive predictive value. SMOTE showed the best classification outcome for both decision tree and FURIA. However, this paper mentions that there is a need for good training samples that display data properties consistent with class labels assigned to them [1].

Medical image analysis problems can be classified into three types: disease or abnormality detection, region of interest segmentation and disease classification. To balance and improve the quality of image datasets, adversarial learning has become popular. This has been used to balance datasets by synthesizing images for lesion detection, Covid-19 detection, breast cancer segmentation, retinal fundus multi-disease and such. Since Ian Goodfellow et al. introduced GAN, numerous researches focused on improving its performance for image synthesis.

B. Image Synthesis using GAN models

Considering medical images, CNN is the preferred type of AI network which relies on a large training dataset for image recognition and processing. This is implemented using a GAN. General structure of a GAN includes a generator *G* that takes a noise vector ***z*** as input and output is a synthetic sample. The discriminator then takes the synthetic input *G(z)* and true sample *x* to predict whether they are real or fake. The purpose is to generate fake images that fool the discriminator [5].

He Huang et al. summarized the taxonomy of methods used in image synthesis using GANs. There are three main approaches for GAN image synthesis which are direct methods, iterative methods and hierarchical methods. For image-to-image translation, they discuss supervised and unsupervised techniques. Among these, CycleGAN proved to be an unsupervised model that demonstrated a potential for performing high-quality image-to-image translation using unpaired data [5].

CycleGAN has a pair of generators that transform images from domain *A* to *B*, G_{AB}, and *B* to *A*, G_{BA}. There are also 2 discriminators D_A and D_b, and has an adversarial loss function for G_{AB} and D_B as defined:

$$L_{GAN}(G_{AB}, D_B) = E_b \sim_{PB(b)} [logD_B(b)] + E_a \sim_{PA(a)} \left[1 - log\left(D_B(G_{AB}(a))\right)\right]$$

Moreover, CycleGAN uses techniques to optimize the training process applying log loss for stabilization and uses SimGAN's strategy to reduce model oscillation [5].

C. CycleGAN for Medical Image Synthesis

Series of researchers implemented CycleGAN and modified it according to their requirements to perform medical image synthesis. Karim Armanious et al. introduced Cycle-MedGAN by expanding CycleGAN with two new non-adversarial loss functions similar to the ones introduced in MedGAN. Qualitative and quantitative comparison of the results showed that this method enhanced translated outputs for the tasks of PET to CT translation and MR motion correction [2].

Zhiwei Qin et al. performed GAN-based image on ISIC 2018 dataset by using a modified style-based GAN's architecture and compared it with GAN, DCGAN, StyleGAN and SL-StyleGAN. StyleGAN has the highest IS score while SL-StyleGAN has highest FID, Precision and Recall. Classification results showed that sensitivity and precision for CNN was highest for Melanocytic nevi classification [3]. Similarly, Hasib Zunair et al. used CycleGAN and VGG-GAP as well as their own MelNet classifiers for melanoma detection. They were able to overcome low inter-class variation in skin lesion dataset with an AUC of 81.18%, sensitivity of 91.76% and FN of 22 [18].

Mina Rezaei et al. worked on conditional GAN for unbalanced medical image semantic segmentation which addressed pixel level classification of healthy and unhealthy regions for LiTS-2017, MDA231, PhC-HeLa and BraTS-201 datasets [7]. Similarly, Liyan Sun et al. proposed ANT-GAN to perform image synthesis for lesion detection which they tested on BratS dataset and LiTS dataset [8]. All of this research work focused on a particular type of disease and developed a modified GAN for each case. Our ambition is to create a CycleGAN model that succeeds in performing image-to-image translations despite the presence of label noise in the dataset.

D. Classification in the Presence of Label Noise

Label noise refers to a form of class noise where the input data is incorrectly classified. When a model is trained on a dataset with label noise, it reduces the accuracy of the prediction and might increase the complexity of the model and the number of required training samples [19]. The two types of noise we looked at in this project are Noisy at Random (NAR) and Noisy Completely at Random (NCAR). NAR is a form of noise that is uniformly distributed where every class has equal probability of being corrupted. NCAR, on the other hand, is not uniformly distributed where targeted corruption is not equally likely.

III. METHODOLOGY

For our purpose, we have introduced these two types of noise in our dataset and compared the classification results of the noisy dataset with the base dataset. The dataset we used for this project is ISIC 2016. It has a total of We used a CycleGAN to generate images for the lesser class, i.e malignant, using benign images.

In addition to generating these images, we introduced the two types of label noises, NAR and NCAR to observe the effect of these on classification. The classifier we used for this project is AlexNet. AlexNet was introduced by Alex Krizhevsky et al. in 2012 when it won the Imagenet Large-Scale Visual Recognition challenge. It successfully classified 1.2 million high resolution images and won the top 5 error rate of 15.3% [20]. This classification model and the introduction of NCAR and NAR noise on the ISIC-2016 dataset is a novel approach that we have taken for this project.

A. Generating Synthetic Images Using CycleGAN

CycleGAN is an unsupervised training model for image-to-image translation. The architecture of CycleGAN is an extension of the Pix2Pix model but the primary advantage of CycleGAN over Pix2Pix is that it can be implemented on unpaired image datasets. Additionally, it allows us to convert images in the reverse direction, from target to source for example.

The architecture of the CycleGAN has 2 generators and 2 discriminators. We implemented Unet structure for our

generator which is known to be the most powerful network for medical image segmentation [22]. It is a U-shaped network that has 4 encoders and 4 decoders connected by a bridge. The generator loss that we used for our model was 2.0311.

For the discriminator, we used PatchGAN structure which focuses on penalizing structure at the scale of local images. This type of network is effective in finding high frequency details by focusing only on the patches of the image [5]. In our case, the discriminator loss was set to 0.29636.

B. Introducing NCAR and NAR Noise

In our approach, we introduced label noise in the form of NCAR and NAR to observe its effect on our model's prediction. Noise at Random, NAR, generally assumes that the label noise occurs randomly and it is independent of sample content. Also, the distribution of the noisy labels is not uniform which means some classes might be more prone to label noise than others. Considering dataset as $D = \{(\{(x0, y\hat{}0, y0), ..., (xn, y\hat{}n, yn)\},$ where each sample is x_i,true label is y_i, actual label as $y\hat{}i$ and classier as $f(x_i)$ [21].

$$P(\hat{y} = c \mid x) = P(\hat{y} = c) = \sum_{\acute{c}\epsilon C} P(\hat{y} = c \mid y = \acute{c}) P(y = \acute{c})$$

Another type of noise we implemented was NCAR. In this type of noise model, the probability of error is the same for all classes. We can denote the error probability with E and say that the probability of finding a label $y\hat{}$ of class c in the set of all classes C is [23]

$$P(\hat{y} = c \mid x) = P(\hat{y} = c)$$
$$= P(E = 0)P(y = c) + P(E = 1)P(y6 = c)$$

C. Classification Using AlexNet

Among the classifiers out there, AlexNet has been successful in classifying the large ImageNet dataset with great accuracy. The AlexNet architecture has a total of eight layers among which five are convolutional and three are fully connected layers. ReLu activation function is used for all layers and the final output layer uses Softmax activation function. Moreover, to prevent overfitting, there are three dropout layers used in this architecture.

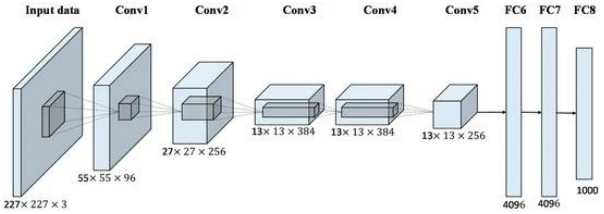


Figure 1: AlexNet Architecture that shows the convolutional and fully connected layers

Using AlexNet, we performed the classification for our datasets. We used this to classify the base dataset as well as the dataset with NAR and NCAR noises. Then we compared the results to identify how label noises affect classification of melanoma.

D. Loss Functions and Optimizer

As we are performing binary classification, we used the Binary Cross Entropy classification. This loss function compares each predicted output probability with actual class output that can be 0 or 1. Based on the distance from the expected value, the score for the penalty is calculated. The formula for binary cross-entropy loss can be denoted by

$$Loss = -\frac{1}{output\ size} \sum_{i=1}^{output\ size} y_i \times log\ \hat{y}_i + (1 - y_i) \times log(1 - \hat{y}_i)$$

For the optimizer, we used two optimizers which were Stochastic Gradient Descent (SGD) and Adagrad optimizer. SGD is an iterative method that is popular for improving machine learning models, especially artificial neural networks. Moreover, we used the Adagrad optimizer which is also a gradient based optimizer. An advantage of Adagrad is that convergence is faster. We used Adagrad optimizer for our base data and data with NCAR noise added. For the data with NAR noise, we use the SGD optimizer.

IV. EXPERIMENTS

In this section, comprehensive experiments are driven to evaluate the performance of three-stage-approach on a dataset for skin lesion analysis.

Dataset: Experiment carried out based on International Skin Image Collaboration (ISIC-2016) dataset which is publicly available dermatology image analysis benchmark challenge for skin lesion analysis on melanoma detection [21]. The dataset has a mixed image of both benign and malignant skin lesions. Which is divided into two-part training and testing. Training and Testing image are 900 and 379 images respectively. These images have differences in texture to both background and foreground with poor contrast, here noteworthy mention that in training images have 727 benign and 173 malignant cases, approximately ratio is 1:4. [fig: -2].

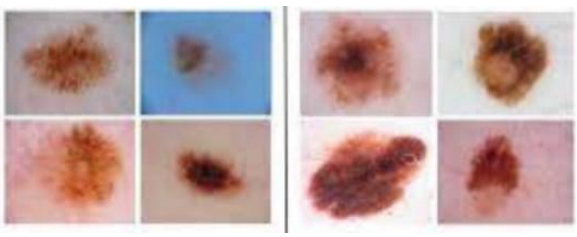


Figure 2: Sample benign and malignant image from ISIC-2016 dataset Looe that Left side is malignant and right side is benign.

This figure shows that both benign and malignant similarity is high and one more point note that benign samples are not visually different [18]. The histogram of training data shows that there is a class imbalance problem where the majority class (“benign”) is much larger than minority class (“malignant"). Not only class imbalance on training set but also in testing set as follows benign have 304 images and malignant have 75.

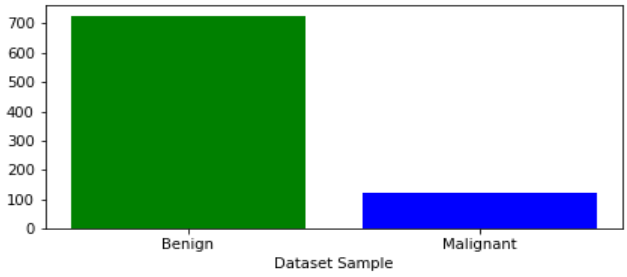


Figure 3: Histogram shows the class imbalance between two classes (malignant and benign).

Training Process: Since we are working with binary classification problems with imbalance data, we used AlexNEt model with binary cross entropy as loss function. It actually calculates the score which penalizes the probabilities based on distance from predicted value.

We train CycleGAN with a Unet generator and PatchGAN discriminator. Used optimization as Adam with 100 epochs, learning rate is 0.001, batch-size obviously 1 and set regularization parameter as ƛ= 8[fig-4]. We used only one classification model named AlexNet. In AlexNet Implementation we used an Adagrad optimizer with learning rate 0.001 and mini-batch 16. To prevent overfitting, we used regularization parameter as Dropout =0.4. After tuning some optimizer like RmsProp, adam and so on we chose Adagrad because its learning rate changes each iteration and is very effective.

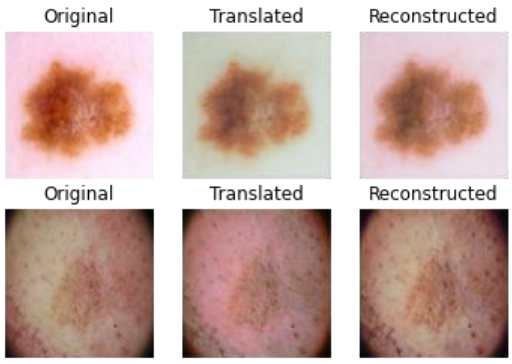


Figure-4: Original, Translated and Reconstructed image after completing 72 epochs with CycleGAN.

After classification with Alexnet, we classified again same dataset with artificially injected noise. In NAR noise, we create 10 and 20 percent noise from the dataset and replace it with real data then apply the classification model to observe accuracy with noise. As the same procedure we follow for NCAR noise. Same amount of noise injected but used difference Optimizer and learning rate for classification to check if the model is doing well with other parameters or not. Stochastic Gradient Descent (SGD) optimizer and learning rate 0.0001 and 0.01 factor used to reduce the learning rate once loss freezes.

V. Results and Discussions

To classify the result of our classification we used several metrics to make comparison between results. We used Receiver Operating Curve (ROC), Area under the ROC curve, Sensitivity. Sensitivity is basically the number of positive instances which are correctly predicted, such as

$$Sensitivity = \frac{TP}{TP + FN}$$

Where TP, FN, TN are denoted as true positive and false negative, true negative respectively. TN classifies the number of correctly predicted benign lesions and TP is the correctly predicted number of malignant lesions. Sensitivity also known as Recall. Its measure is how often a classifier misses positive prediction and most useful metrics to evaluate the result of the medical image classification.

Another important metric to sum up the information contains the ROC curve which differentiates True Positive rate (TPR) between True Negative rate (TNR) = FP+FP/TN. Basically, the AUC curve measures the model based on area. If a larger area is covered that means the model is more than the small area cover model.

In this letter we applied three models, AlexNet with Synthesis data, AlexNet with NAR label noise data and also NCAR label noise data. In the beginning we applied AlexNet with original data which has imbalance classes but output was bad as expected. Conducting CycleGAN on the dataset we make more malignant images, actually in original dataset we have 727 benign images and 173, so we applied CycleGAN on benign images and produced 727 malignant images. Then we used AlexNet with Noise NAR and NCAR and the output as follows:

Method	AUC (%)	Recall (%)	FN
VGG-GAP	79.08	84.46	55
VGG-GAP +Augment-10x	79.56	86.09	51
Alexnet-NAR (ours)	81.03	87.39	49
AlexNet_NCAR (ours)	82.32	89.74	47

AlexNet (ours)	87.427	93.26	24

Table 1: Classification result of AlexNet ours model and previous Model.

First, we develop the AlexNet Model with Adagrad Optimizer (learning rate=0.001). In training time, we notice that loss decreases gradually and after 100 epochs loss becomes 0.1798 and training accuracy is 0.9816 which is pretty good at Computer vision, but when we check the testing accuracy which was 0.84. We failed because our model is overfitting. That's why training accuracy is higher than testing accuracy. Overfitting happens because of high variance and high variance reduction by collecting more data or using regularization. Applying regularization to reduce variance is tough for this model because AlexNet already has built-in Dropout=0.4. Regularization and CycleGAN are already used so we have an augmentation method left to increase the dataset size and reduce variance and we implement it. Then after several iterations of augmentation then we applied the model again and got AUC 87.39[fig-5].

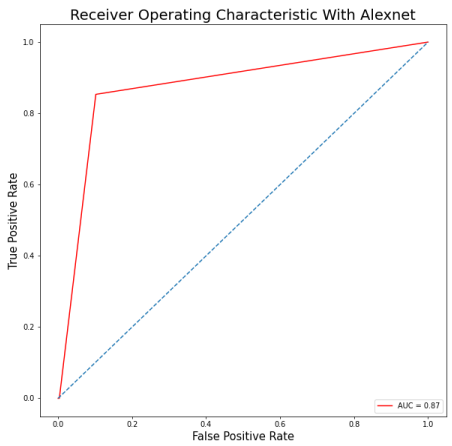


Figure 5: ROC curve with AlexNet model along with the corresponding AUC values

In our ALexNet-NAR model we create 0 to 20% label noise and applied each even number (2,4,6 until 20) noise to the dataset and check classification result. We already know that Neural Networks work robustly at label noise. For this purpose, >10 label noise doesn't affect the accuracy but when we apply <10 then accuracy decreases gradually. Finally, when we injected 20% label noise to the dataset then the AUC value 81.03 fig [6].

Previously, work conducted with the same dataset with CaycleGan and then they introduced VGG-GAP and VGG-GAP with 10x Augmentation, Optimizer Adadelta on learning rate 0.0001 and also highly mention the focal loss [18]. We change the whole procedure to choose a model to classify the image. Optimizer Adadelta is an extension of Adagrad. Where adagrad changes the learning rate in each iteration on the other hand adadelta decreases the learning rate aggressively. Let's check the loss function, they used customized focal loss but we use binary cross entropy. Which are equivalent to each other but note that the point is binary cross entropy works very well particularly in the medical sector. In the discussion on choosing a model we take AlexNet and they work with VGG-GAP. VGG defines here the VGG-16 model and in the second last layer they take Global Average Pooling(GAP) instead of a fully connected layer. AlexNet introduced before VGG-16 and compared to AlexNet VGG-16 have less parameters.VGG-16 actually introduces a receptive field with (3x3 stride of 1) instead of AlexNet's (11x11, stride of 4). If the dataset has high bias then using more hidden units or training longer could be an advantage. For the sake of several reason we chose our model and after tuning some parameters we get better results for our synthesis dataset.

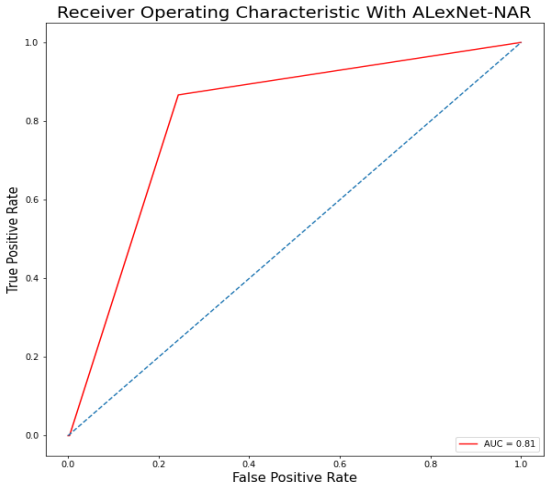


Figure 6: ROC curve with AlexNet_NAR model along with the corresponding AUC values

VI. Conclusion

In this project, we tried to address the class imbalance issue for medical images. Computer vision research has proposed several approaches to address this issue among which conditional images synthesis is one. The introduction of GAN has led to the progress of image synthesis with great quality. One of the popular models of GAN is CycleGAN which is an unsupervised model for performing unpaired image-to-image translation. Our goal was to observe the performance of CycleGAN on the base dataset and then add label noises to see its impact on performance.

We used the ISIC 2016 dataset which is a binary classification dataset with low inter-class variability. Among the two classes, benign and malignant, there are 900 benign and 379 malignant images which were used with 70% for training and 30% for testing purposes. In our experiment, we applied CycleGAN where the generator used the Unet

architecture and the discriminator used the PatchGAN architecture with a loss of 2.0311 for the generator and 0.29636 for the discriminator.

Initially, we trained the CycleGAN and performed classification on the resulting data using AlexNet classifier and binary cross-entropy loss function. Then to identify the impact of label noise, we introduced NAR and NCAR label noises and performed the classifications on them again to observe the impact. We compared our results with previous models such as VGG-GAP, VGG-GAP + Augment10x and found out that our AlexNet classifier without NAR and NCAR had the highest AUC and recall.

This project further justifies the use of CycleGAN in generating good quality synthetic medical images. By introducing label noises, we observed that it has little effect on the AUC and recall score which also ensures that neural networks are quite robust in tackling label noises. For future studies, this model can be applied to other datasets which are prone to label noise or class imbalance to test and see its performance on classifying diseases other than melanoma.

References

[1] Mostafizur Rahman & Darryl N. Davis (2013). Addressing the Class Imbalance Problem in Medical Datasets. [Online]. Available: https://www.researchgate.net/publication/239608168_Addressing_the_Class_Imbalance_Problem_in_Medical_Datasets

[2] Karim Armanious, Chenming Jiang, Sherif Abdulatif, Thomas Kustner, Sergios Gatidis, Bin Yang (2019). Unsupervised Medical Image Translation Using Cycle-MedGAN. [Online]. Available: https://arxiv.org/pdf/1903.03374.pdf

[3] Zhiwei Qin, Zhao Liu, Ping Zhu, & Yongbo Xue (2020). A GAN-based image synthesis method for skin lesion classification. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S0169260720302418

[4] Ding, Saisai & Zheng, Jian & Liu, Zhaobang & Zheng, Yanyan & Chen, Yanmei & Xu, Xiaomin & Lu, Jia & Xie, Jing. (2021). High-resolution dermoscopy image synthesis with conditional generative adversarial networks. Biomedical Signal Processing and Control. 64. 102224. 10.1016/j.bspc.2020.102224.

[5] He Huang, Philip S. Yu, & Changhu Wang (2018). An Introduction to Image Synthesis with Generative Adversarial Nets. [Online]. Available: https://arxiv.org/abs/1803.04469

[6] M. Rezaei, H. Yang, K. Harmuth and C. Meinel, "Conditional Generative Adversarial Refinement Networks for Unbalanced Medical Image Semantic Segmentation," 2019 IEEE Winter Conference on Applications of Computer Vision (WACV), 2019, pp. 1836-1845, doi: 10.1109/WACV.2019.00200.

[7] Augustus Odena, Christopher Olah, and Jonathon Shlens. 2017. Conditional image synthesis with auxiliary classifier GANs. In Proceedings of the 34th International Conference on Machine Learning - Volume 70 (ICML'17). JMLR.org, 2642–2651.

[8] L. Sun, J. Wang, Y. Huang, X. Ding, H. Greenspan and J. Paisley, "An Adversarial Learning Approach to Medical Image Synthesis for Lesion Detection," in IEEE Journal of Biomedical and Health Informatics, vol. 24, no. 8, pp. 2303-2314, Aug. 2020, doi: 10.1109/JBHI.2020.2964016

[9] G. SANCHEZ, V. GUIS, R. MARXER and F. BOUCHARA, "Deep Learning Classification with Noisy Labels," 2020 IEEE International Conference on Multimedia & Expo Workshops (ICMEW), 2020, pp. 1-6, doi: 10.1109/ICMEW46912.2020.9105992.

[10] J. Deng, W. Dong, R. Socher, L. Li, Kai Li and Li Fei-Fei, "ImageNet: A large-scale hierarchical image database," 2009 IEEE Conference on Computer Vision and Pattern Recognition, 2009, pp. 248-255, doi: 10.1109/CVPR.2009.5206848.

[11] Skandarani, Youssef & Jodoin, Pierre-Marc & Lalande, Alain. (2021). GANs for Medical Image Synthesis: An Empirical Study.

[13] N. C. F. Codella et al., "Skin lesion analysis toward melanoma detection: A challenge at the 2017 International symposium on biomedical imaging (ISBI), hosted by the international skin imaging collaboration (ISIC)," 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018), 2018, pp. 168-172, doi: 10.1109/ISBI.2018.8363547.

[14] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. 2017. ImageNet classification with deep convolutional neural networks. Commun. ACM 60, 6 (June 2017), 84–90.

[15] Hawkes N. Cancer survival data emphasise importance of early diagnosis BMJ 2019; 364 :l408 doi:10.1136/bmj.l408

[16] Banik, Debapriya & Bhattacharjee, Debotosh. (2021). Mitigating Data Imbalance Issues in Medical Image Analysis. 10.4018/978-1-7998-7371-6.ch004.

[18] Zunair, Hasib & Hamza, A.. (2020). Melanoma detection using adversarial training and deep transfer learning. Physics in Medicine & Biology. 65. 10.1088/1361-6560/ab86d3.

[19] Frénay, Benoît & Verleysen, Michel. (2014). Classification in the Presence of Label Noise: A Survey. Neural Networks and Learning Systems, IEEE Transactions on. 25. 845-869. 10.1109/TNNLS.2013.2292894.

[20] Yan, Wenjun & Wang, Yuanyuan & Gu, Shengjia & Huang, Lu & Yan, Fuhua & Xia, Liming & Tao, Qian. (2019). The Domain Shift Problem of Medical Image Segmentation and Vendor-Adaptation by Unet-GAN.

[21] Codella, Noel & Gutman, David & Celebi, M. Emre & Helba, Brian & Marchetti, Michael & Dusza, Stephen & Kalloo, Aadi & Liopyris, Konstantinos & Mishra, Nabin Kumar & Kittler, Harald & Halpern, Allan. (2017). Skin Lesion Analysis Toward Melanoma Detection: A Challenge at the 2017 International Symposium on Biomedical Imaging (ISBI), Hosted by the International Skin Imaging Collaboration (ISIC).

[22] Guillaume Sanchez, Vincente Guis, Ricard Marxer, Frederic Bouchara. Deep learning classification with noisy labels. ICME Workshop, Jul 2020, Londres, United Kingdom. ffhal-02552375f