# Medical Image Dataset Rebalancing via Conditional Deep Convolutional Generative Adversarial Networks (cDCGANs)

Adam Sun and Kevin Li
{adsun, kevinli8}@stanford.edu

## Background

Imbalanced datasets are common in the medical imaging context where normal data outnumbers disease data.
Deep learning models trained on imbalanced data tend to be biased towards majority class, leading to more false negatives, which is particularly problematic since this means medical conditions may be left undetected.
Classic approaches include:

- Oversampling
- Undersampling
- Class weights
- SMOTE methods
- Data augmentation

To avoid overfitting that comes with classic oversampling methods we propose using a GAN to generate synthetic examples from random noise, adding to the diversity of minority class.

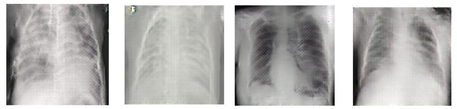
## Problem

Our goal is to rebalance image datasets while reducing bias and improving overall model performance.

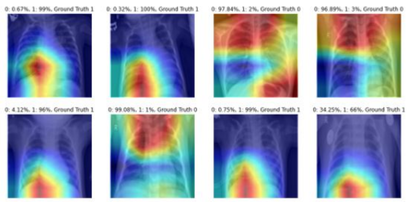
Our GAN uses random noise to generate new synthetic pneumonia chest x-rays that are added to the dataset, and we assess the different methods by training a ResNet-18 model that takes the chest x-rays as input and outputs a binary classification of healthy or pneumonia.

Due to our imbalanced dataset, we evaluate performance based not on accuracy, but precision, recall/sensitivity, specificity, and F1 score.

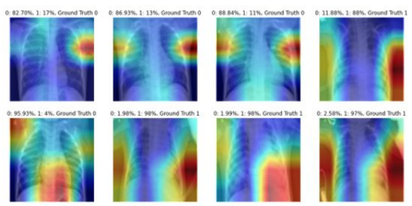
## Visualizations



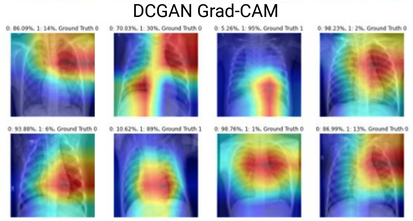
DCGAN Pneumonia cDCGAN Healthy vs Pneumonia



Baseline Grad-CAM



DCGAN Grad-CAM



cDCGAN Grad-CAM, $\alpha$ = 0.5

## Dataset

Dataset was taken from Kermany et al. consisting of pediatric chest X-rays from Guangzhou. We supplemented with healthy chest X-rays from public NIH dataset ChestX-Ray8. We randomly removed pneumonia images to created imbalanced situation: 4385 normal, 1460 pneumonia. The test set is more balanced to be an unbiased evaluation of model performance.

Healthy

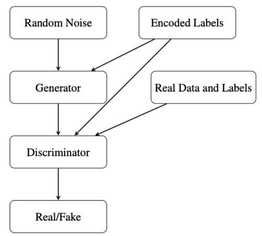
Pneumonia

## Methods

We use a Generative Adversarial Network, which is used to learn a mapping from random noise to the data space of a distribution. A generator network, which tries to generate a realistic data point, plays a minimax game against a discriminator network which seeks to discriminate between real and fake data. The objective function from the original paper:

$$\min_G \max_D V(D,G) = \mathbb{E}_{\boldsymbol{x}\sim p_{\text{data}}(\boldsymbol{x})}[\log D(\boldsymbol{x})] + \mathbb{E}_{\boldsymbol{z}\sim p_{\boldsymbol{z}}(\boldsymbol{z})}[\log(1 - D(G(\boldsymbol{z})))]$$

We use a deep convolutional GAN (DCGAN), which has strided convolution, batchnorm, and ReLU to improve GAN stability and image comprehensibility. We train it on the minority class of pneumonia images and then use the generator to generate realistic examples from the minority class distribution and oversample the minority class to rebalance an imbalanced dataset.




CGAN architecture

Additionally, we explore conditional DCGAN as a way to utilize the entire dataset. By conditioning on labels, we can then drive the generated images towards the pneumonia class. We also introduce a hyperparameter $\alpha$ that represents the probability of a randomly generated healthy label.
We compare ResNet-18 performance on the rebalanced dataset with its performance on datasets rebalanced by more classic methods.

## Experiments

| | Precision | Recall/ Sensitivity | Specificity | F1 |
|---|---|---|---|---|
| Baseline | 0.730 | 0.842 | 0.850 | 0.782 |
| Class Weights | 0.859 | 0.855 | 0.915 | 0.857 |
| Under-sam-pling | 0.709 | 0.865 | 0.843 | 0.779 |
| Over-sam-pling | 0.863 | 0.845 | 0.917 | 0.854 |
| DCGAN | 0.821 | **0.910** | 0.898 | 0.863 |
| cDCGAN $\alpha = 0.50$ | **0.893** | 0.857 | 0.934 | **0.874** |
| cDCGAN $\alpha = 0.25$ | 0.710 | 0.706 | **0.949** | 0.803 |
| cDCGAN $\alpha = 0.75$ | 0.821 | 0.901 | 0.898 | 0.859 |

## Analysis

We see that:

- DCGAN and the cDCGAN variants were all able to outperform the baseline in almost all metrics.
- In particular, both our DCGAN and cDCGAN ($\alpha$ = 0.5) approaches outperform class weights, oversampling, and undersampling (the most commonly used methods) in terms of F1 score.
- DCGAN and cDCGAN ($\alpha$ = 0.25) have the highest sensitivity and specificity, respectively. This uncovers the possibility of using these models in conjunction to generate a better dataset.
- We find that $\alpha$ can be used to tune sensitivity and specificity.
- Our Grad-CAMs uncover that the DCGAN ResNet is not learning relevant features inside the chest cavity, while the baseline and cDCGAN do a bit better in this respect.

## Conclusions and Future Work

In this work, we explored the usage of Deep Convolutional Generative Adversarial Networks (DCGANs) and conditional Deep Convolutional Generative Adversarial Networks (cDCGANs) for oversampling of the minority class in imbalanced medical image datasets.

Based on the results, we conclude that GAN-based approaches are a valid way to oversample an imbalanced training dataset, despite the Grad-CAMs indicating that the model may be learning irrelevant features.

In the future, we would like to tune the GAN models more carefully, and train models multiple times to compensate for the stochastic nature of model training, as well as curate a more diverse imbalanced dataset. We may offer our GAN-generated images to an actual experienced radiologist for review for usefulness, relevance, and realisticness, resulting in a better annotated dataset.