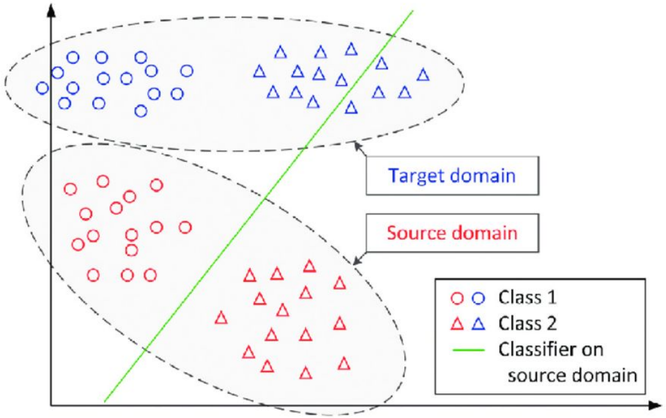
# Domain Adaptation in Medical Imaging: WGAN + CNN-Transformer

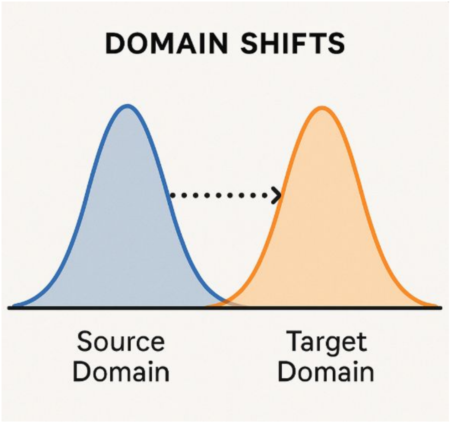
Saad Alrajhi, Drew Kulischak, Nick Russert, Francis Fernandez, Chao Jung Wu

# Motivation: Domain Shifts

- When models are trained and tested on their same domains or distributions, they are able to perform really well.
- However, when these models attempt to test on other domains or distributions, they perform poorly.
- The domain differences in medical scans could be based on modality, patient population differences, or region.



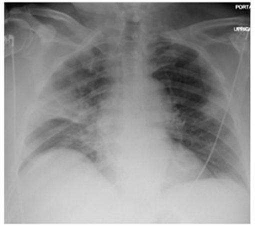

Credit

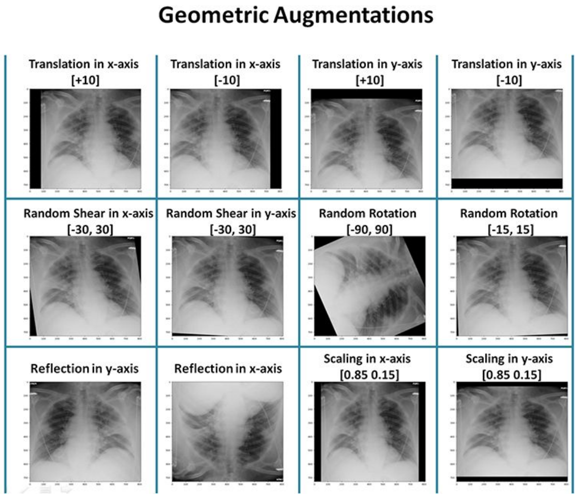



GPT-image

# Motivation: Data Scarcity

- In medical images, there is a data scarcity problem
- Because of the lack of available medical images, scientists apply different techniques to get the most out of existing medical data






| | **Egypt** | **Poland** |
|---|---:|---:|
| **Malignant** | 210 | 102 |
| **Benign** | 570 | 154 |
| **TOTAL** | 780 | 256 |

# Approach: Datasets

We define two distinct domains (distributions) that share the same modality of ultrasound medical images of tumors:

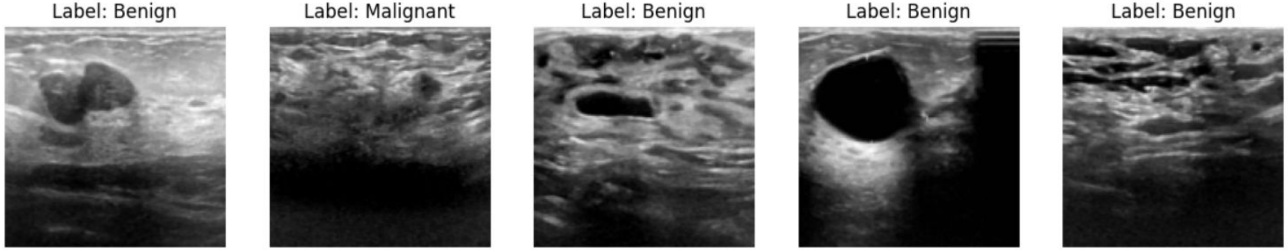
1. **Egypt**, a collection of ultrasound images of breast cancer from Egyptian medical centers (*our training dataset)*
2. **Poland,** a collection of ultrasound images of breast cancer from Polish hospitals (*our testing dataset)*

- The tasks is to classify whether the tumors are benign (1) or malignant (0)

# Approach: Datasets

Poland


Egypt



# Goals and Contributions

**Goals:**

1. Understand why some models yields higher accuracies in the general classification task.
2. Build a GAN model that can **generate realistic looking synthetic images**.
3. Understand how different **ratios of real and synthetic data help models adapt their accuracy** to other domains

**Contributions:**

1. A modified WGAN model **(BIG-WGAN)**
2. **CNN-Transformer** based on TransMed
3. Experiments within and across domains using **CNNs,ViT, and a CNN-Transformer**

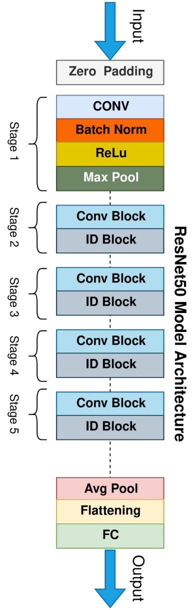
# Domains Differences

Training and Testing on different domains causes models to perform much worse the target domain. Here is an example using a Resnet:

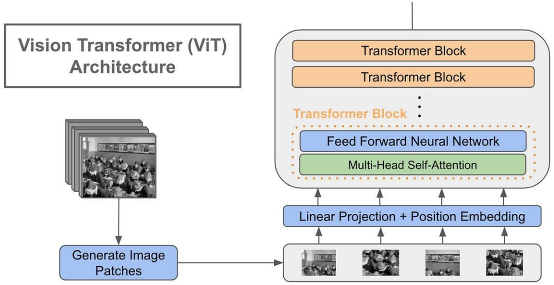
| | Tested on Souce | Tested on Target | Domain Gap |
|---|---|---|---|
| Trained on Egypt | 80.26% | 62.50% | **17.76%** |
| Trained on Poland | 68.06% | 49.01% | **19.07%** |

# Classification Models

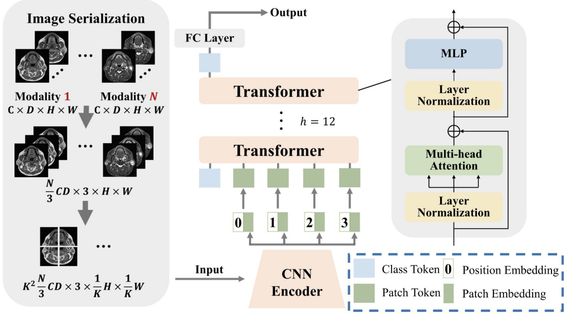
## ResNet50




## Vision Transformer




## CNN-Transformer

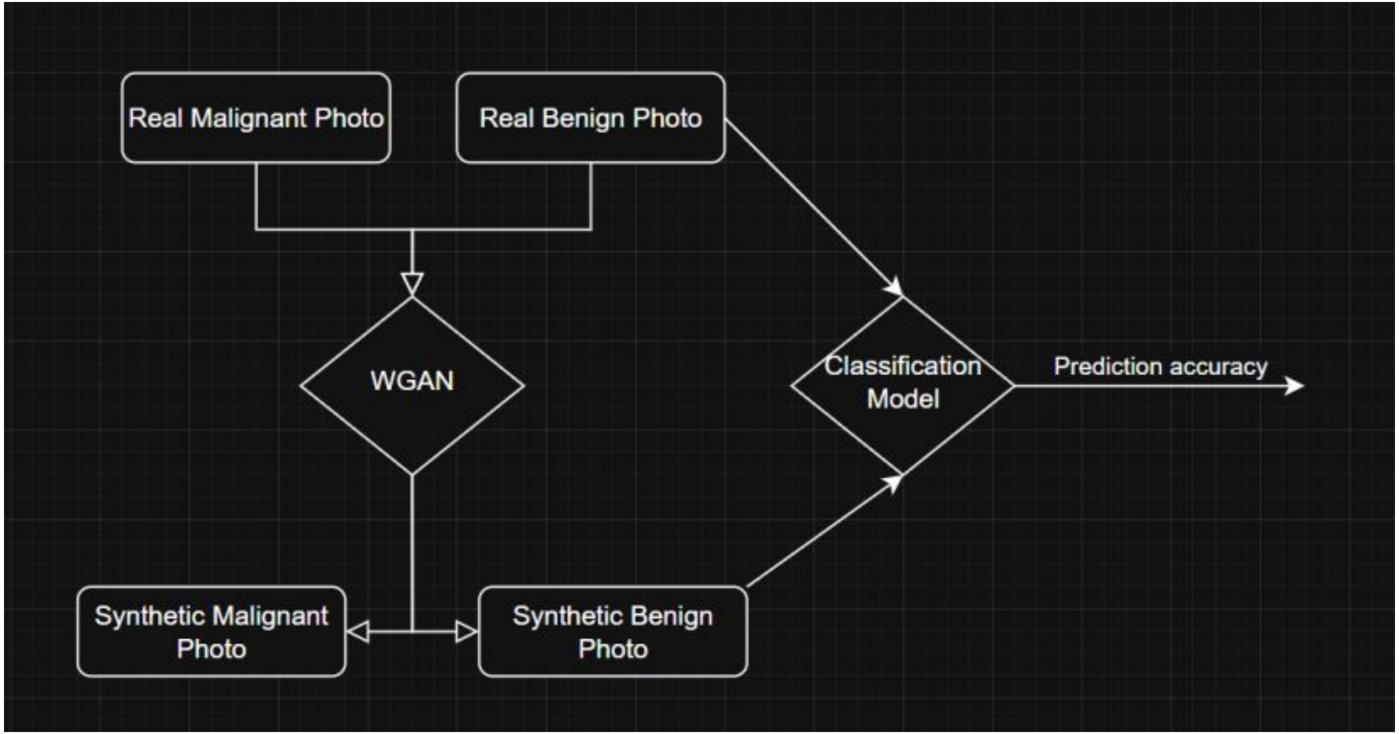


# Synthetic Images Generation

We utilize a WGAN, a GAN model with the following attributes:

- Provides smoother gradients
- A score metric (wasserstein distance) instead of a sigmoid percentage in the discriminator.
- Weight clipping, which prevents the gradients from exploding

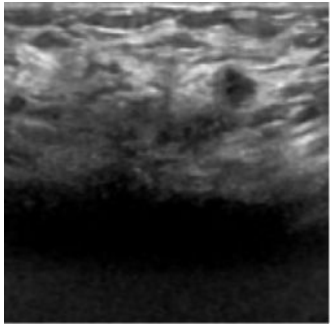
# Methodology



# GANs Weaknesses

- Its difficult to find a GAN model that is able to generate “good” fake images for smaller datasets, especially for this problem. We modify the WGAN and call it BIG-WGAN.

Real



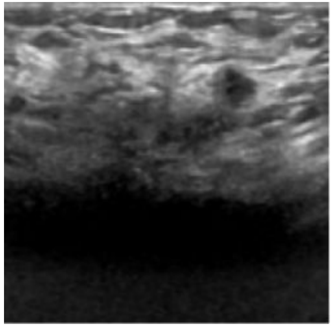
Fake

[GAN author]()

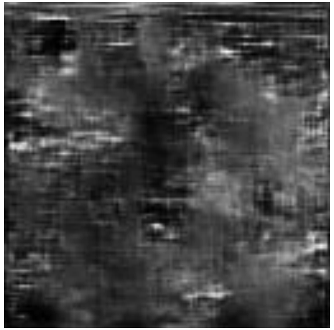
# Big WGAN

- We add more convolution layers to our Generator

Real



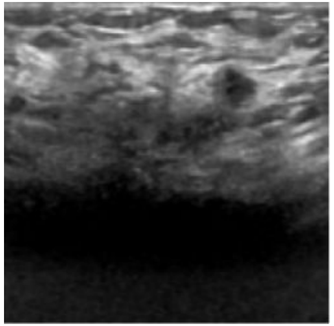
Fake



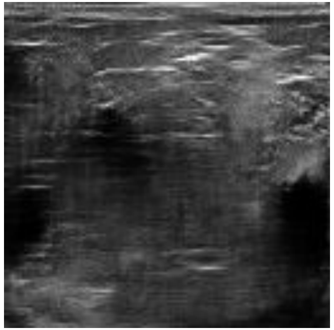
# Big WGAN

- More Convolutions in the discriminator

Real



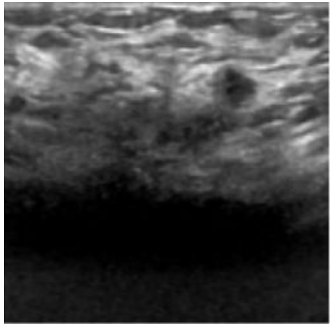
Fake



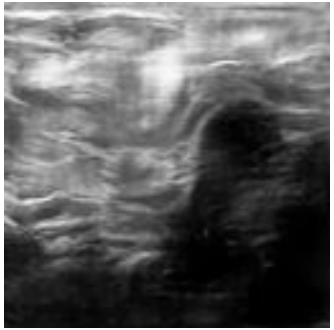
# Big WGAN

- Spectral norm added to the discriminator

Real

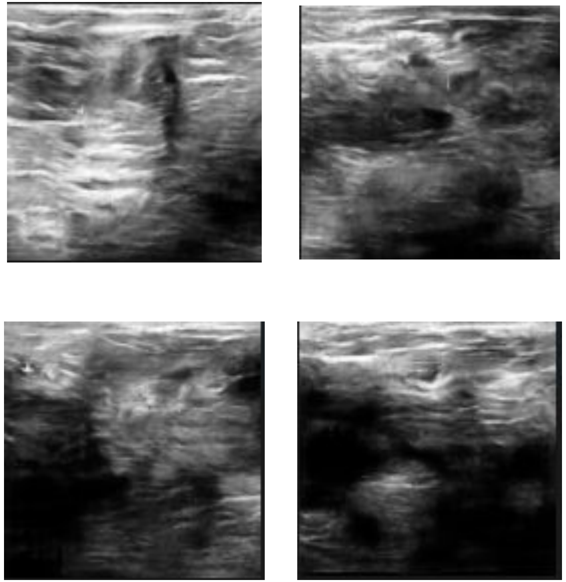


Fake

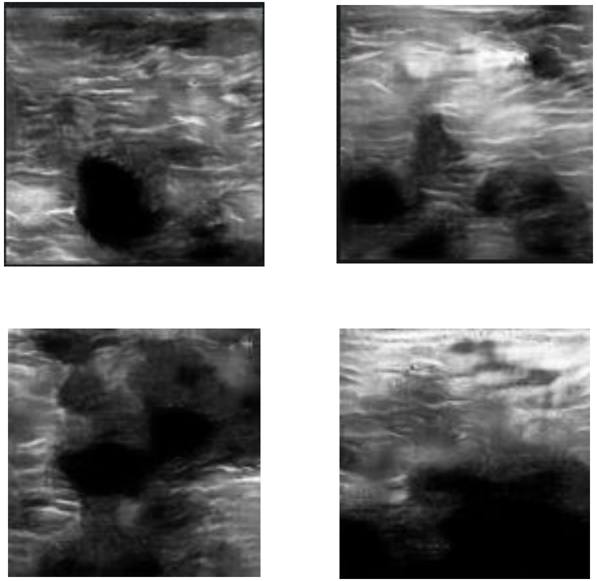


# Synthetic Data

Benign:



Malignant:



# Results: ResNet50

ResNet model trained on Egypt

| Train on | Real | Fake | Test on | Accuracy | Domain Gap |
|---|---|---|---|---|---|
| Real only | 100 | 0 | Egypt | 71.4% | N/A |
| Real only | 100 | 0 | Poland | 65% | 6.4% |
| **Real + Fake (4:1)** | **100** | **25** | **Poland** | **66%** | **5.4%** |
| Real + Fake (2:1) | 100 | 50 | Poland | 63% | 8.4% |
| Real + Fake (3:2) | 100 | 75 | Poland | 64% | 7.4% |
| **Real + Fake (1:1)** | **100** | **100** | **Poland** | **66%** | **5.4%** |

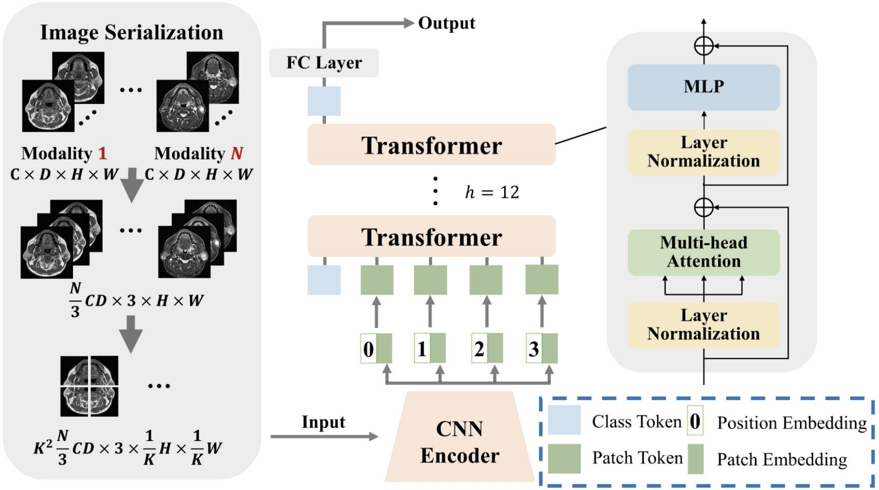
# Results: Vision Transformer

Vision Transformer model only trained on Egypt

| Train Domain | Real | Fake | Test Domain | Accuracy | Domain Gap |
|---|---|---|---|---|---|
| Real only | 100 | 0 | Egypt | 70% | N/A |
| Real only | 100 | 0 | Poland | 55% | 15% |
| Real + Fake (4:1) | 100 | 25 | Poland | 50% | 20% |
| **Real + Fake (2:1)** | **100** | **50** | **Poland** | **58%** | **12%** |
| Real + Fake (3:2) | 100 | 75 | Poland | 55% | 15% |
| Real + Fake (1:1) | 100 | 100 | Poland | 55% | 15% |

# CNN-Transformer

- The CNN takes 1x128x128 images and transforms them into a 256x32x32 patches
- Then the 32x32, or 1024 “patches” are fed into the transformer
- Great at discovering long-distance relationships for many small patches




Inspired by [TransMed]

# Results: CNN-Transformer

CNN-Transformer model trained on Egypt

| Train Domain | Real | Fake | Test Domain | Accuracy | Domain Gap |
|---|---|---|---|---|---|
| Real only | 100 | 0 | Egypt | 76.4% | N/A |
| Real only | 100 | 0 | Poland | 63.6% | 12.8% |
| Real + Fake (4:1) | 100 | 25 | Poland | 65% | 11.4% |
| Real + Fake (2:1) | 100 | 50 | Poland | 65% | 11.4% |
| **Real + Fake (3:2)** | **100** | **75** | **Poland** | **65.4%** | **11%** |
| Real + Fake (1:1) | 100 | 100 | Poland | 65% | 11.4% |

# Comparing The Models

Training on Egypt’s images and testing on Poland

| **Train Domain** | **Test Domain** | **ResNet50 Domain Gap** | **Vision Transformer Domain Gap** | **CNN-Transformer Domain Gap** |
|---|---|---|---|---|
| Real only | Poland | 6.4% | 15% | 12.8% |
| Real + Fake (4:1) | Poland | **5.4%** | 20% | 11.4% |
| Real + Fake (2:1) | Poland | 8.4% | **12%** | 11.4% |
| Real + Fake (3:2) | Poland | 7.4% | 15% | **11%** |
| Real + Fake (1:1) | Poland | **5.4%** | 15% | 11.4% |

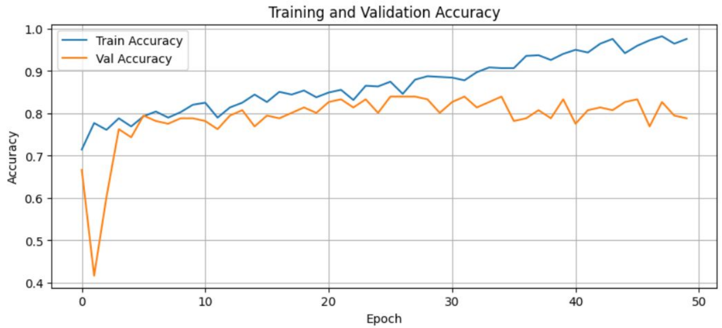
# Comparing The Models by Accuracy

- All the training is done on Egypt, testing on both **Egypt** and **Poland**.
- Some of the augmentation is done by horizontal flipping of both real and synthetic images.

| Train set | Real (624) | Real (780) | Real (780) + Synthetic (400) | Real (780) + Synthetic (1580) |
|---|---|---|---|---|
| Test on | Egypt | Poland | Poland | Poland |
| **Resnet** | 80.26% | 61.3% | 60.15% | **65.62%** |
| **ViT** | 74.3% | **60.5%** | 57.4% | 60.2% |
| **CNN-Transformer** | 83% | 63.5% | 62.27% | **63.8%** |

# Training (CNN-Transformer) on the source

- The model is overfitting, likely because of the small amount of training data
- Most of our data is dominant on the benign side.
- In real world scenarios, such a low recall is highly problematic


| | precision | recall | f1-score |
|---|---|---|---|
| Malignant (0) | 0.7027 | 0.5417 | 0.6118 |
| Benign (1) | 0.8151 | 0.8981 | 0.8546 |

# Discussion: CNNs, Transformers

- It seems that transformers perform poorly for medical image classification, perhaps an ever bigger patch size is needed in the transformers.
- Domain gaps don’t account for the general accuracy of the model, they are just a way to show how accuracies are different across domains.
- CNN based models are better at learning from both synthetic and ultrasound data

# Discussion: Datasets and results

- Generating good fake images is extremely difficult with smaller datasets.
- It's hard to classify if tumors are benign or malignant within the same domain, let alone other domains.

## Future Ideas:

  - Explore distance metrics between images
  - Distance metrics perform well in image comparisons

# References


1. Al-dhabyani, W., et al. (2020). **Dataset of breast ultrasound images**. *Science Direct.* Paper. Dataset.
2. Yang, J, et al. (2022). **MedMNIST v2 – A large-scale lightweight benchmark for 2D and 3D biomedical image classification**. *Cornell University* Paper.
3. Pawłowska, A., et al. (2024). **Curate benchmark dataset for ultrasound based breast lesion analysis.** *Sci Data.* Paper. Dataset.
4. Dosovitskiy, A., et al. (2020). **An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale**. *Cornell University.* Paper.
5. He, K., et al. (2015). **Deep Residual Learning for Image Recognition**. *Cornell University.* Paper.
6. Kebaili, A., et al. (2023). **Deep Learning Approaches for Data Augmentation in Medical Imaging: A Review.** *MDPI*. Paper.
7. Guan, H., Liu, M. (2021). **Domain Adaptation for Medical Image Analysis: A Survey**. *PubMed Central.* Paper.
8. Dai, Y., Gao, Y. (2021). **TransMed: Transformers Advance Multimodal Medical Image Classification.** C*ornell University*. Paper.
9. Arjovsky, M., et al. (2017). **Wasserstein GAN**. *Cornell University.* Paper

# Contributions

**Saad Alrajhi:** BIG-WGAN, CNN-Transformer

**Drew Kulischak:** ResNet, Vision Transformer

**Nick Russert:** Gan Exploration, Synthetic Data

**Francis Fernandez:** ResNet, GANs

**Chao Jung Wu:** ResNet