A large, semi-transparent watermark of a microscopic tissue sample is visible across the slide. The tissue shows various cell types, some with dark brown nuclei, suggesting a specific marker or stain. A small orange horizontal bar is located in the top left corner.

Ben Pfeffer

# Generating Cancer Images to Improve Cancer Diagnosis Accuracies

---



# Problem

- Early diagnosis crucial
- Cancer misdiagnosis rate up to 28% [1]
- Plenty of room for improvement

# Problem

---

- HIPAA
- Lack of medical data
- AI lagging in healthcare [2]



# Problem

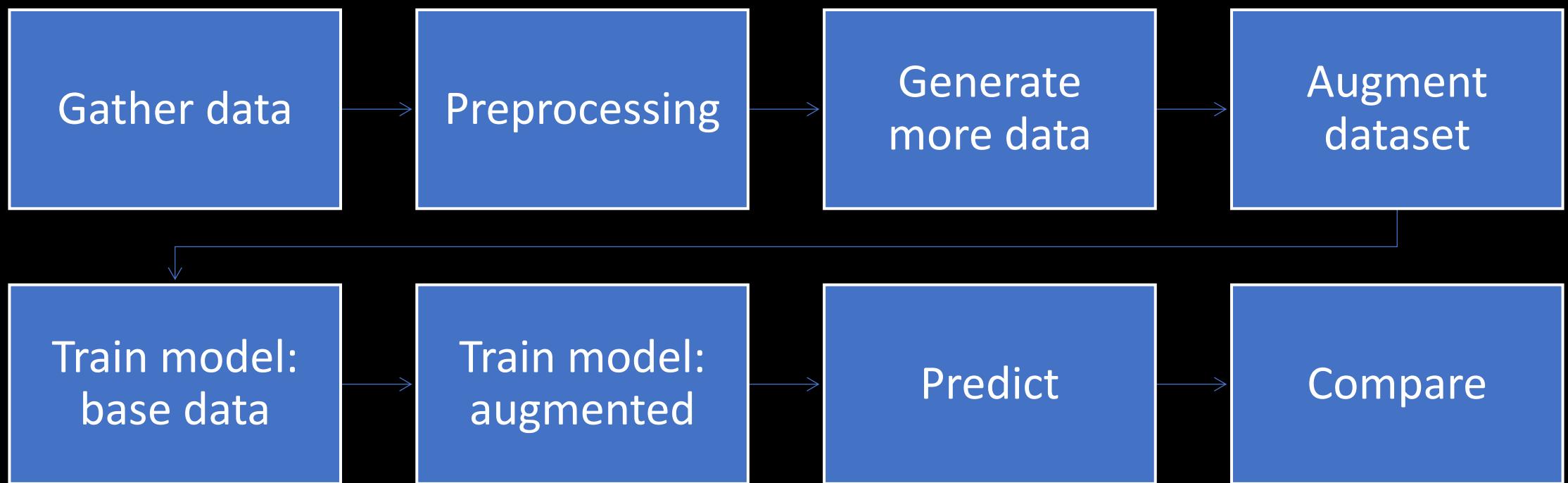
- Need. More. Data.
- Generative Adversarial Networks (GANs) [3]

Image data  
+ Ordinal label

= Ordinal Convolutional Neural Network (O-CNN)

# Architecture

---



# Data Collection and Preprocessing

---

# Stanford Tissue Microarray Database

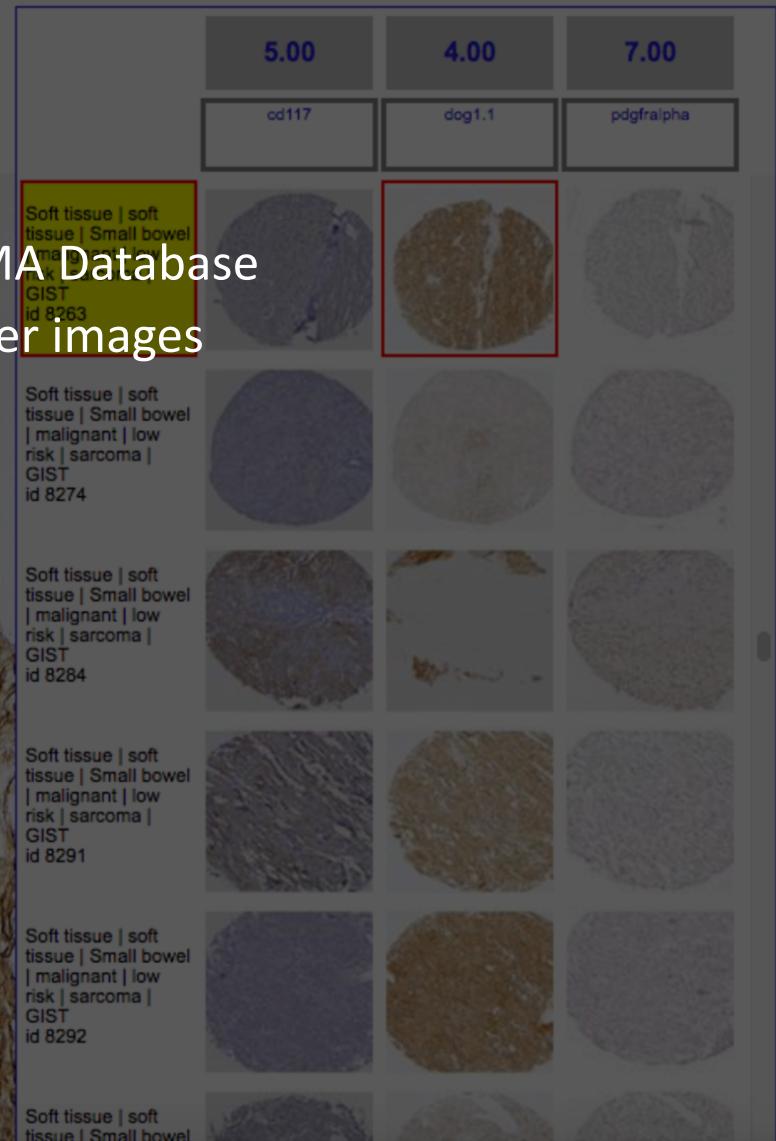
## Evaluate staining patterns across tissues and probes

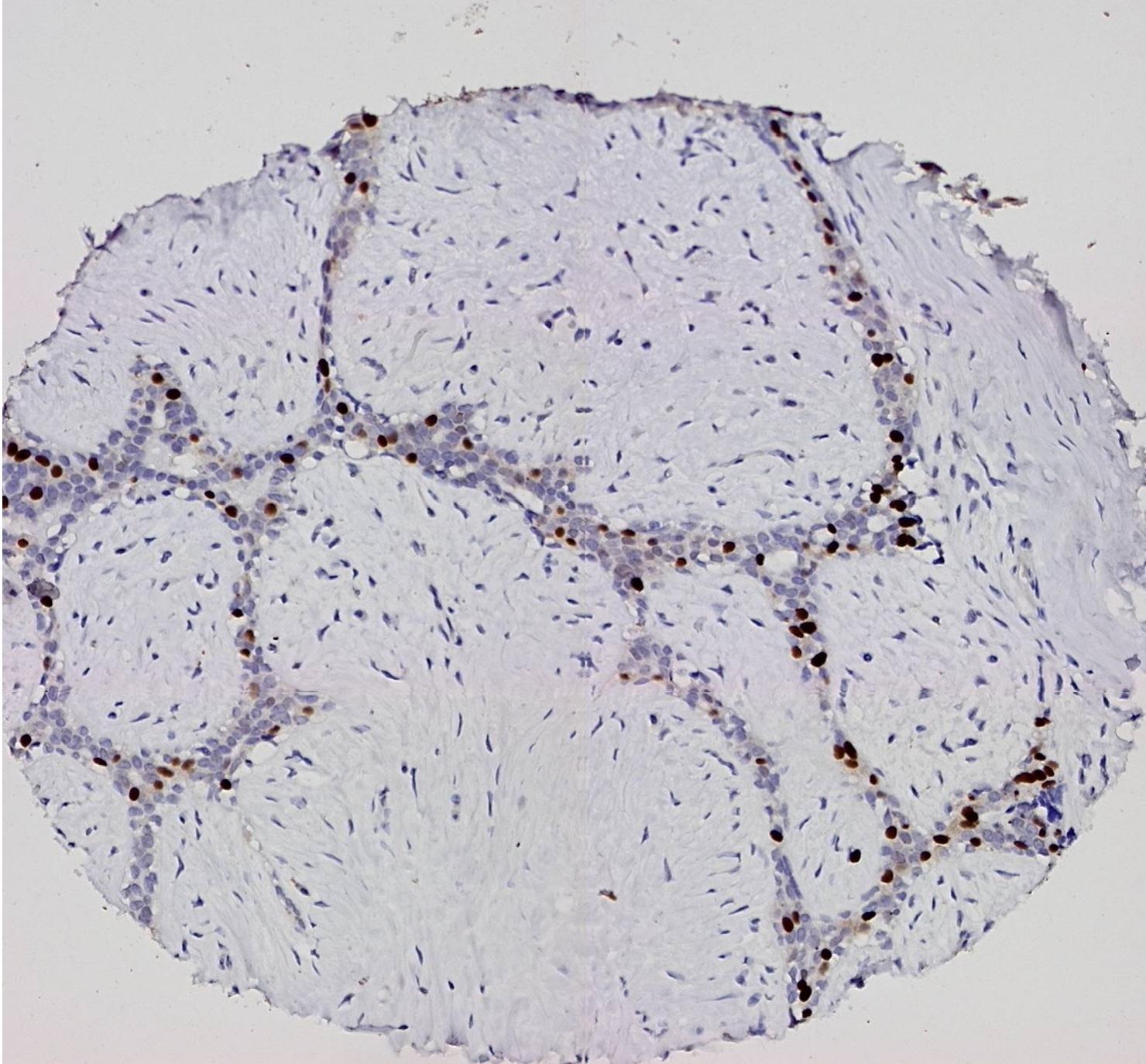
id 8263 | Soft tissue | soft tissue | Small bowel | malignant | low risk | sarcoma | GIST

slice 4.00 dog1.1

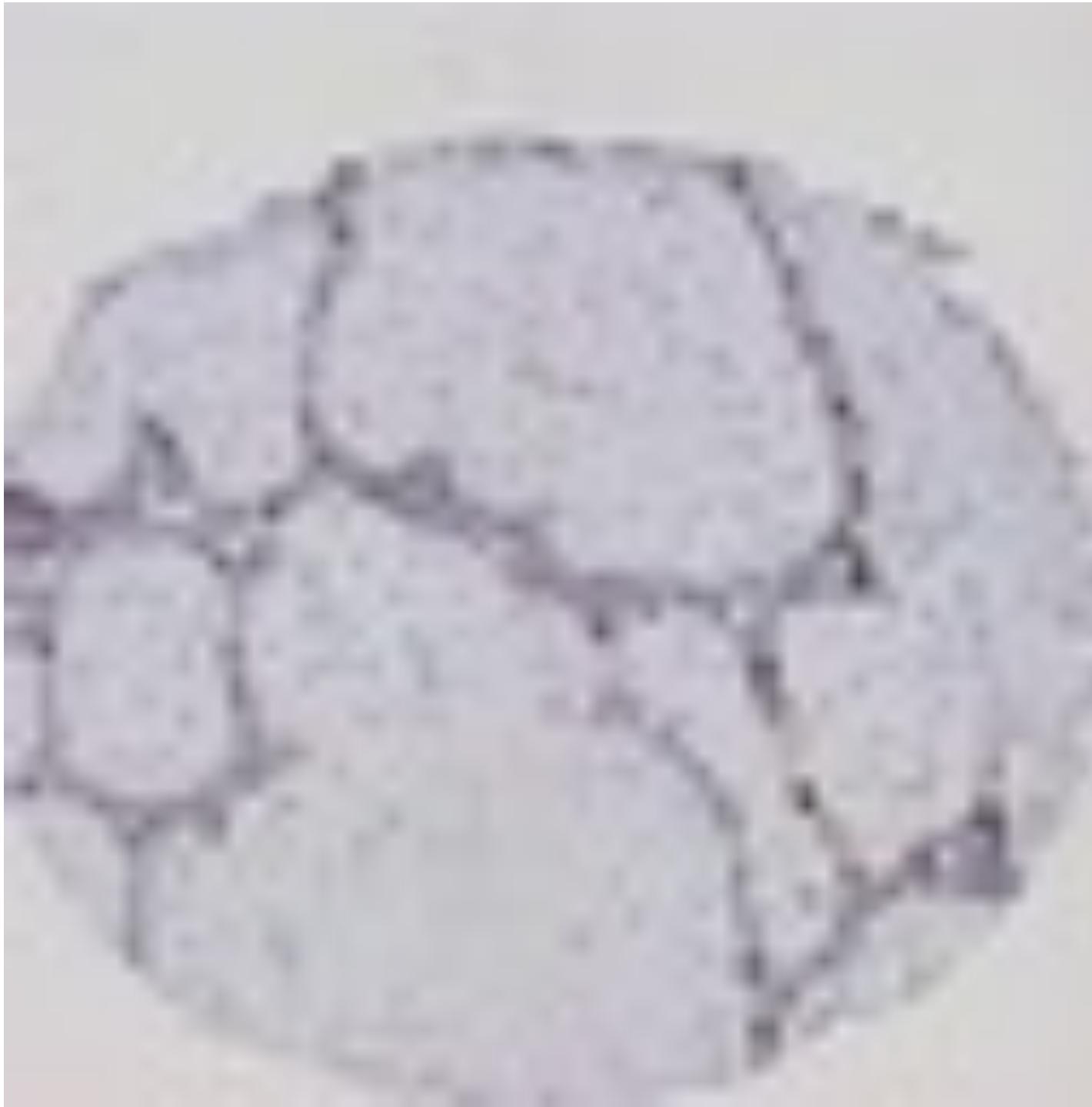
Where to get  
the data?

- Scrape Stanford's TMA Database
  - 665 breast cancer images
  - H-score





1504x1400

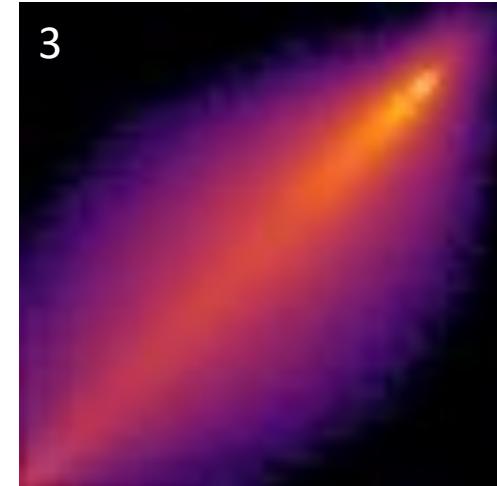
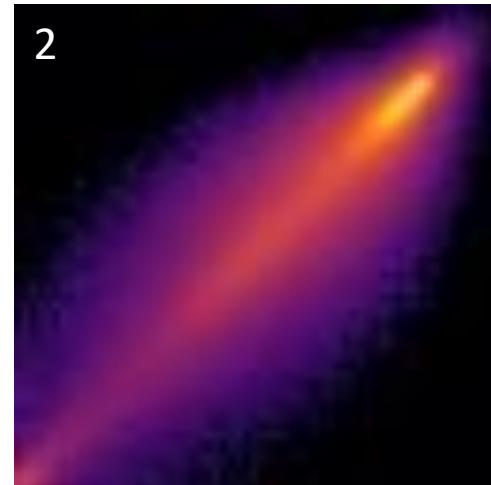
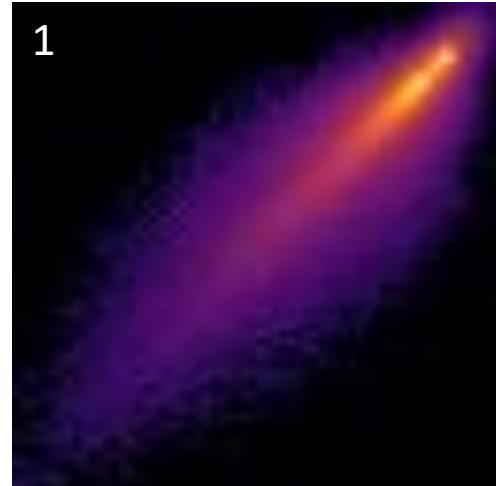
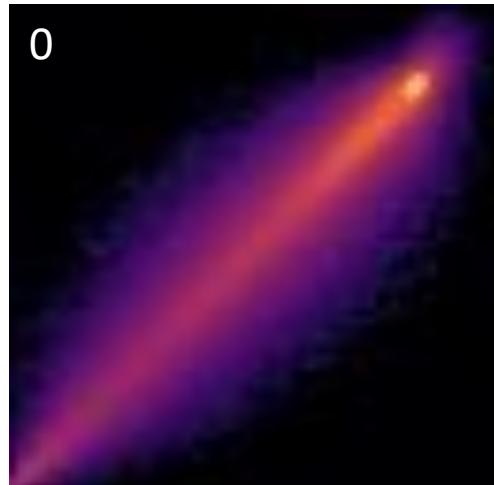


64x64

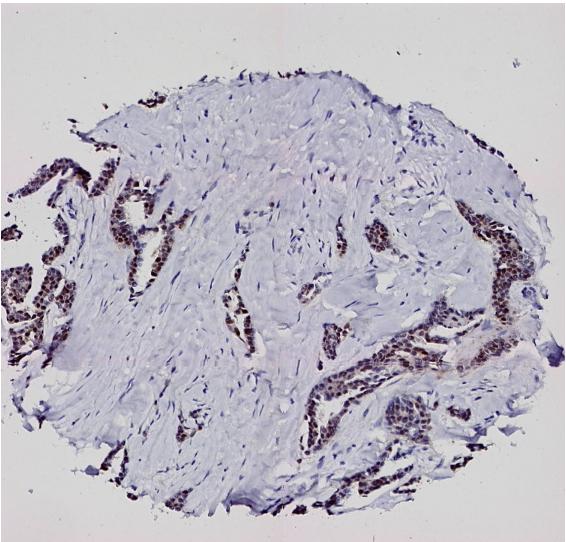
# How to reduce dimensions and maintain relational structure of the data?

- Convert into Gray-Level Co-Occurrence matrices (GLCMs)
  - Grayscale
  - Select direction and distance  1

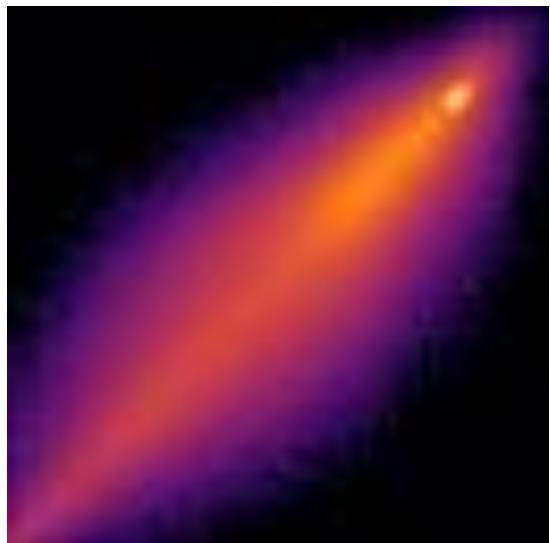
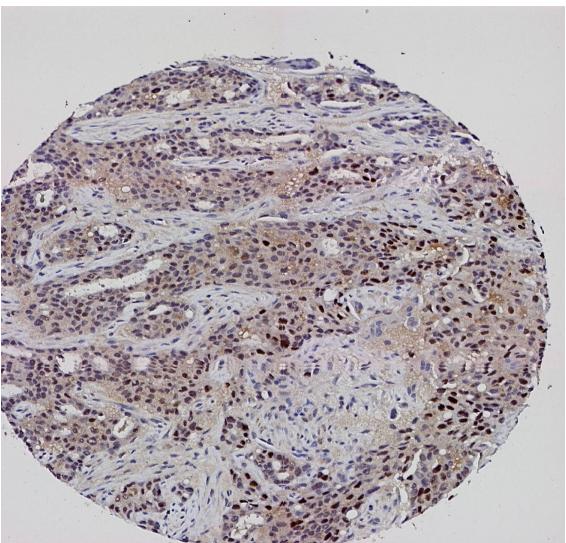
- 64x64

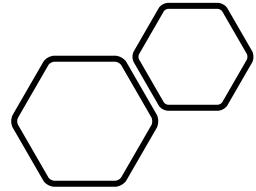


# Different image, similar GLCM



H-Score: 3

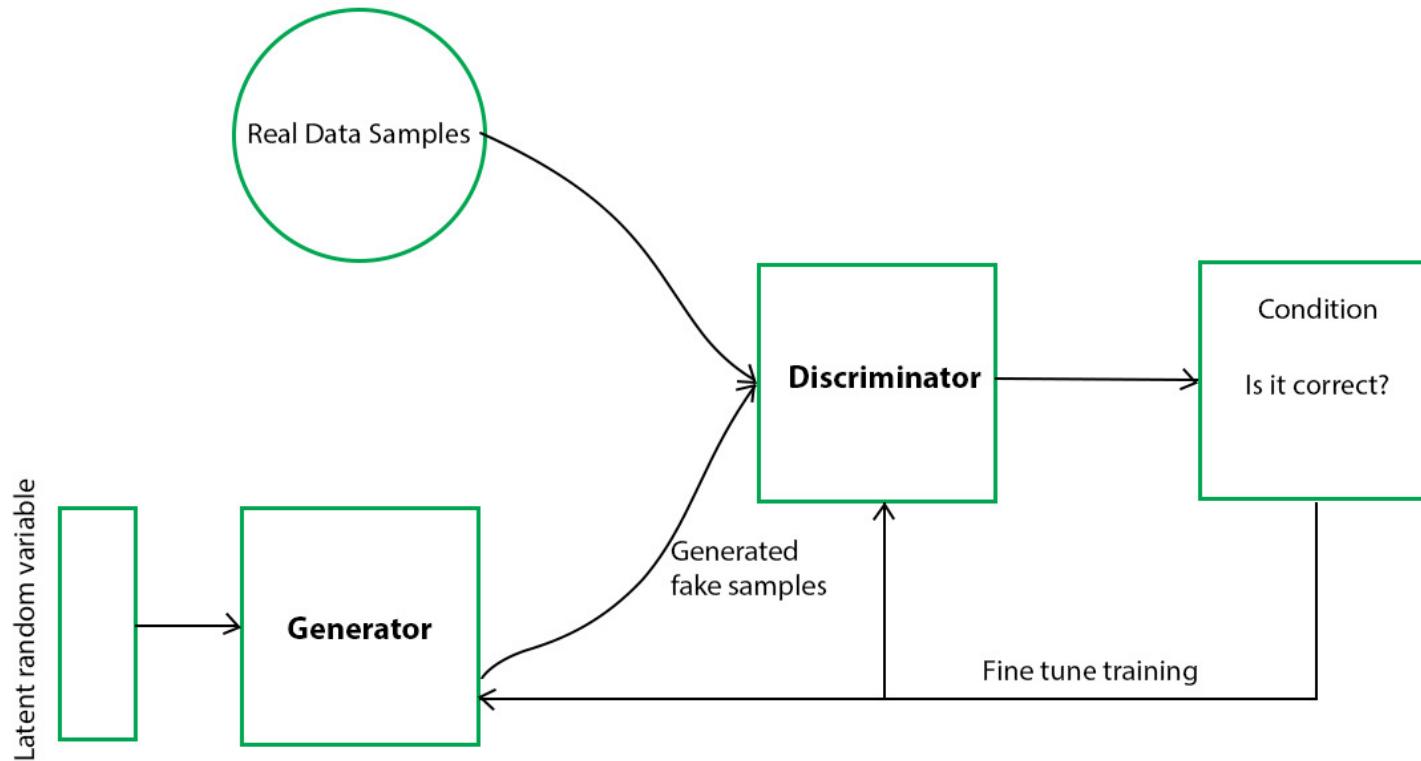




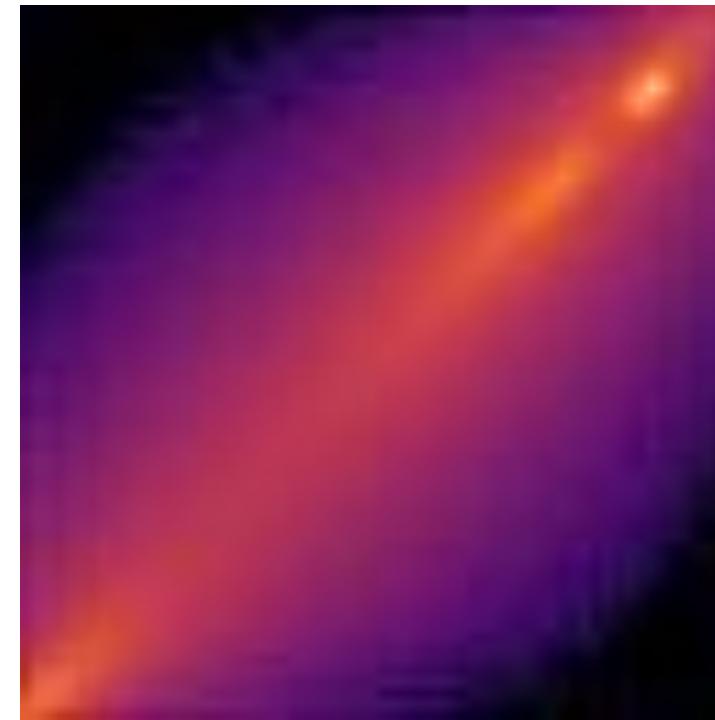
# Data Generation & Augmentation

# DC-GAN

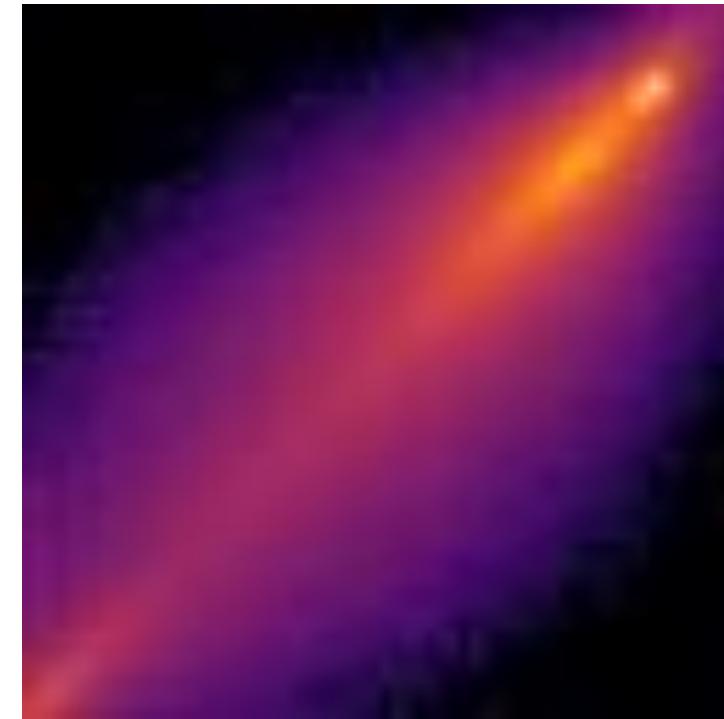
- How do GANs work?



# Which one is Generated?



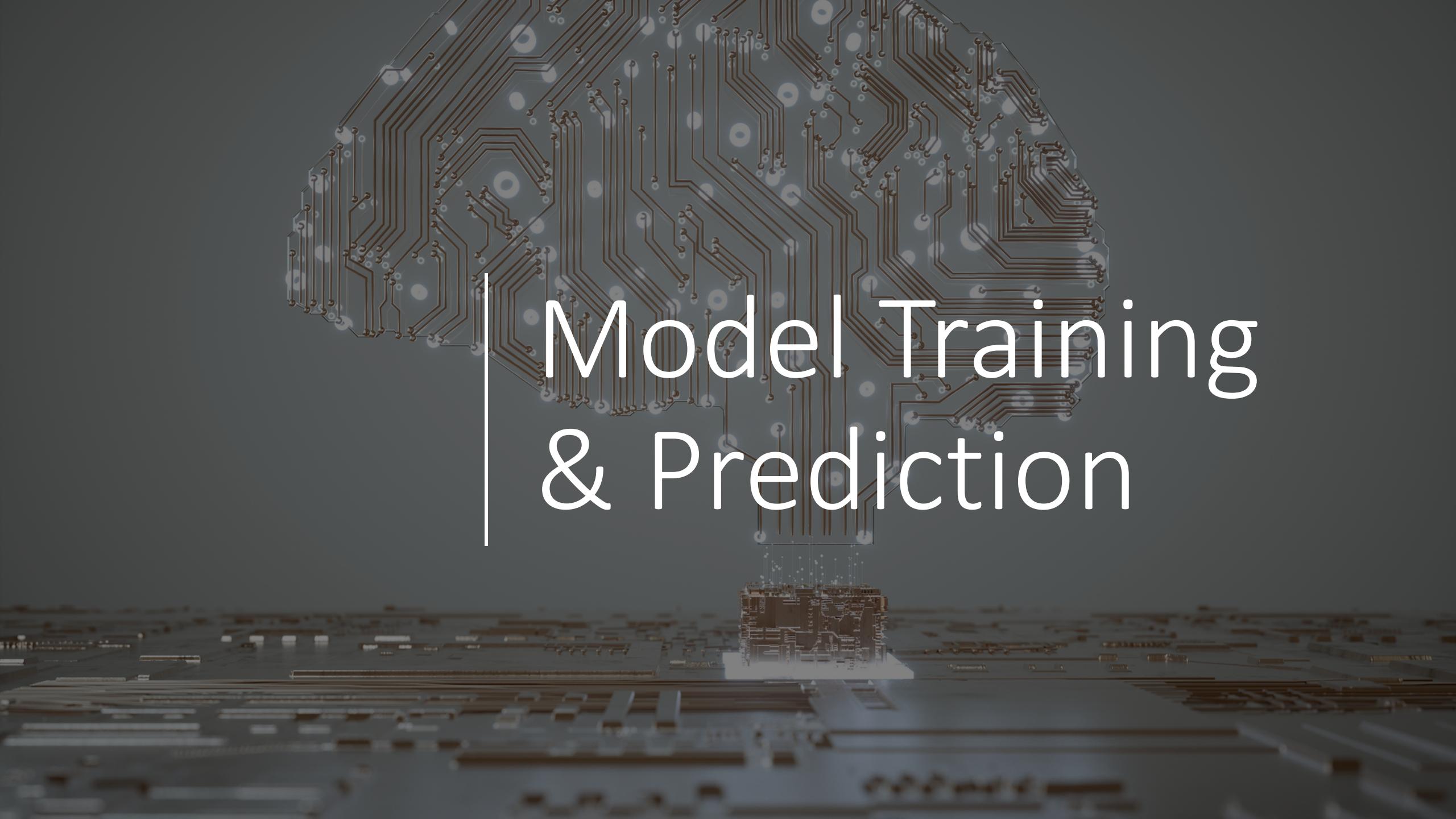
# Which one is Generated?





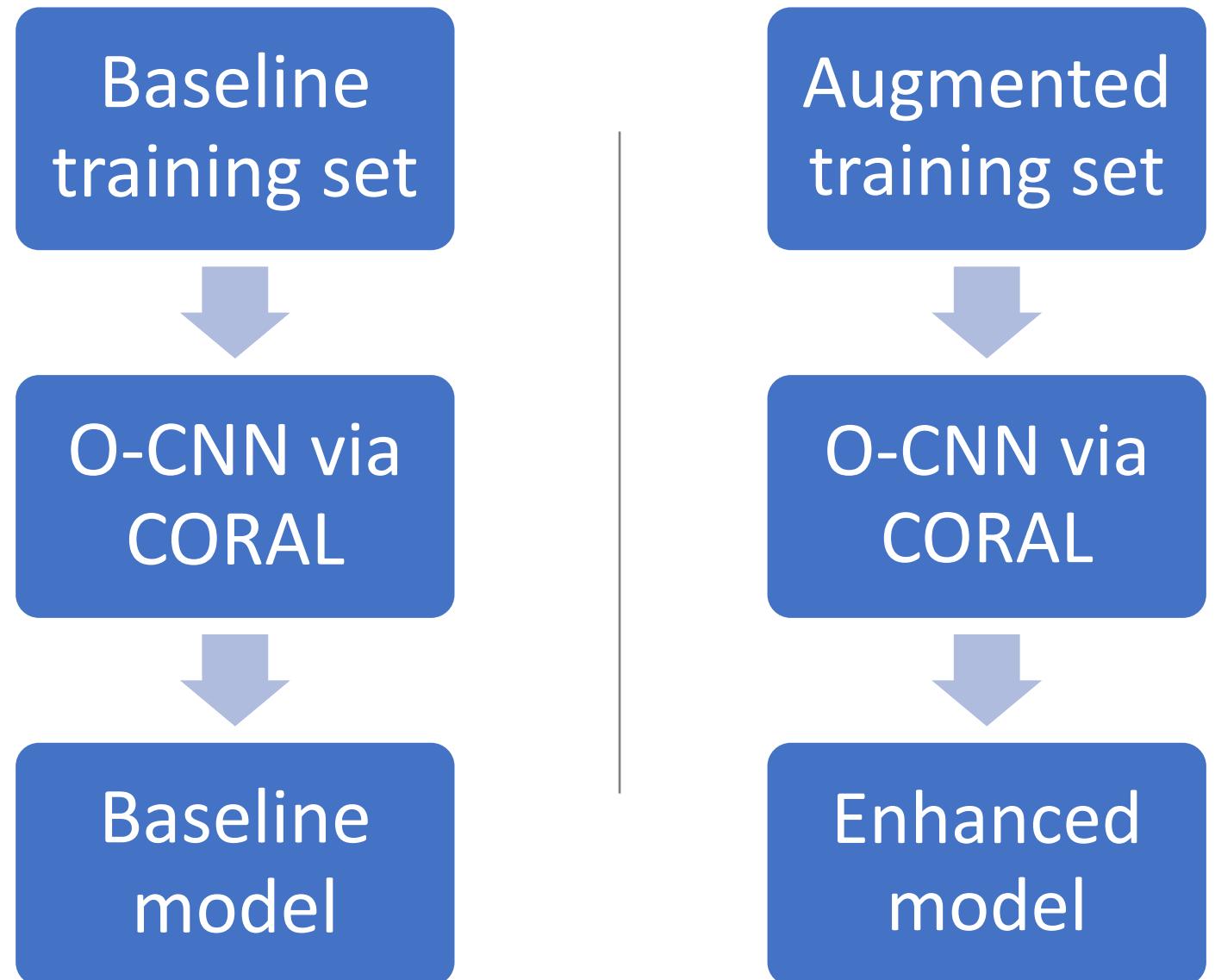
# Augment the dataset

- Duplicate
- Maintain Train/test split
- Augment one training set with these generated images.



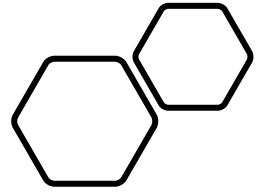
# Model Training & Prediction

Train O-CNN  
on Both  
Datasets



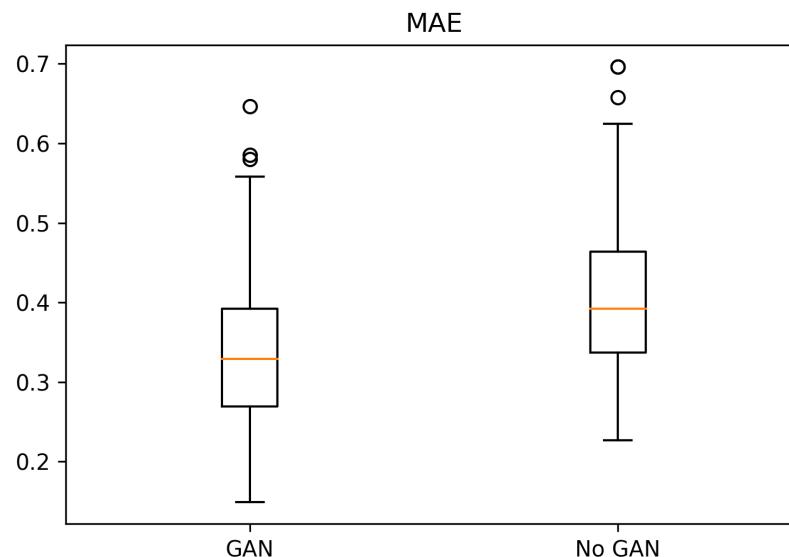
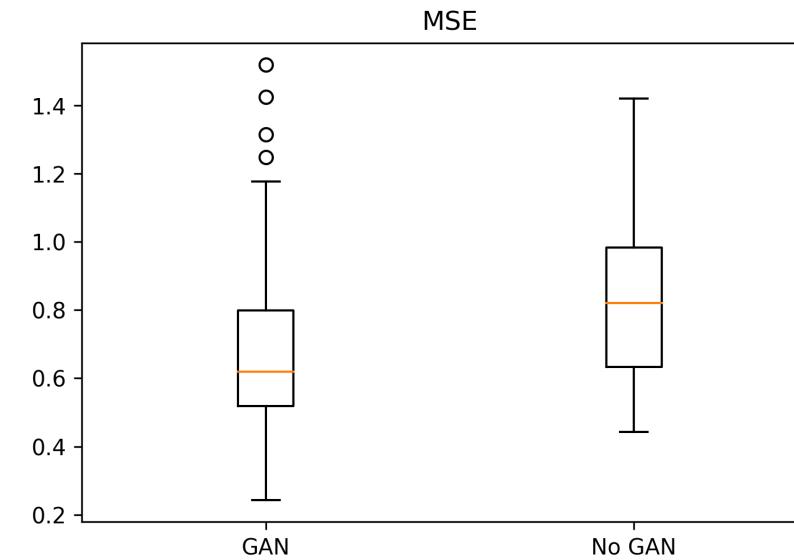
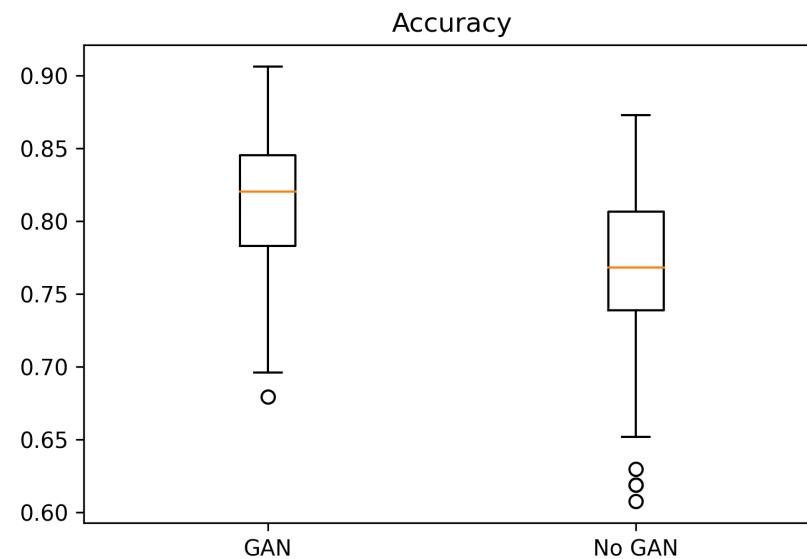
# Compare models

- Predict on identical test sets
- Scored
- Repeated last half of steps 100 times
- Compared
- Accuracy, MAE, MSE
- Convergence



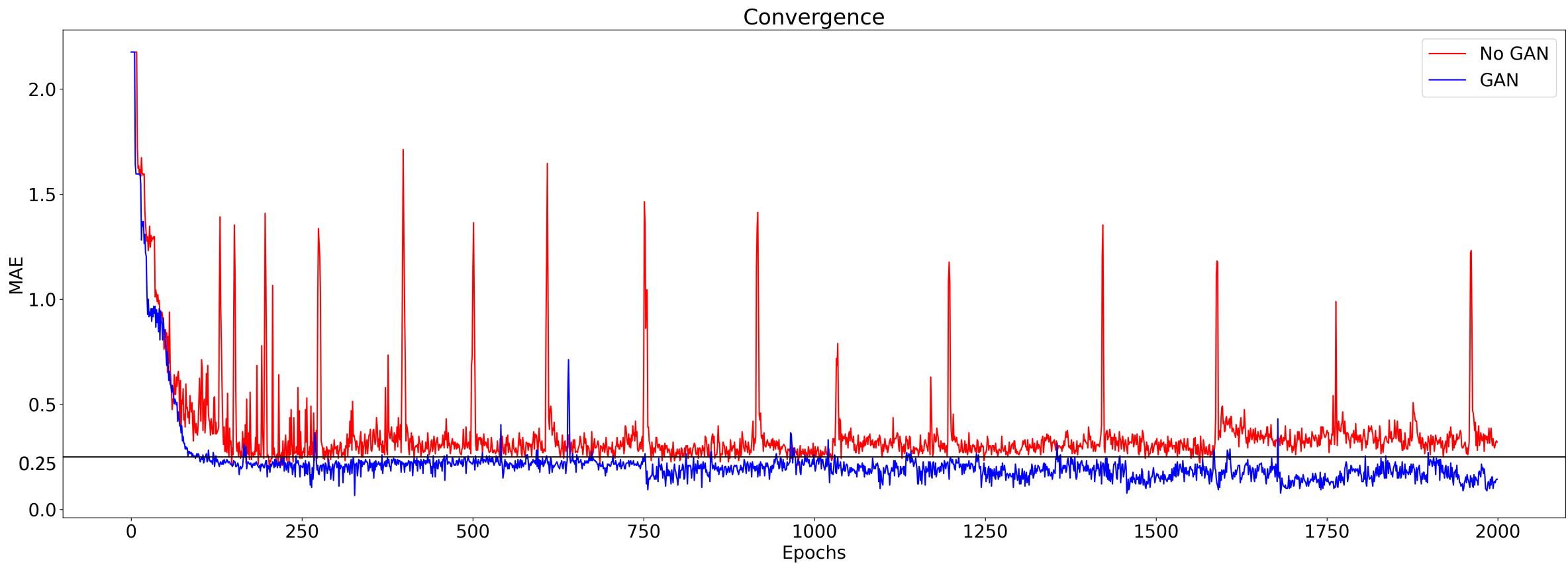
# Findings

# Scoring of the O-CNNs

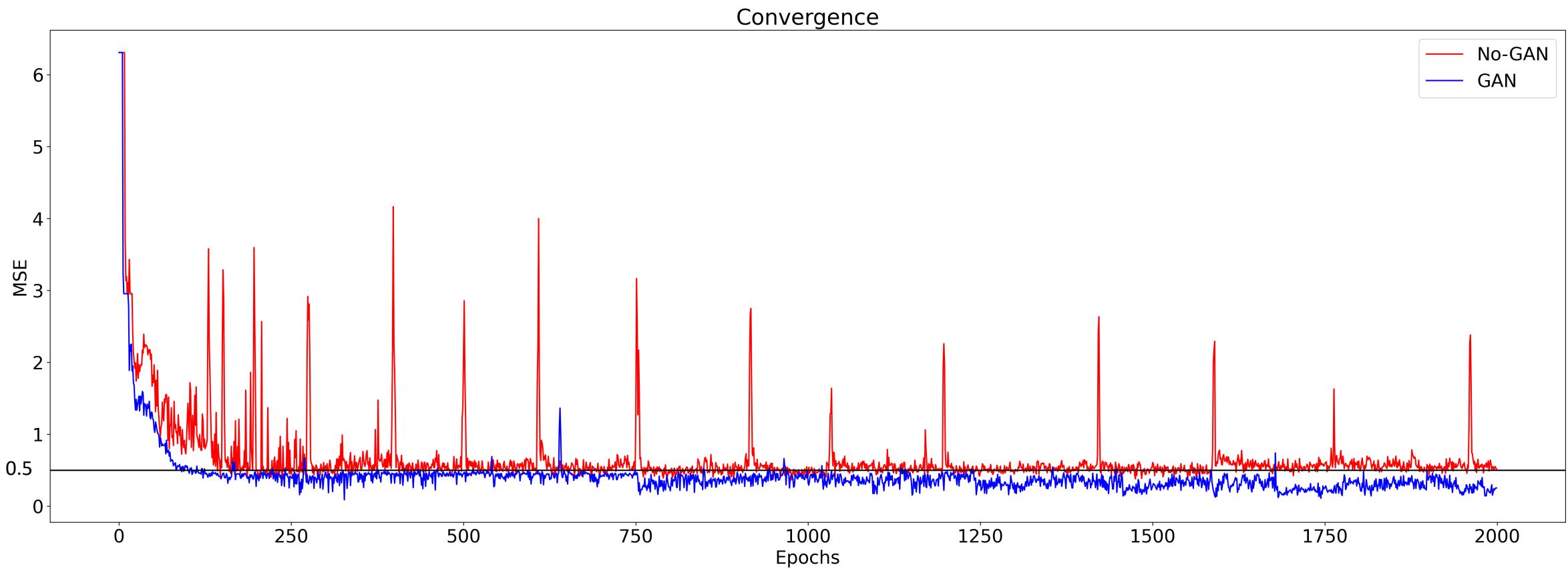


	GAN	No GAN
Accuracy	82.0%	76.8%
MAE	0.329	0.392
MSE	0.619	0.820

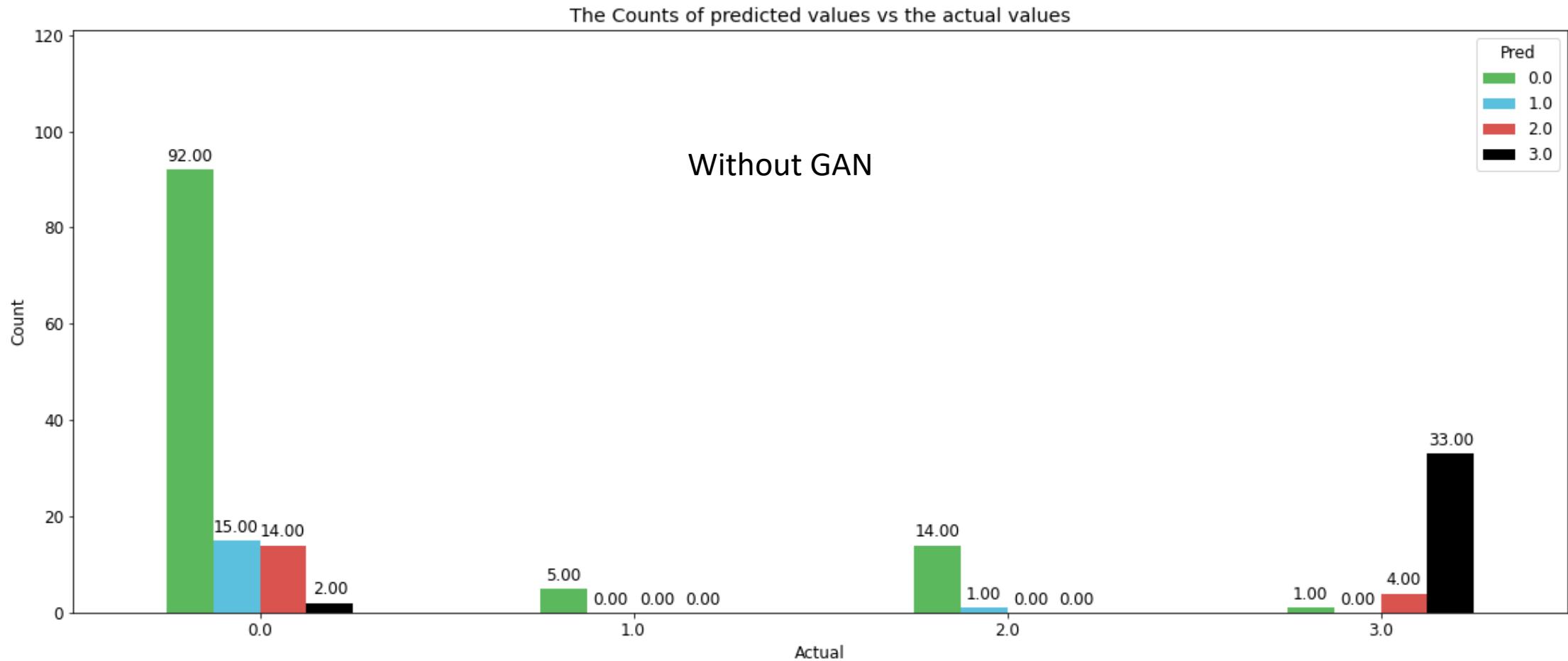
# Analyzing Convergence Rates



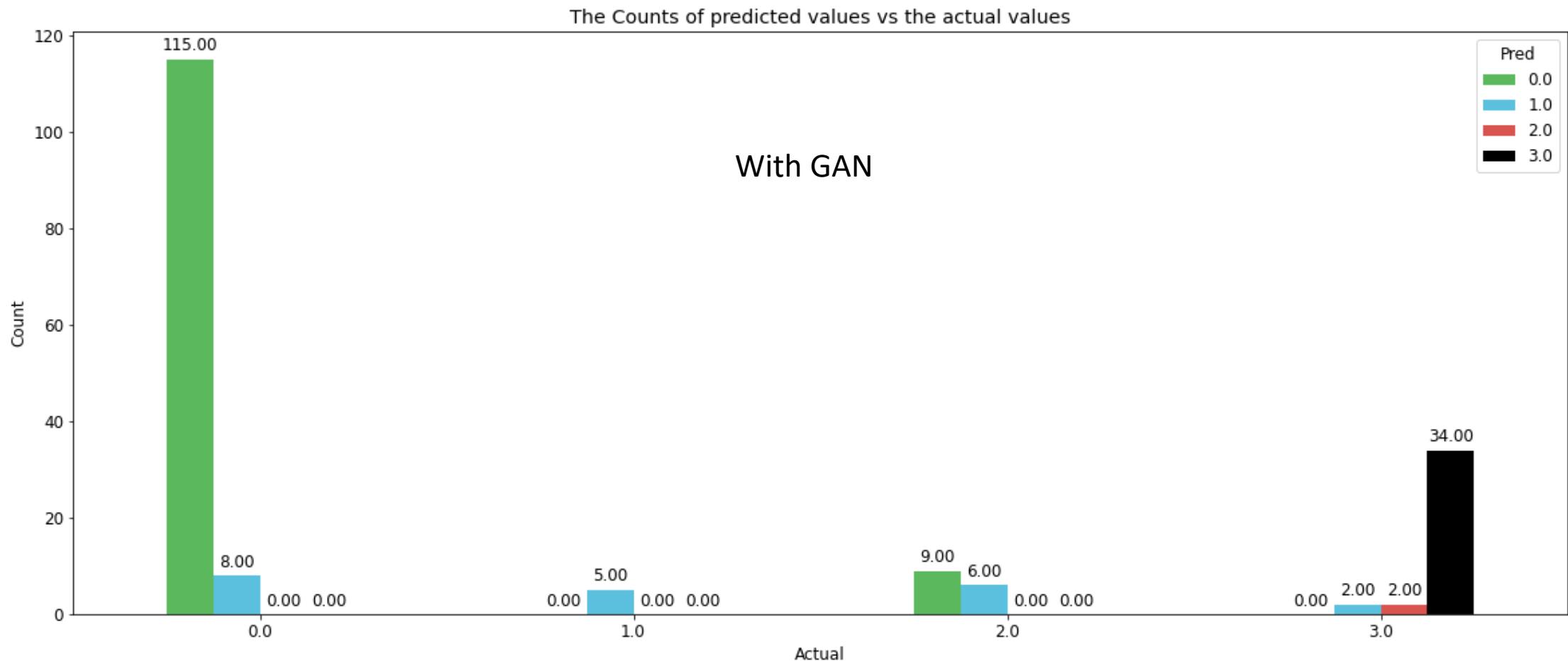
# Analyzing Convergence Rates



# Example results

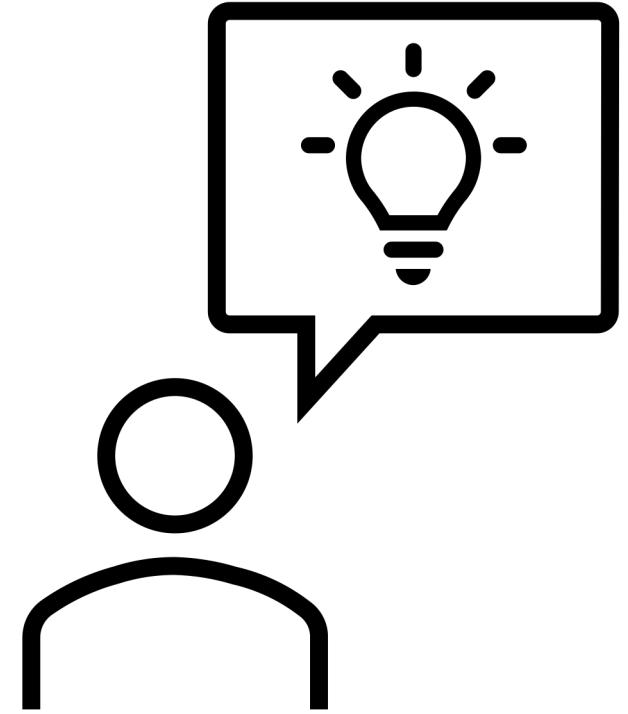


# Example results

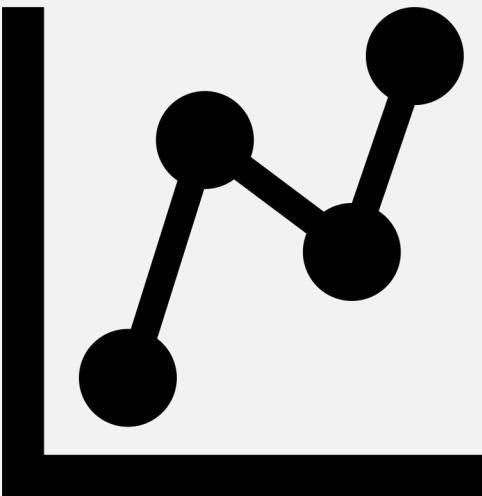


# Takeaways

---



# Result analysis

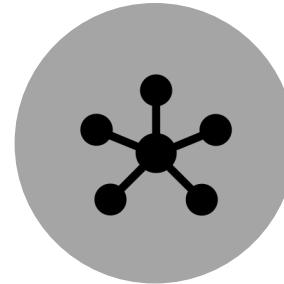


- GANs can be useful for data augmentation
- AI on cancer image data is better with a GAN
- GANs may improve overall training time

# Viable for general use?



Larger datasets



Different GANs



Outside the medical field

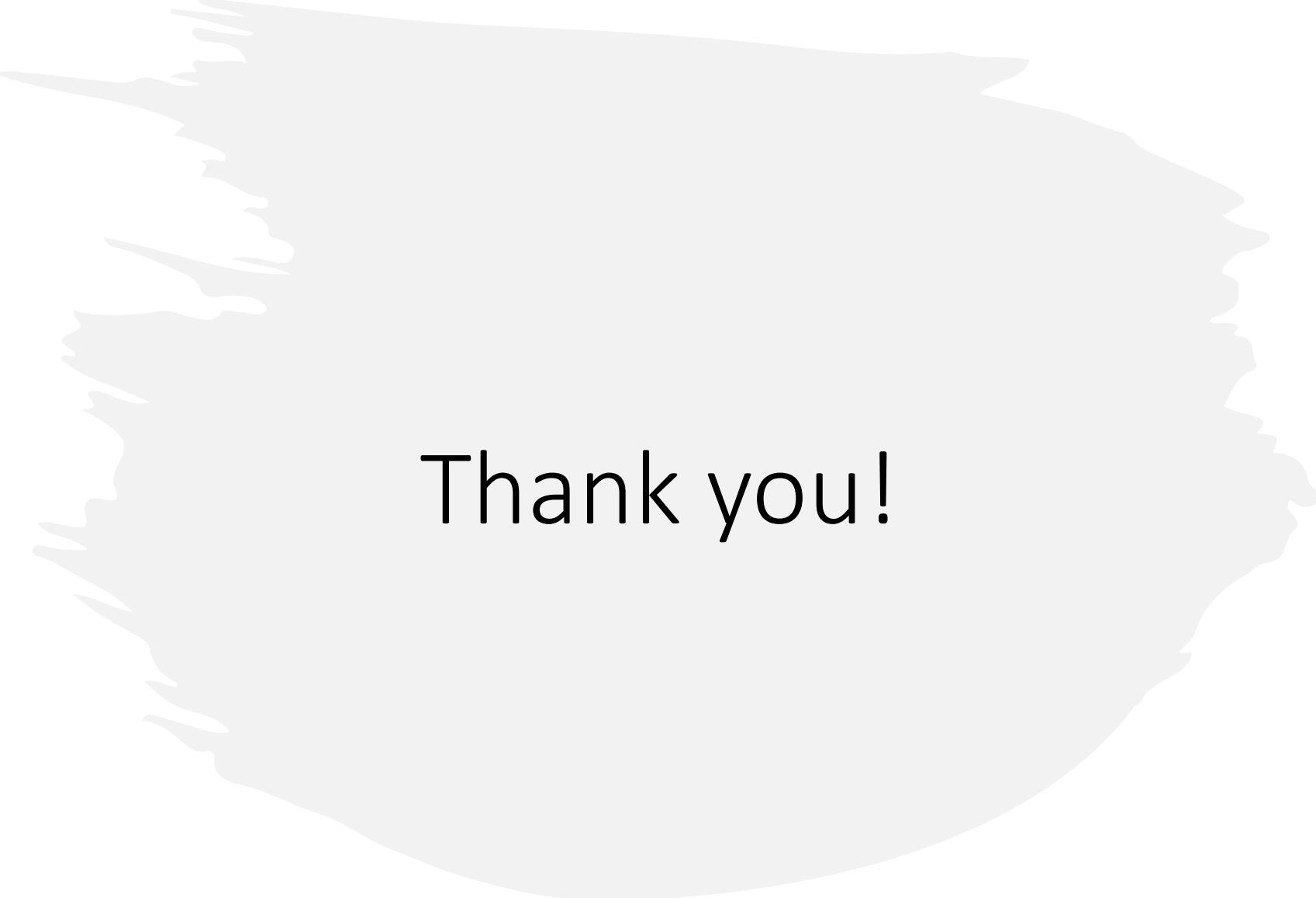


Larger images

# Acknowledgements

- Professor/Mentor Donghui Yan





Thank you!

Questions?

# References

- [1] "How is cancer misdiagnosed - missing the warning signs as a doctor," *Paul & Perkins*. [Online]. Available: <https://paulandperkins.com/cancer-misdiagnosed/>.
- [2] A. Goldfarb and F. Teodoridis, "Why is AI adoption in health care lagging?," *Brookings*, 09-Mar-2022. [Online]. Available: <https://www.brookings.edu/research/why-is-ai-adoption-in-health-care-lagging/>.
- [3] I. J. Goodfellow, "Generative adversarial nets," *arXiv.org*, 10-Jun-2014. [Online]. Available: <https://arxiv.org/pdf/1406.2661.pdf>.
- [4] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative Adversarial Networks," *arXiv.org*, 07-Jan-2016. [Online]. Available: <https://arxiv.org/abs/1511.06434v2>.
- [5] W. Caoa, V. Mirjalilib, and S. Raschkaa, "Rank consistent ordinal regression for neural networks with application to age estimation," *arXiv.org*, 13-Nov-2020. [Online]. Available: <https://arxiv.org/pdf/1901.07884.pdf>.