Speckle Pattern Clustering Strategies for Detecting Glaucoma

Seth Chatterton · Ian Kleinemolen · Josh Marchant

Sandip Mondal

Background and Motivation

- Glaucoma leads **reduced vascular branching** in patients [ref].

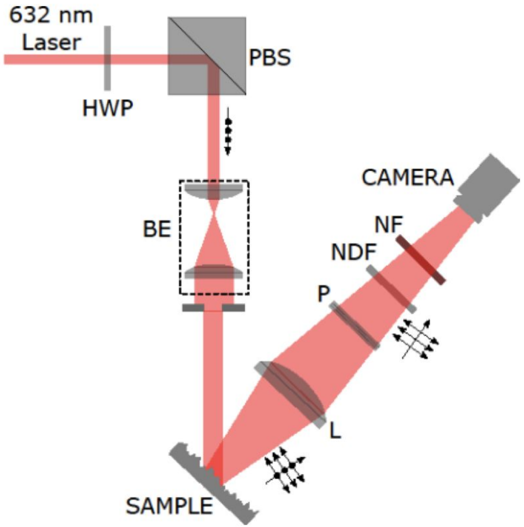
  Alterations in vascular → reduced tissue function

  Reduced tissue function → alterations in vasculature

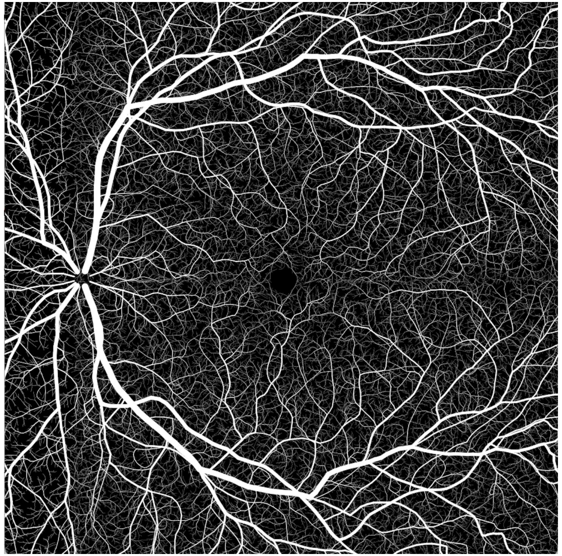
- **Potentially important biomarker** for diagnosis.
- High resolution vasculature imaging using optical coherence tomography (OCT) is technically challenging and expensive, **limiting widespread clinical viability**.

Background

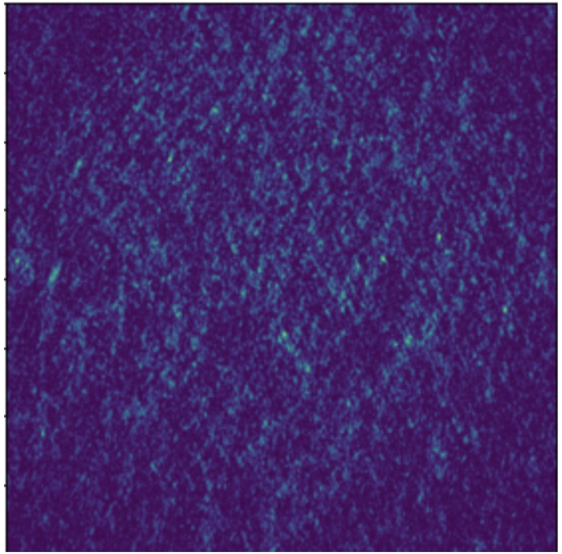
- Instead of directly imaging vasculature, lasers could be used to scatter light off of vasculature and produce unique network-specific speckle patterns.



(simulated) vascular network



(simulated) *diffraction* pattern



Methods

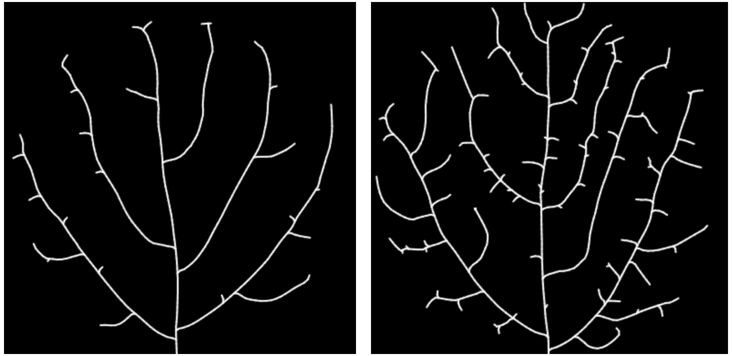
Hypothesis: clustering speckle data or relevant latent variables may enable characterization of higher (“healthy”) vs. lower (“pathological”) branching patterns.

Tools:

- **K-Means**
- **t-SNE**
- **Frequency space** clustering
- Variational autoencoder (**VAE**)

Methods

Data 1 split



Data generation:

3 sets of “toy model” networks:

- **Set 1:** 1%:10%, **Set 2:** 3%:7%, **Set 3:** 5%:6% branching probability

Data split (for each set):

- Train: 75+75, Test: 25+25 pathological+healthy

Diffractio Python toolbox used for generating light diffraction patterns.

- Raleigh Summerfield approximation of diffracted light propagation

Methods

We want to see if there is sufficient signal within our generated diffraction data to distinguish the “healthy” from the “pathological” cases.

- **Clustering (K-Means)**
- **t-SNE**
- **PCA**

All of the above are from the scikit learn library

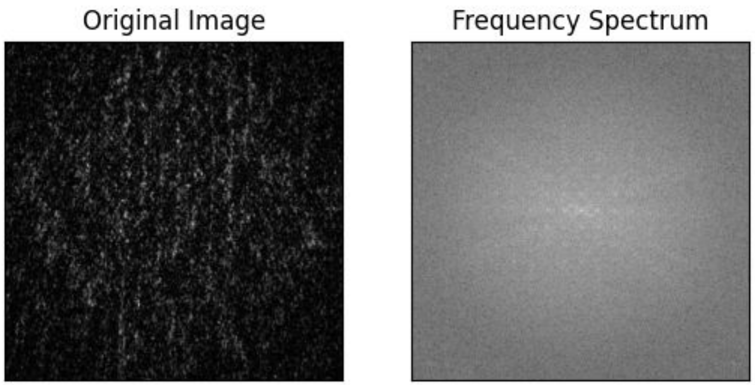
For each of the methods, we flatten the 400x400 far-field diffraction image into a vector of 160,000 elements.

Methods

Frequency Space

We want to explore whether transformation to the frequency space improves the ability to distinguish the “healthy” from the “pathological” cases, especially when diffraction values are normalized to simulate a more representative measurement.

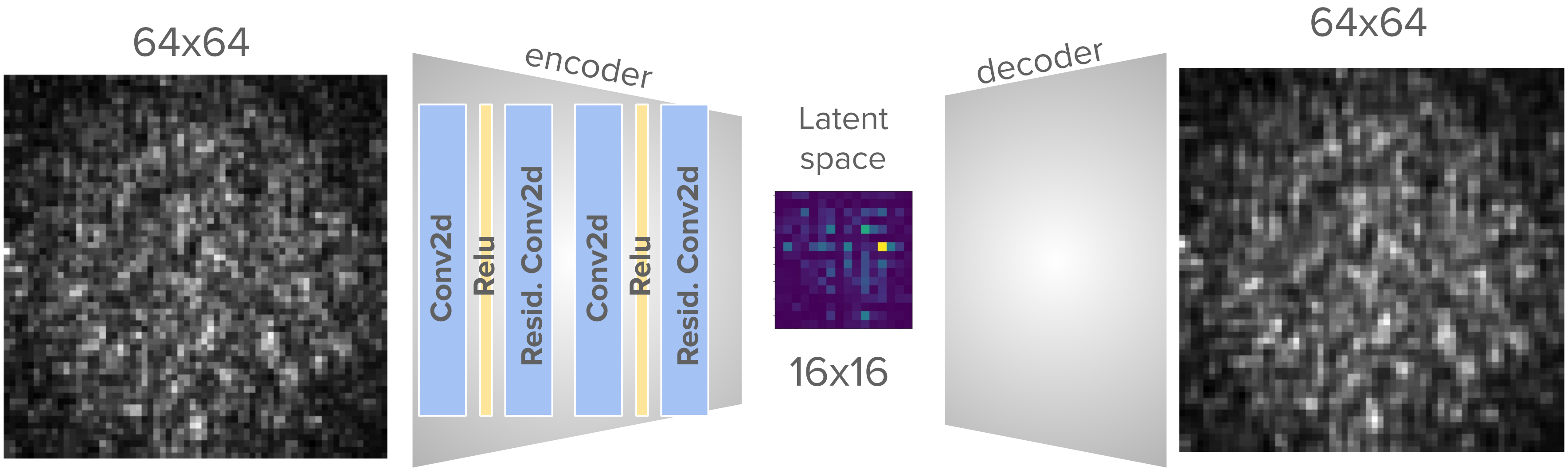
1. FFT run on diffraction images, frequencies shifted, and magnitude extracted
2. PCA used for feature extraction
3. K Means clustering



Methods

VAE:

- Created custom torch-compatible datasets (see link here)
- VAE and vector quantized (VQ)-VAE: github.com/Jackson-Kang/Pytorch-VAE-tutorial
- Trained anew for each dataset



Methods

VAE:

- Aside: vector quantized (VQ)-VAE: github.com/Jackson-Kang/Pytorch-VAE-tutorial

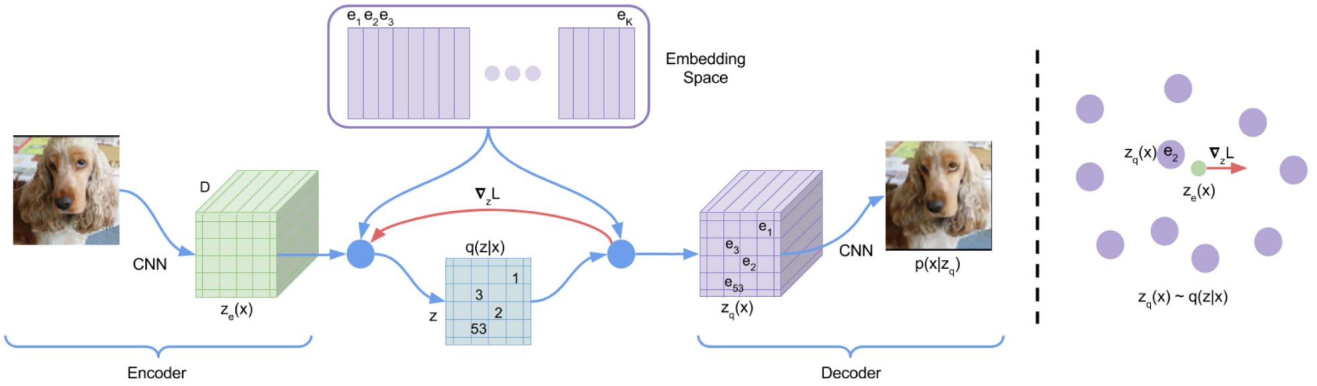


Figure 1: Left: A figure describing the VQ-VAE. Right: Visualisation of the embedding space. The output of the encoder $z(x)$ is mapped to the nearest point e_2. The gradient $\nabla_z L$ (in red) will push the encoder to change its output, which could alter the configuration in the next forward pass.

https://arxiv.org/pdf/1711.00937

Methods

VAE: How should we use the **16x16** latent space?

Ideas for clustering:

Global metric (average, max)

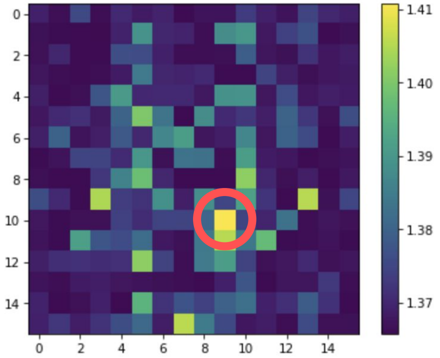
Most critical pixels in the latent space

```
for ii in range(xdim):
  for jj in range(ydim):
    z[:,ii,jj] = 0
    xhat_test = model.decoder(z)
    recon_loss = mse_loss(xhat_test, x)
```

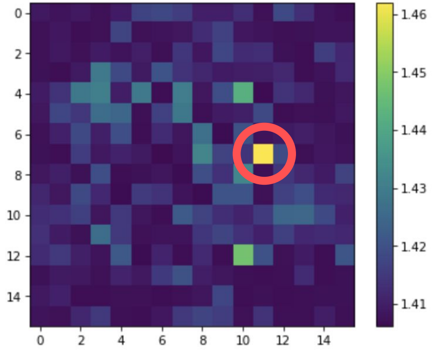
"LIME"-esque interpretability strategy...

"Loss effect landscape"

Path Modl # 1/25

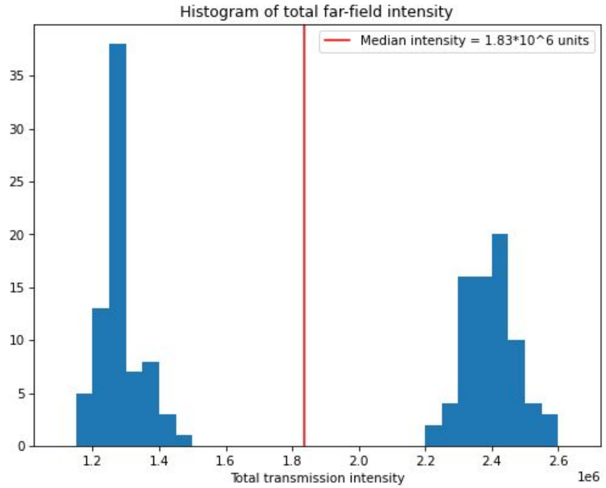


Heal. Modl # 1/25



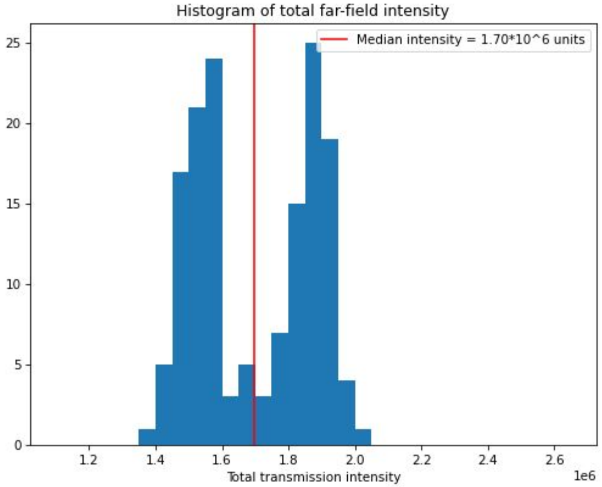
Results: a baseline intensity approach

Total diffraction intensity increases as branching probability increases, because there is more transmission.



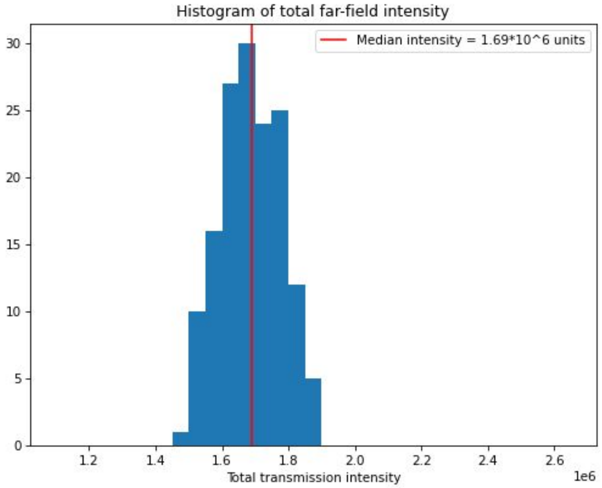
Dataset 1: high separation

```
Training Accuracy: 100.00%
Test Accuracy: 100.00%
```



Dataset 2: some separation

```
Training Accuracy: 98.67%
Test Accuracy: 98.00%
```

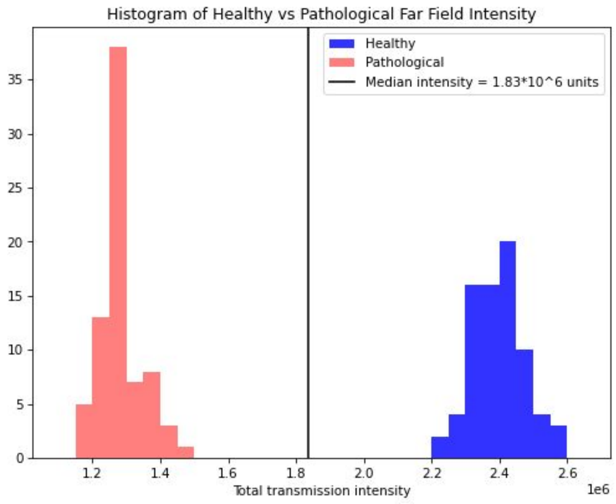


Dataset 3: little separation

```
Training Accuracy: 85.33%
Test Accuracy: 77.08%
```

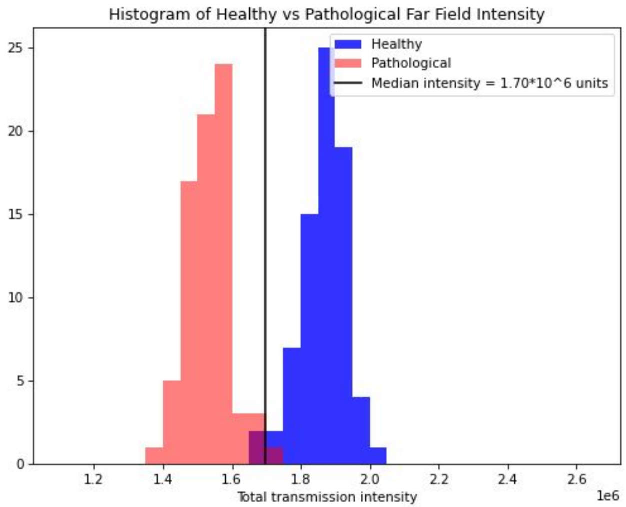
Results: a baseline intensity approach

Total diffraction intensity increases as branching probability increases, because there is more transmission.



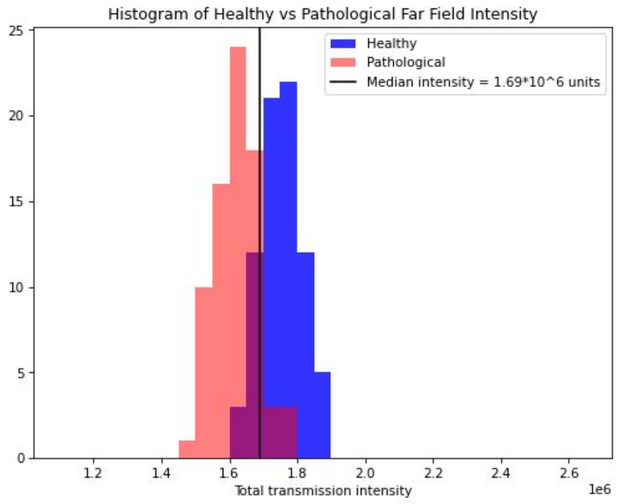
Dataset 1: high separation

`Training Accuracy: 100.00%`
`Test Accuracy: 100.00%`



Dataset 2: some separation

`Training Accuracy: 98.67%`
`Test Accuracy: 98.00%`

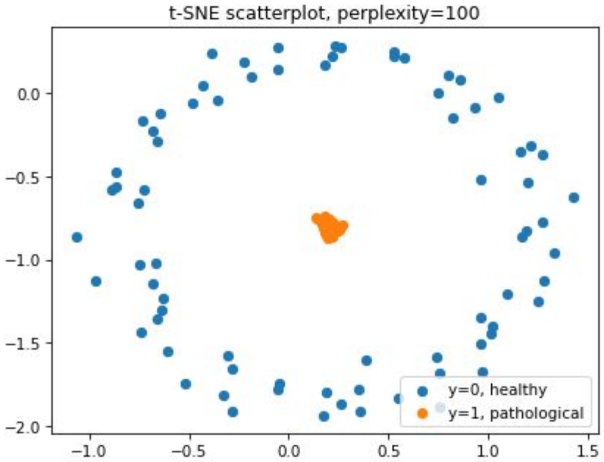


Dataset 3: little separation

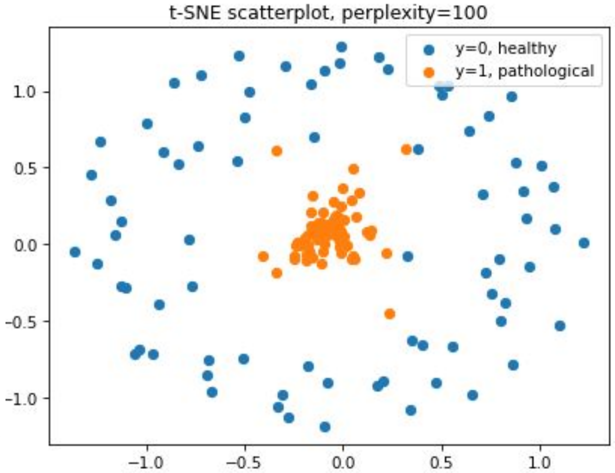
`Training Accuracy: 85.33%`
`Test Accuracy: 77.08%`

Results: t-SNE

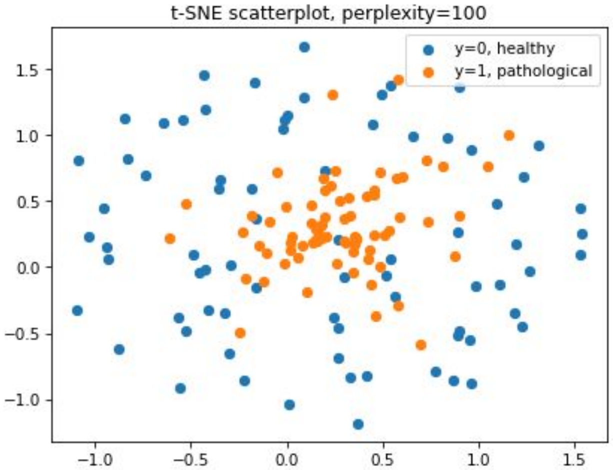
Visual inspection: does it look like we will be able to separate the two classes?



Dataset 1: high separation

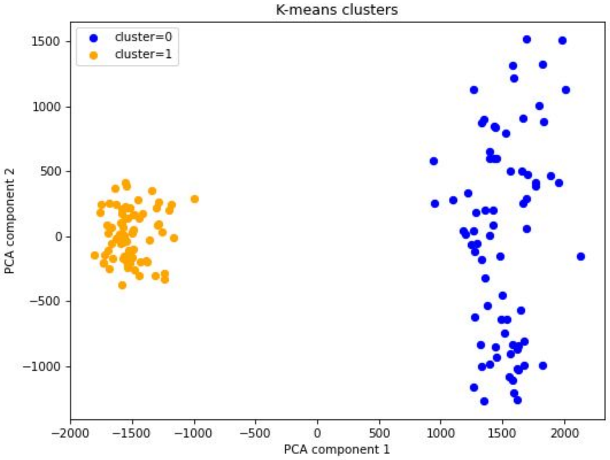


Dataset 2: some separation



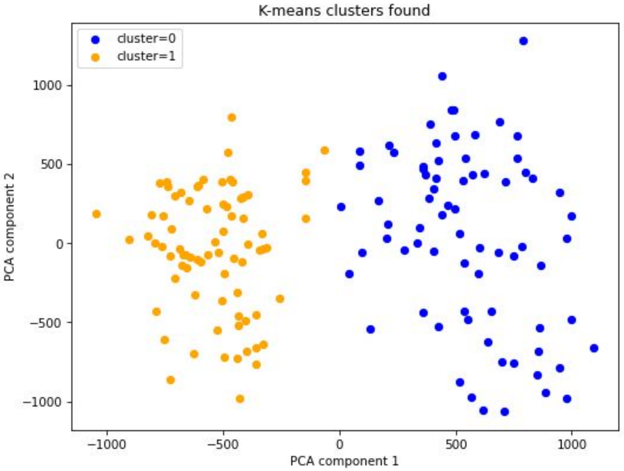
Dataset 3: little separation

Results: k-Means



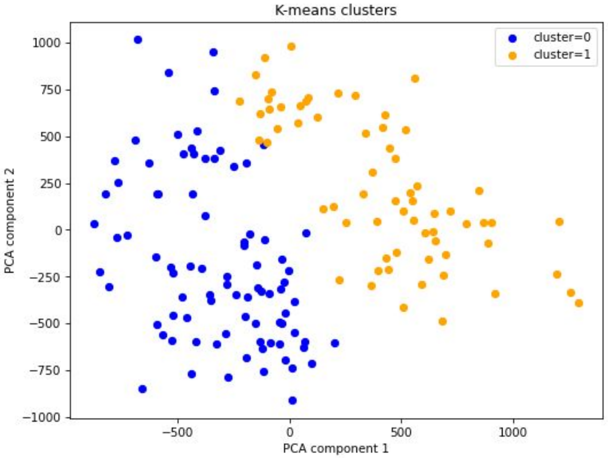
Dataset 1: high separation

`Test Accuracy: 100.00%`
`Train Accuracy: 100.00%`



Dataset 2: some separation

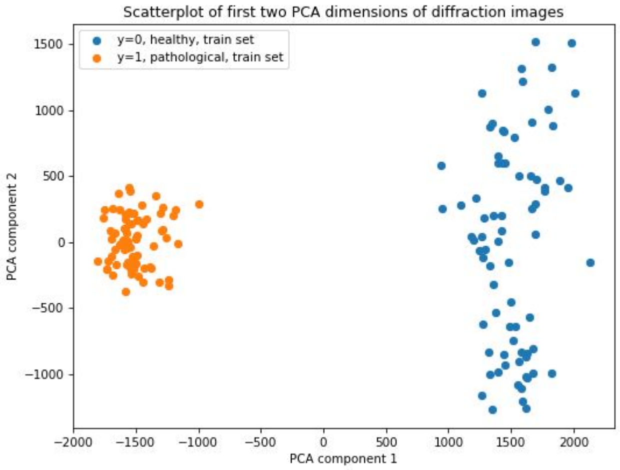
`Test Accuracy: 100.00%`
`Train Accuracy: 98.00%`



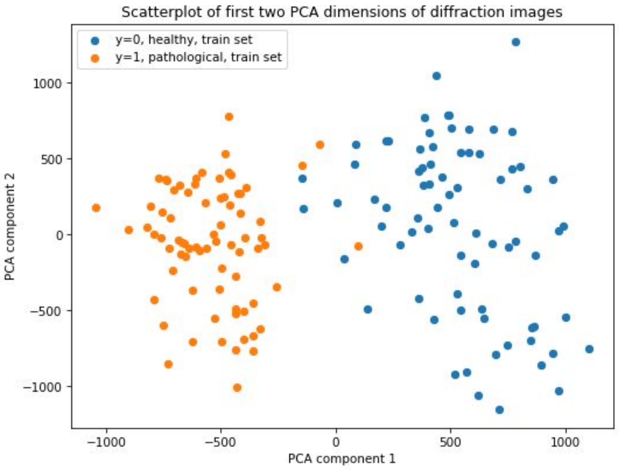
Dataset 3: little separation

`Test Accuracy: 52.08%`
`Train Accuracy: 52.00%`

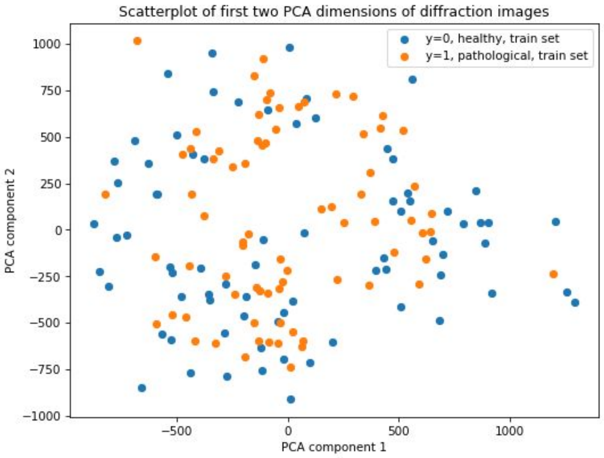
Results: PCA



Dataset 1: high separation



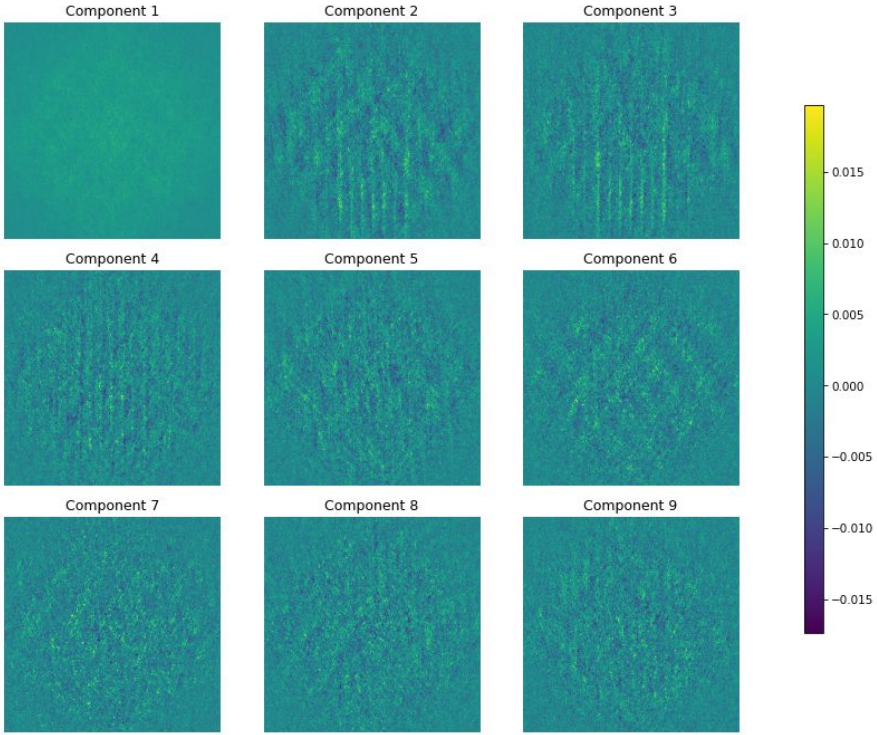
Dataset 2: some separation



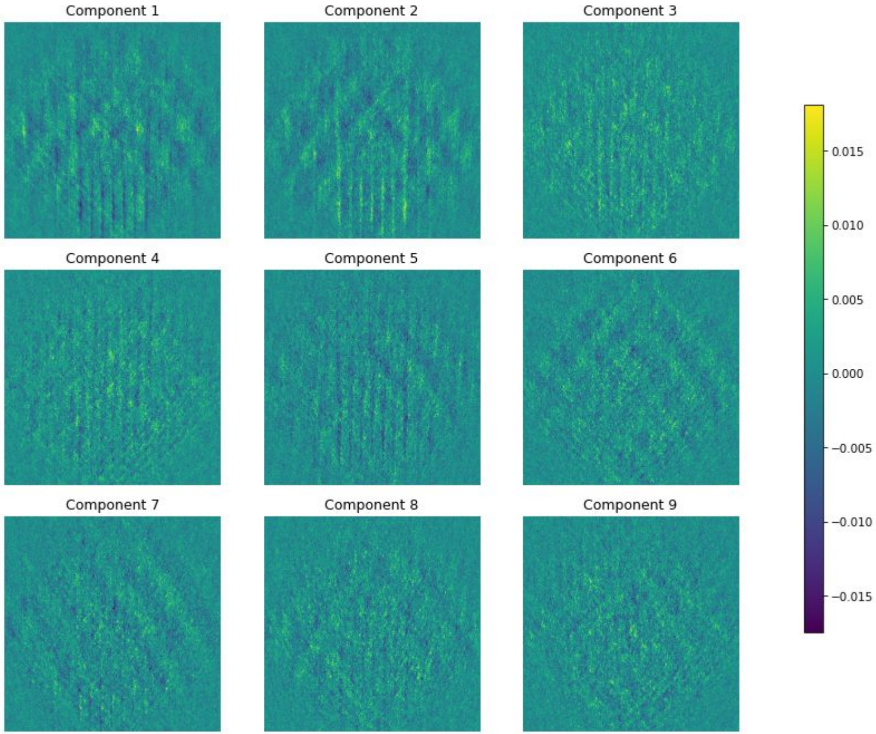
Dataset 3: little separation

Results: PCA Visualization

Principal components for Dataset 1 (high separation)



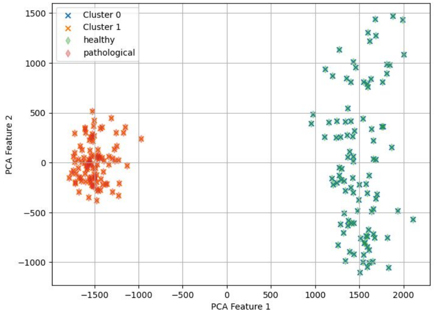
Principal components for Dataset 3 (low separation)

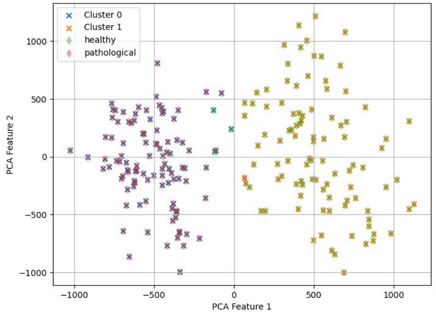


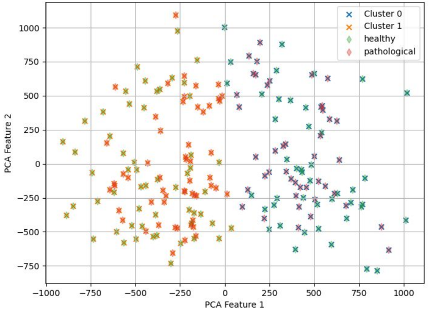
Results: Clustering in Frequency Space

Data1 | Data2 | Data3

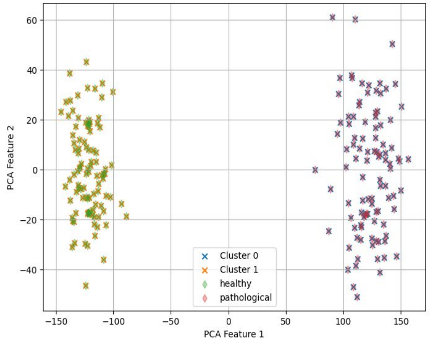
Spatial

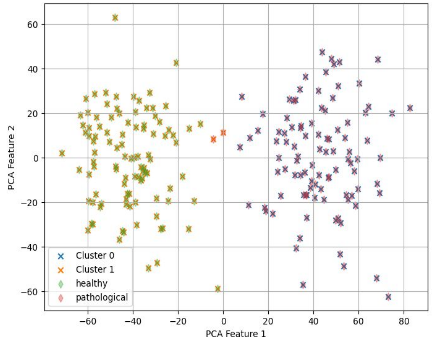


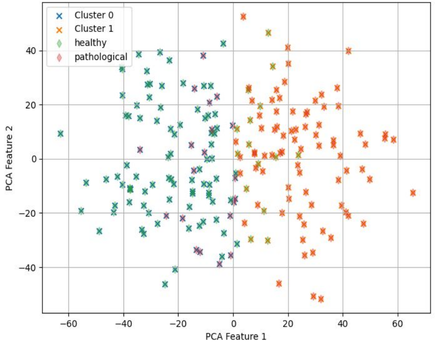




Fourier



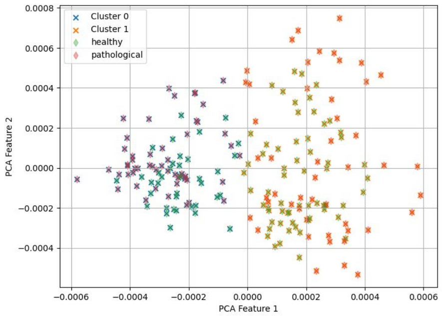


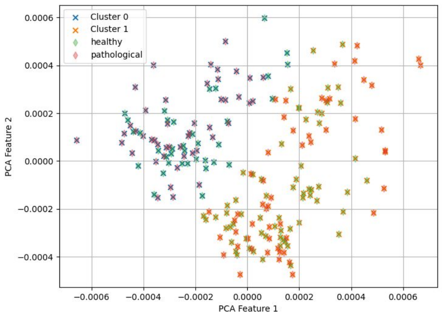


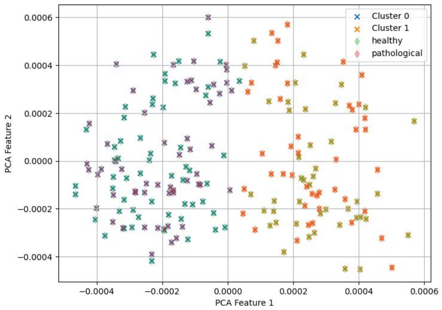
Results: Clustering in Frequency Space with Normalization

Data1 | Data2 | Data3

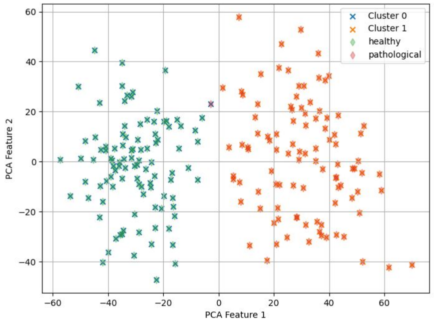
Spatial Normalized

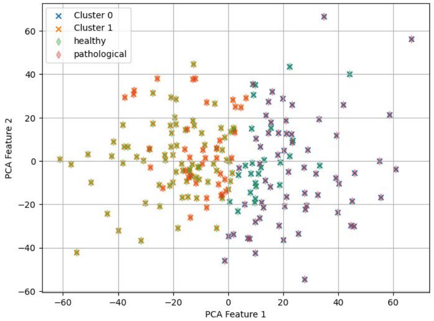


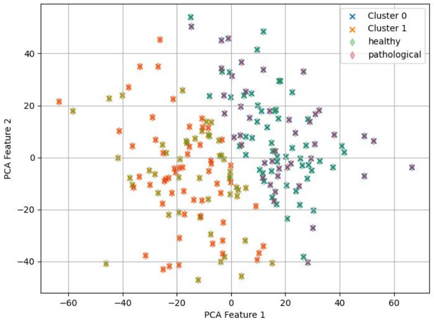




Fourier Normalized

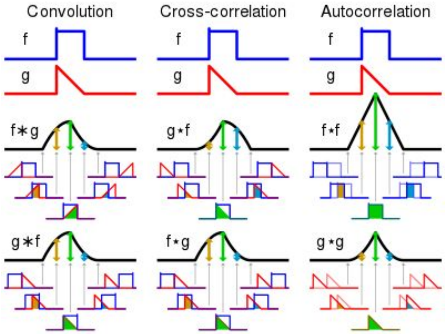






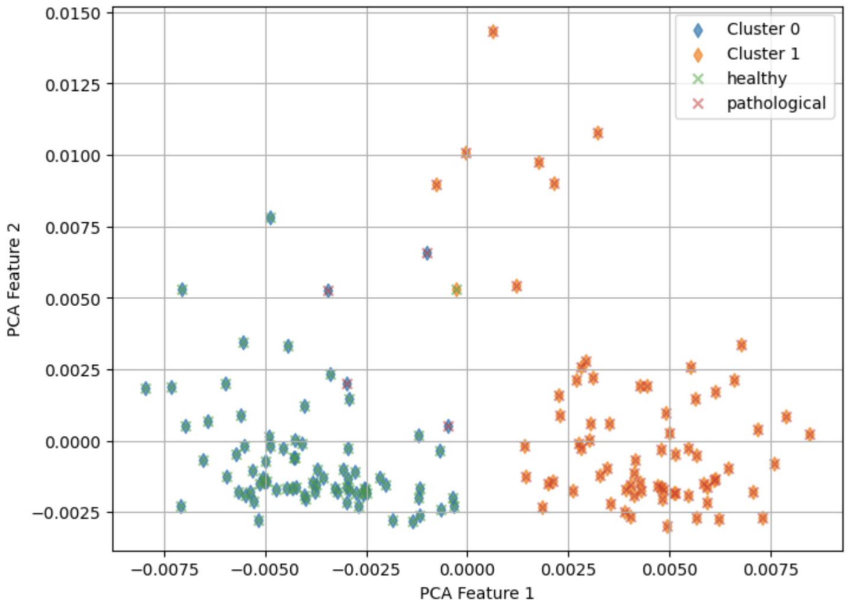
Autocorrelation in Frequency Domain Experiment

Autocorrelation of a signal can represent the structure or “memory” of a signal which could provide additional information on magnitude and phase of fourier transform.



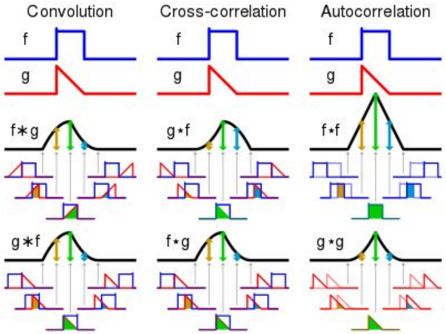
https://en.wikipedia.org/wiki/Autocorrelation

Data1 PSD using Normalized and mean-subtracted Diffraction



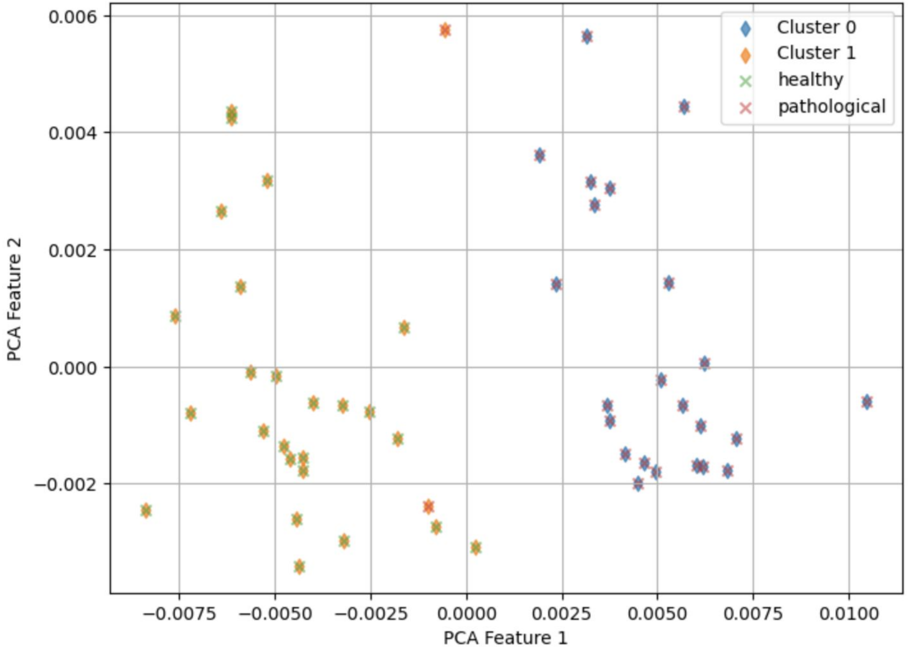
Autocorrelation in Frequency Domain Experiment

Autocorrelation of a signal can represent the structure or “memory” of a signal which could provide additional information on magnitude and phase of fourier transform.



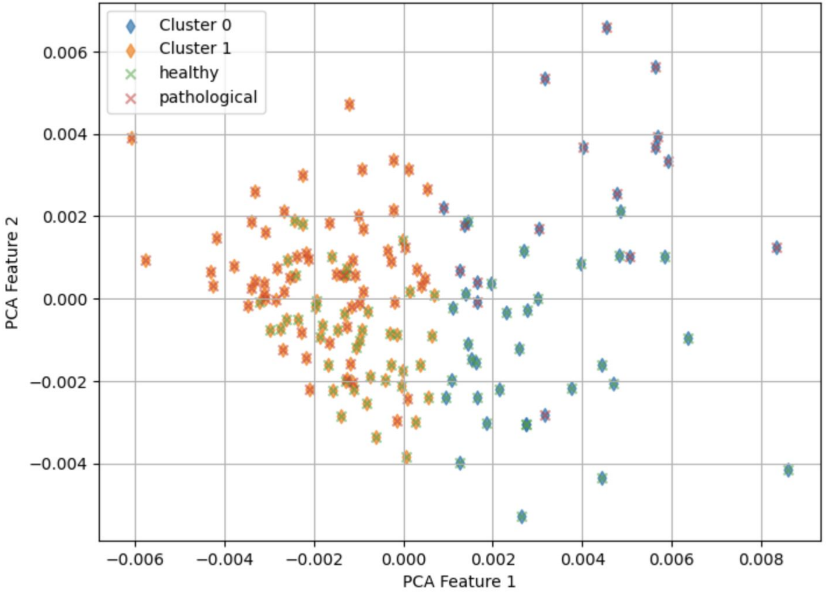
https://en.wikipedia.org/wiki/Autocorrelation

Data1 PSD TEST data using Normalized and mean-subtracted Diffraction

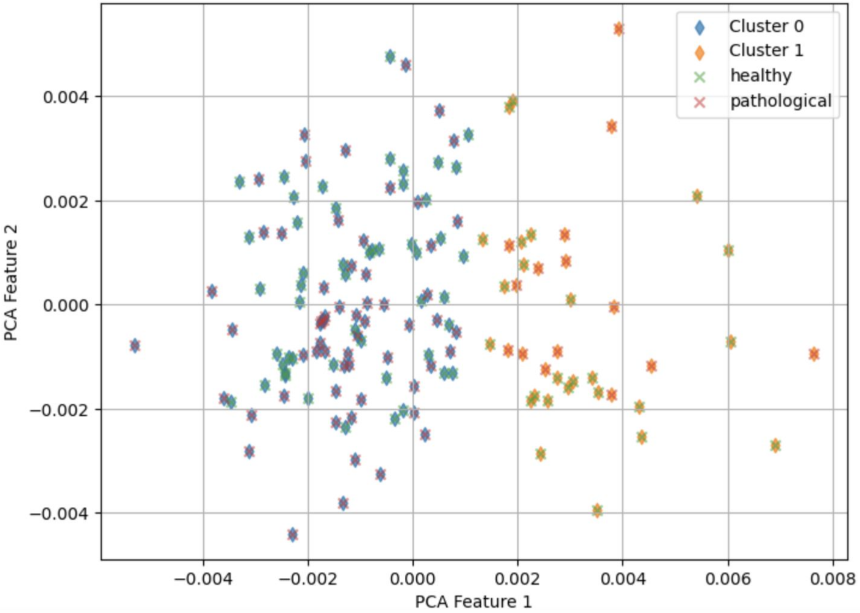


Autocorrelation in Frequency Domain Experiment

Data2 PSD using Normalized and mean-subtracted Diffraction



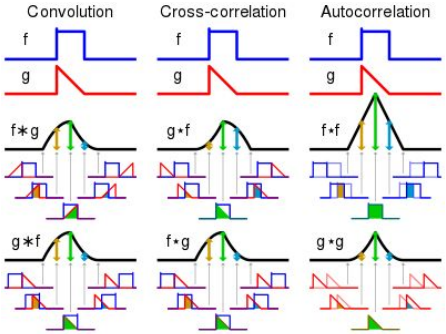
Data3 PSD using Normalized and mean-subtracted Diffraction



https://en.wikipedia.org/wiki/Autocorrelation

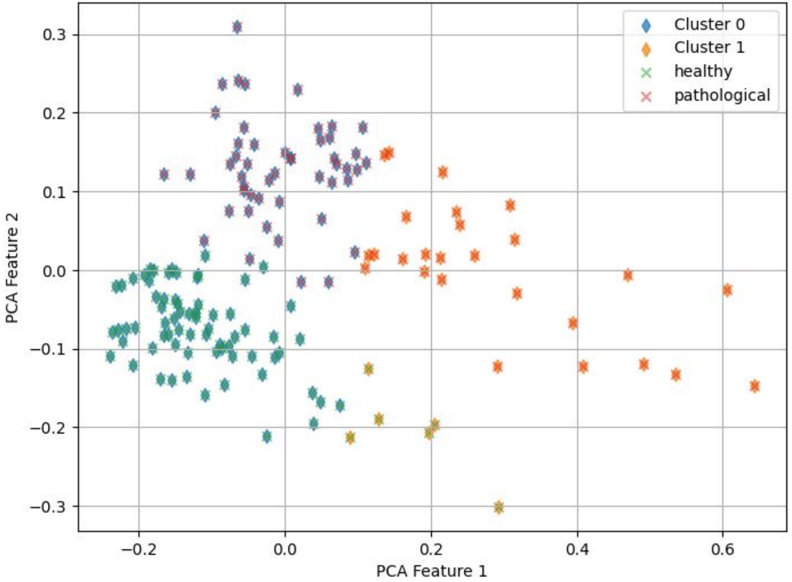
Autocorrelation in Frequency Domain Experiment

Autocorrelation of a signal can represent the structure or “memory” of a signal which could provide additional information on magnitude and phase of fourier transform.



https://en.wikipedia.org/wiki/Autocorrelation

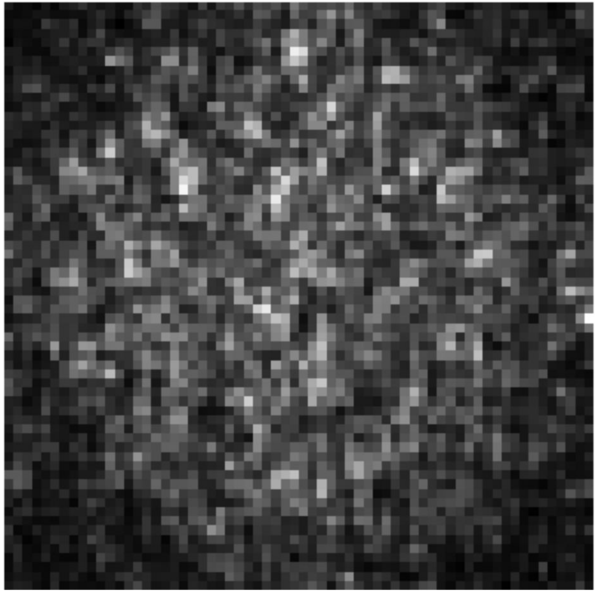
Data1 Autoencoder of Fourier Space
With Normalized Diffraction
(1st Attempt, needs refinement)



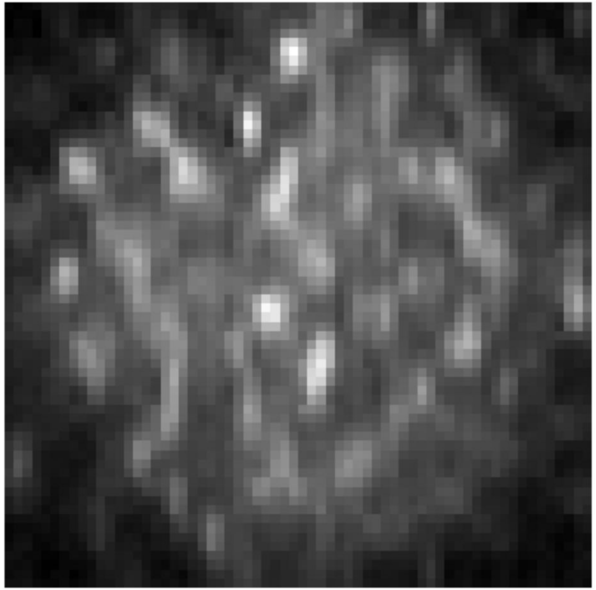
Results: VAE

Set 1 - **not** normalized:

Input



Reconstructed

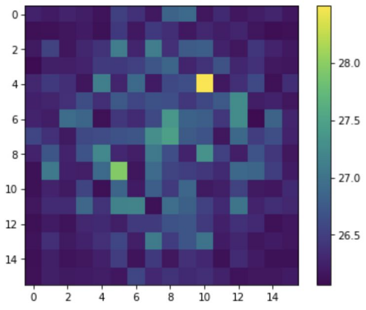


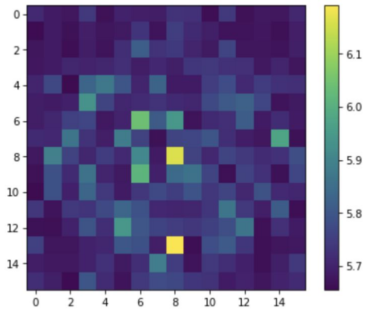
16x16 latent space

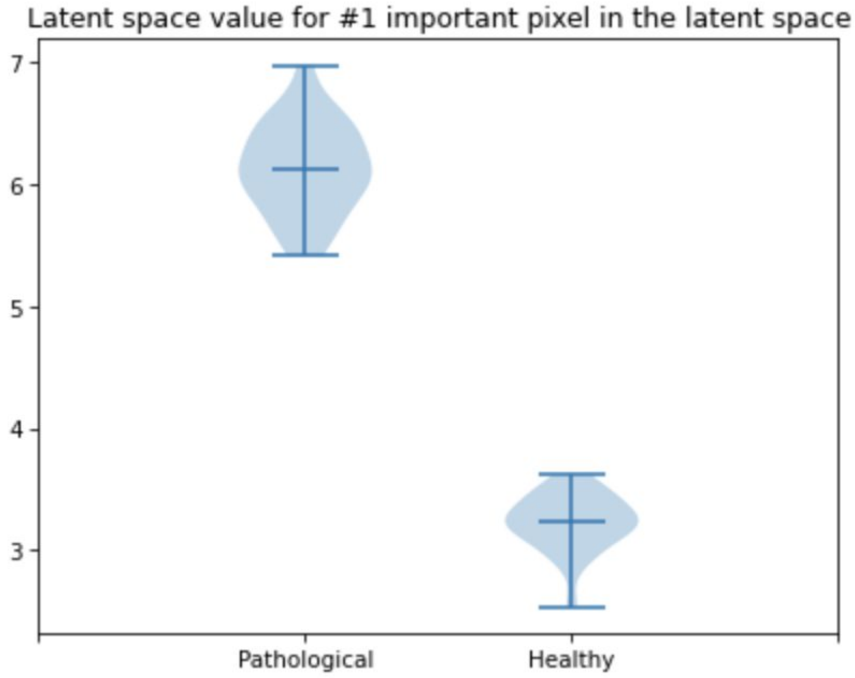
15 epochs

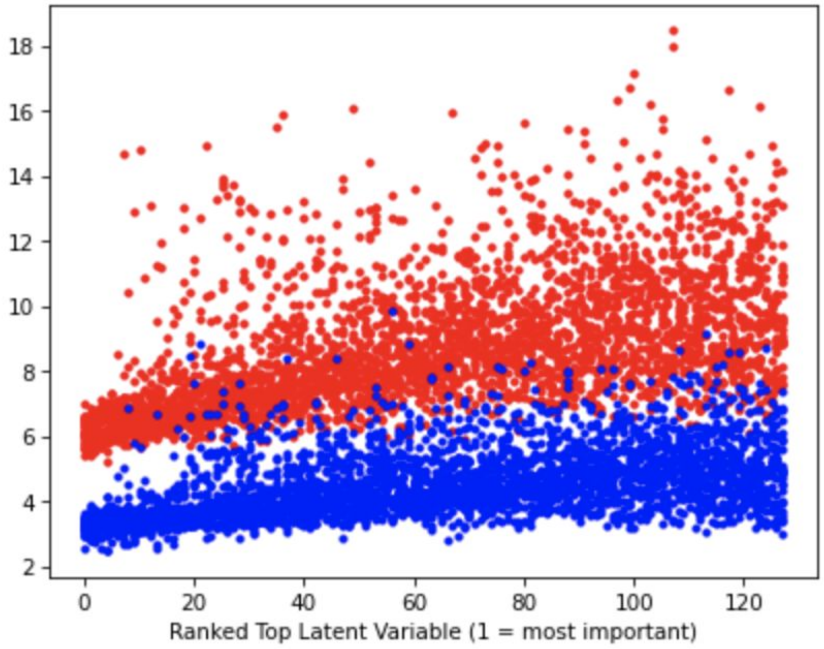
Results: VAE

Set 1 - **not** normalized:



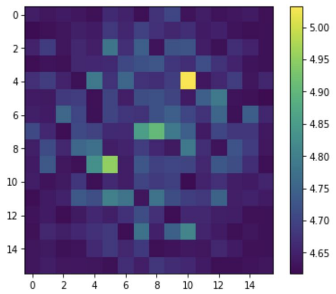


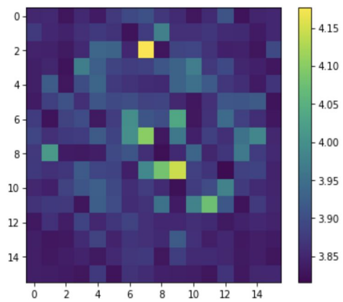


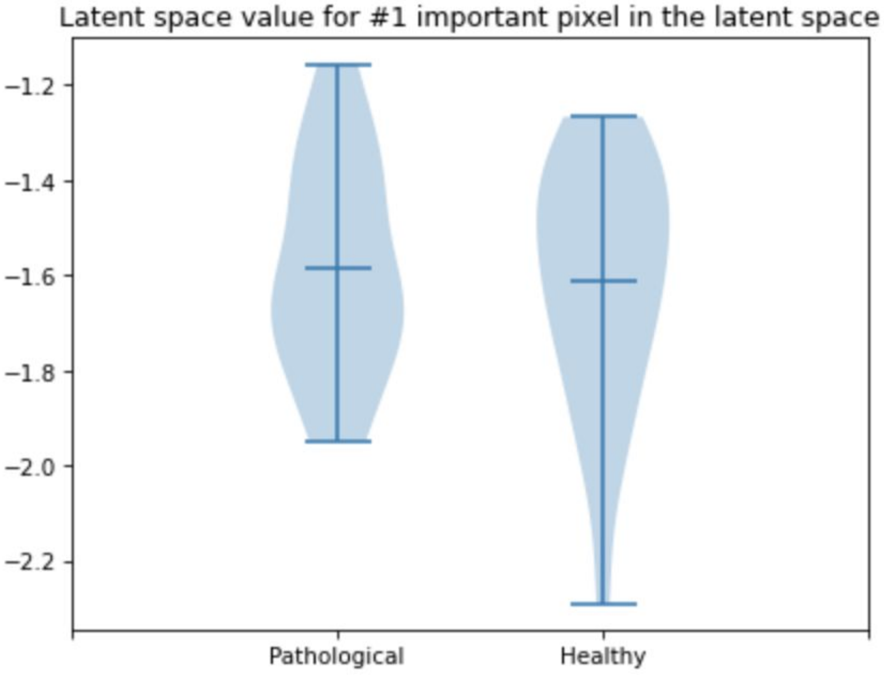


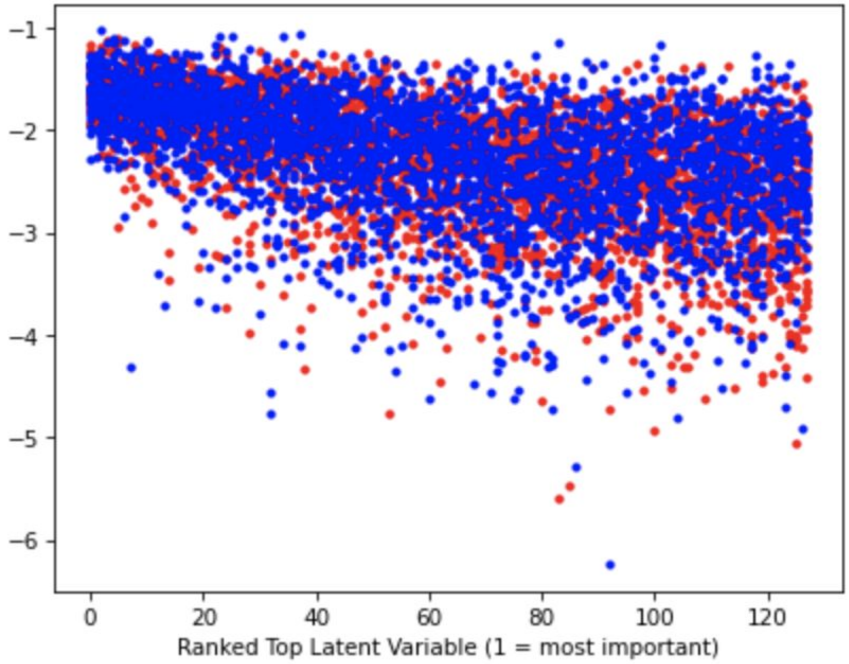
Results: VAE

Set 1 - **normalized:**



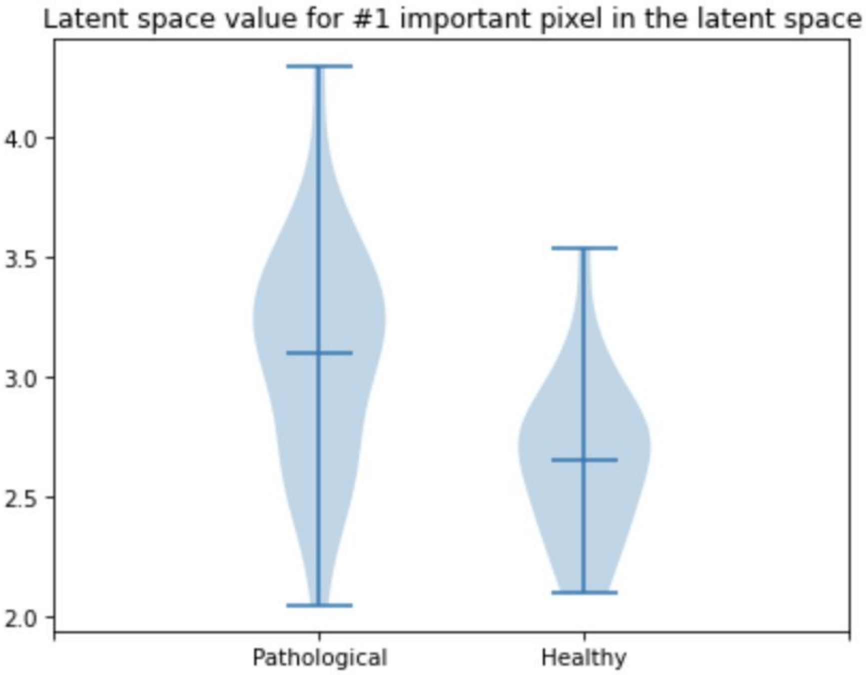


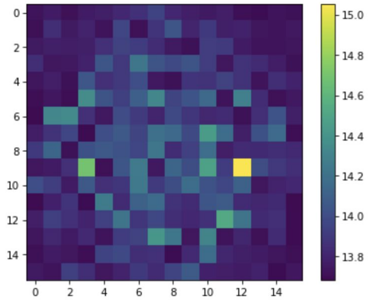


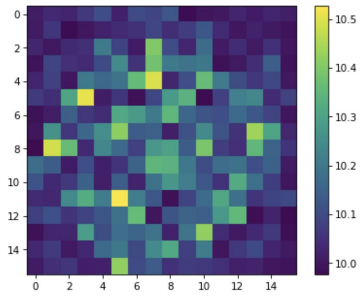


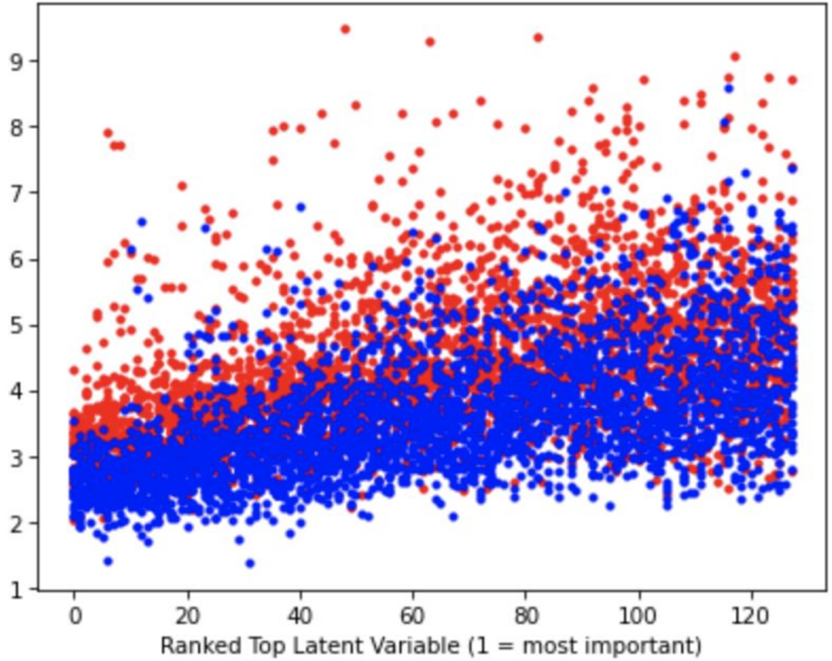
Results: VAE

Set 2, not normalized



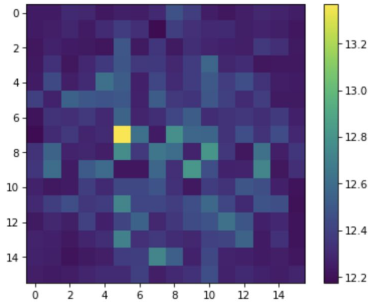


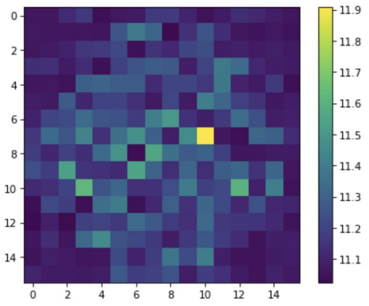


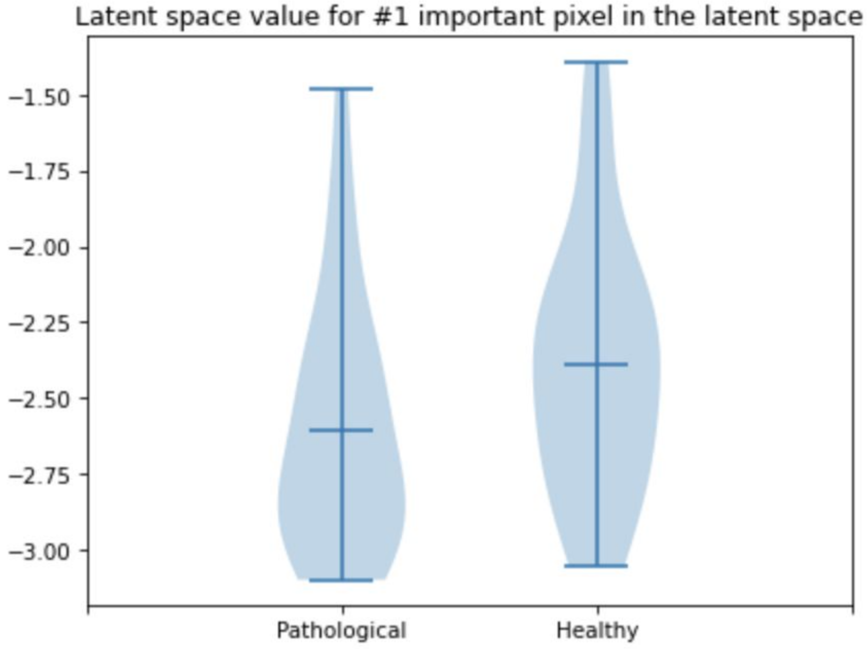


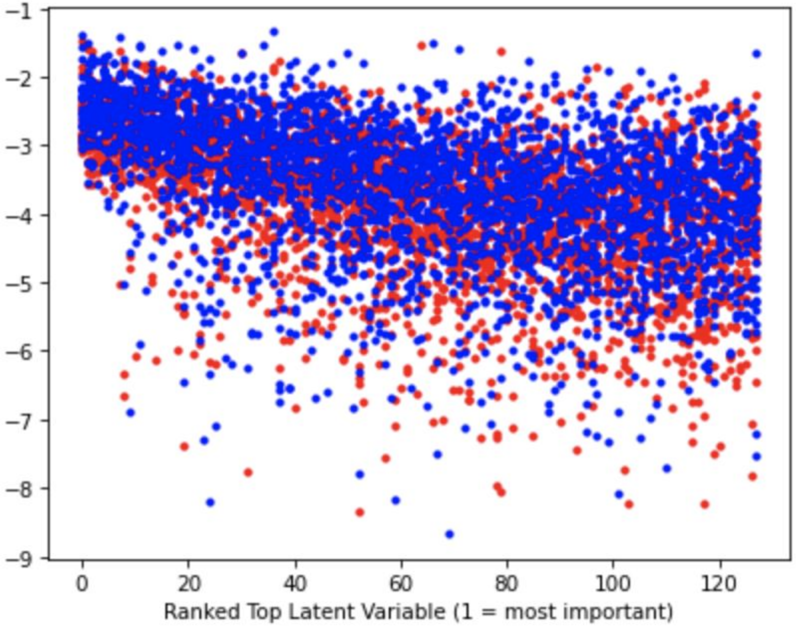
Results: VAE

Set 3, not normalized





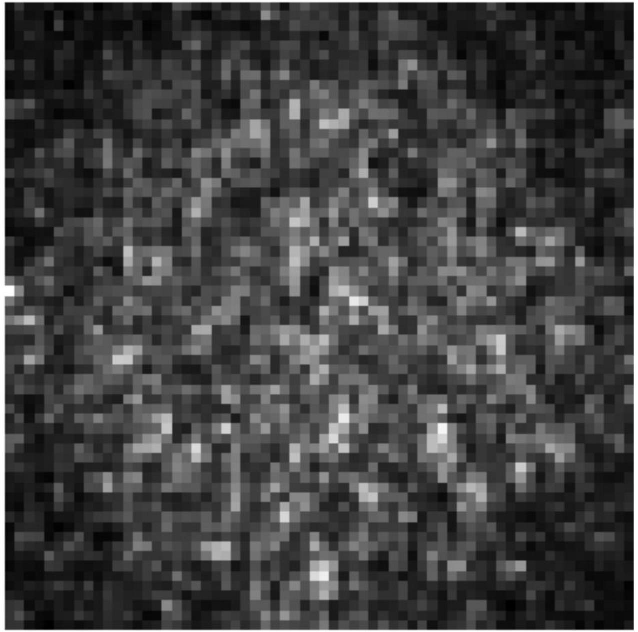




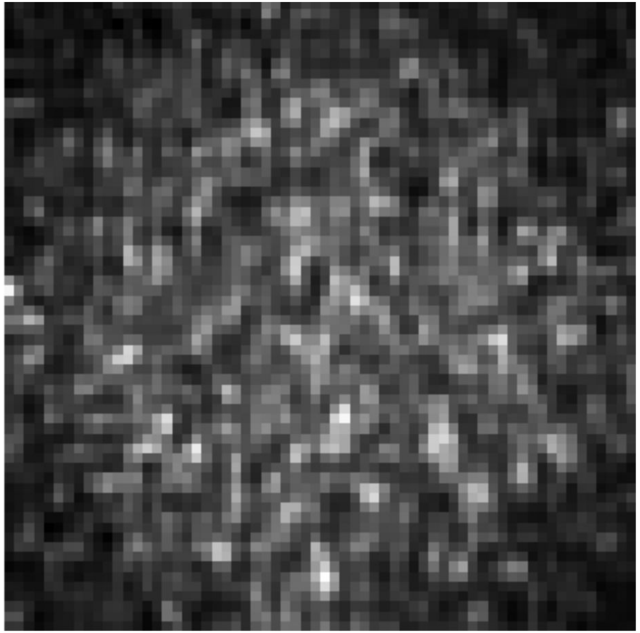
Results: VAE

Set 1, normalized, **larger latent** space (6x16x16)**:**

Input



Reconstructed

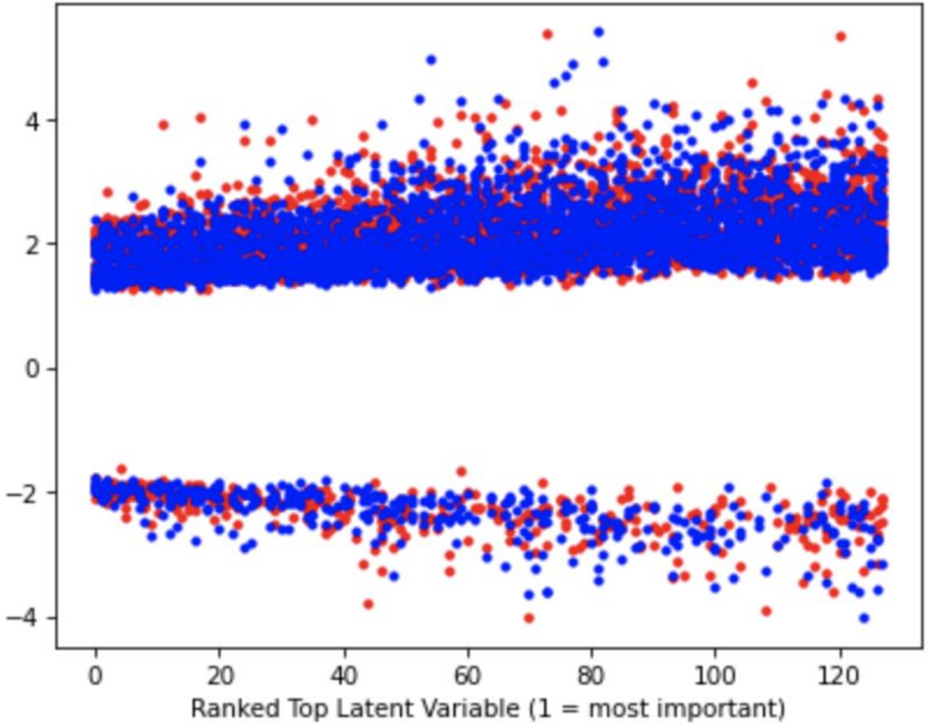


6x16x16 latent space

15 epochs

Results: VAE

Set 1, normalized, larger latent space (6x16x16):

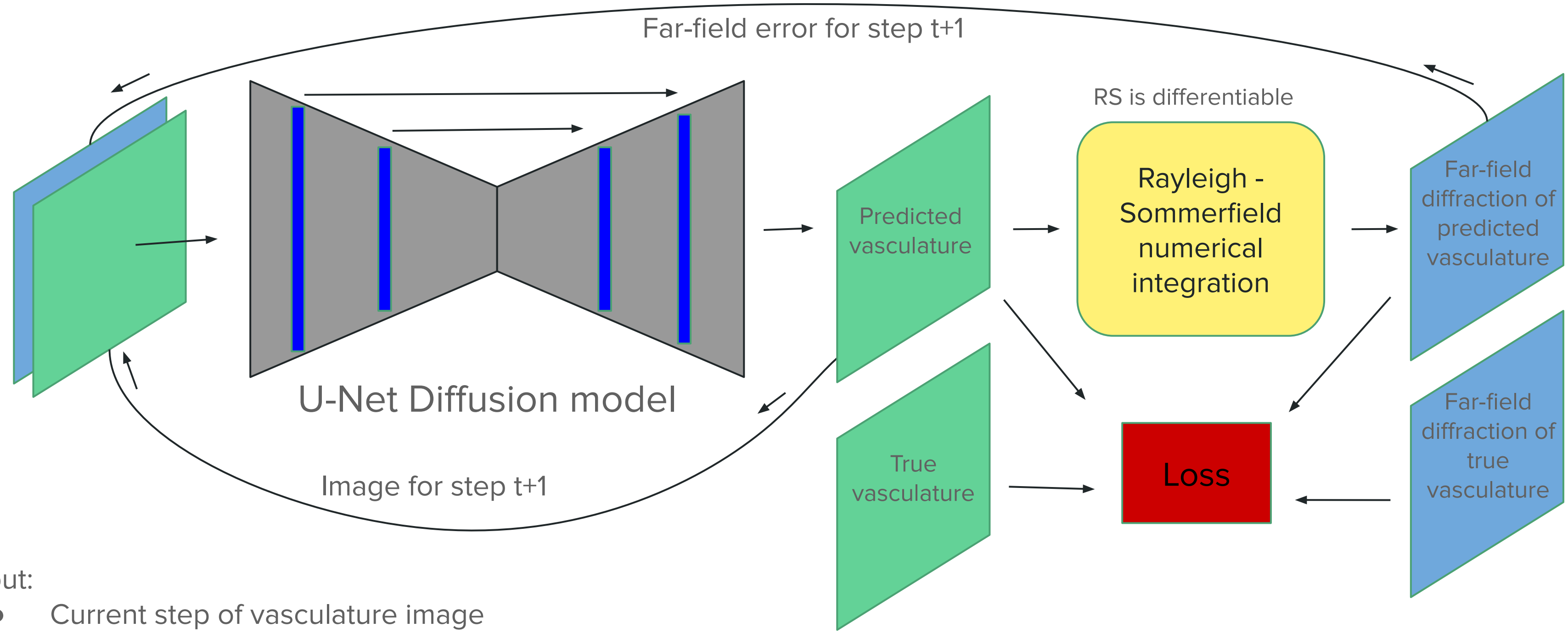


Conclusions + Next Steps

- We used several unsupervised methods to see if differences in the far-field speckle pattern of vasculature with different branching probabilities could be identified
- Some techniques, like t-SNE and k-Means can be used to identify different clusters of vein branching if the difference between “healthy” and “pathological” branching is large enough (and depending on normalization)
- Variational Autoencoders may underperform in our experiments due to the limited amount of data we had access to. Further work is needed to optimal parameters (number of encoding layers, latent variables) and latent variable characterization.

Possible next exploration: Conditional Diffusion Model

Can we generate the true vasculature using the far field speckle pattern?



Input:

- Current step of vasculature image
- Error of noisy vasculature speckle pattern and true vasculature speckle pattern

We think this might work because although we do not have any information about the phase of the light, we do have a good amount of prior knowledge in that we know what vasculature is supposed to look like.