# Mid Project Demo

Samuel Jackson

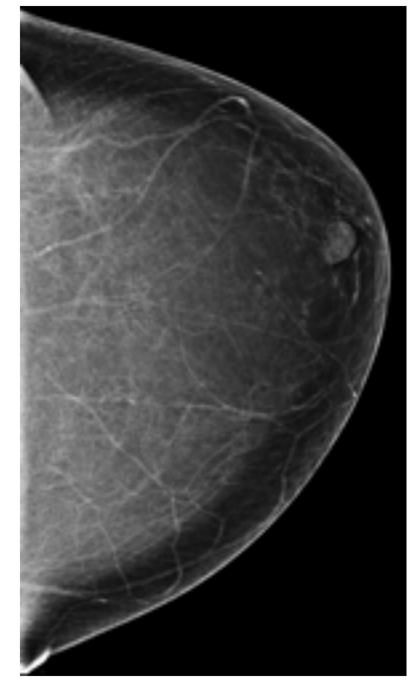
# Visualisation and Topological Aspects of Higher Dimensional Data

# In a Nutshell

- Project aims to examine the effects of mapping a **high** dimensional space to **low** dimensional one
- Specifically, to examine mapping the **feature space** of real and synthetic mammograms to a lower dimensional representation.

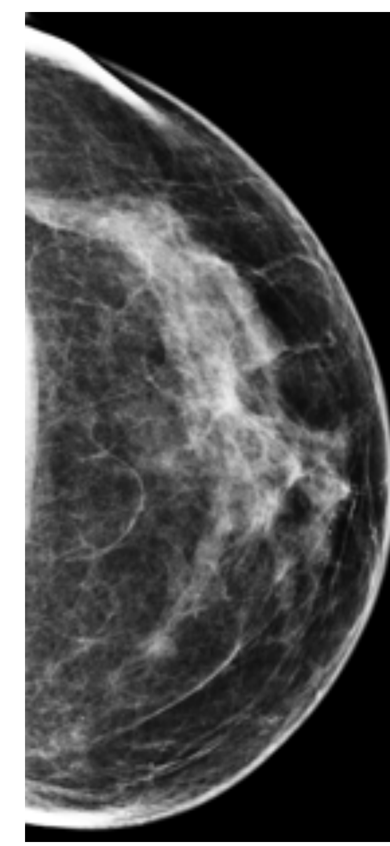
# Background

# Mammogram Image Analysis



- Mammogram: X-ray image of the human breast

- Used for early detection of breast cancer

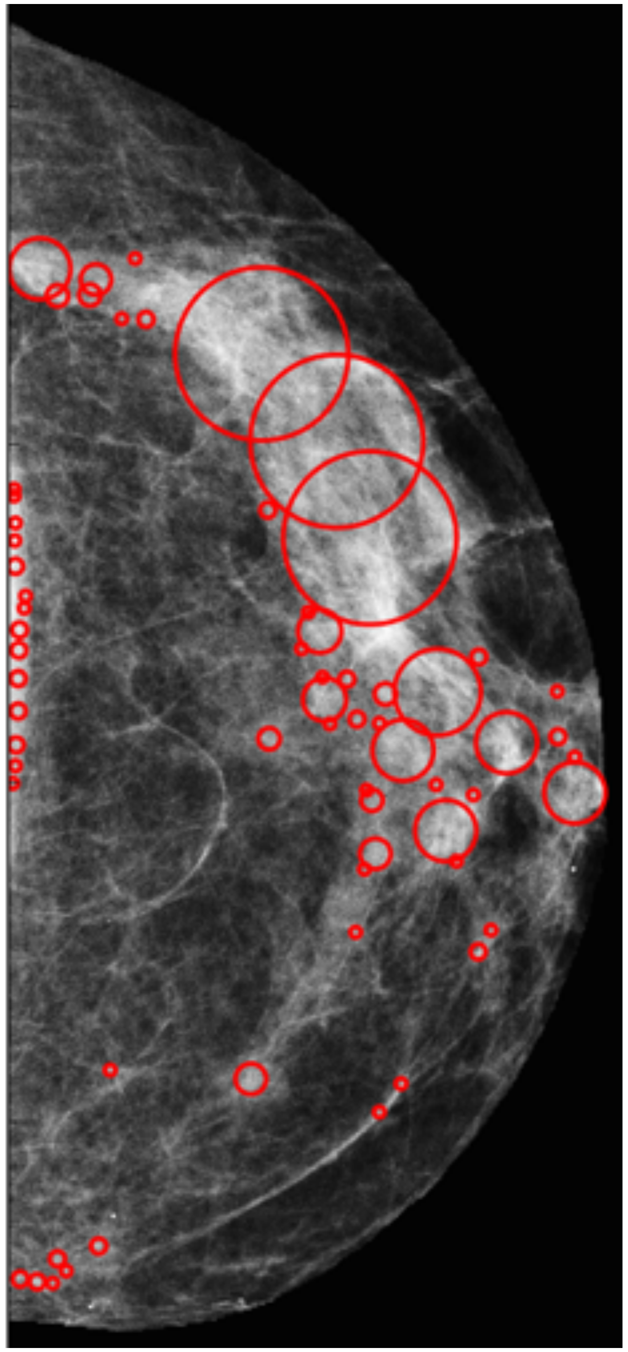


- Risk classed according to the BI-RADS scoring system

- Generally: denser breast == higher risk

# Mammogram Image Analysis (Contd.)

- CADx often focuses on using features
  - Shape features
  - Texture features
  - Intensity features
- From features we can calculate statistics which can characterise tissue of the breast



# Problem

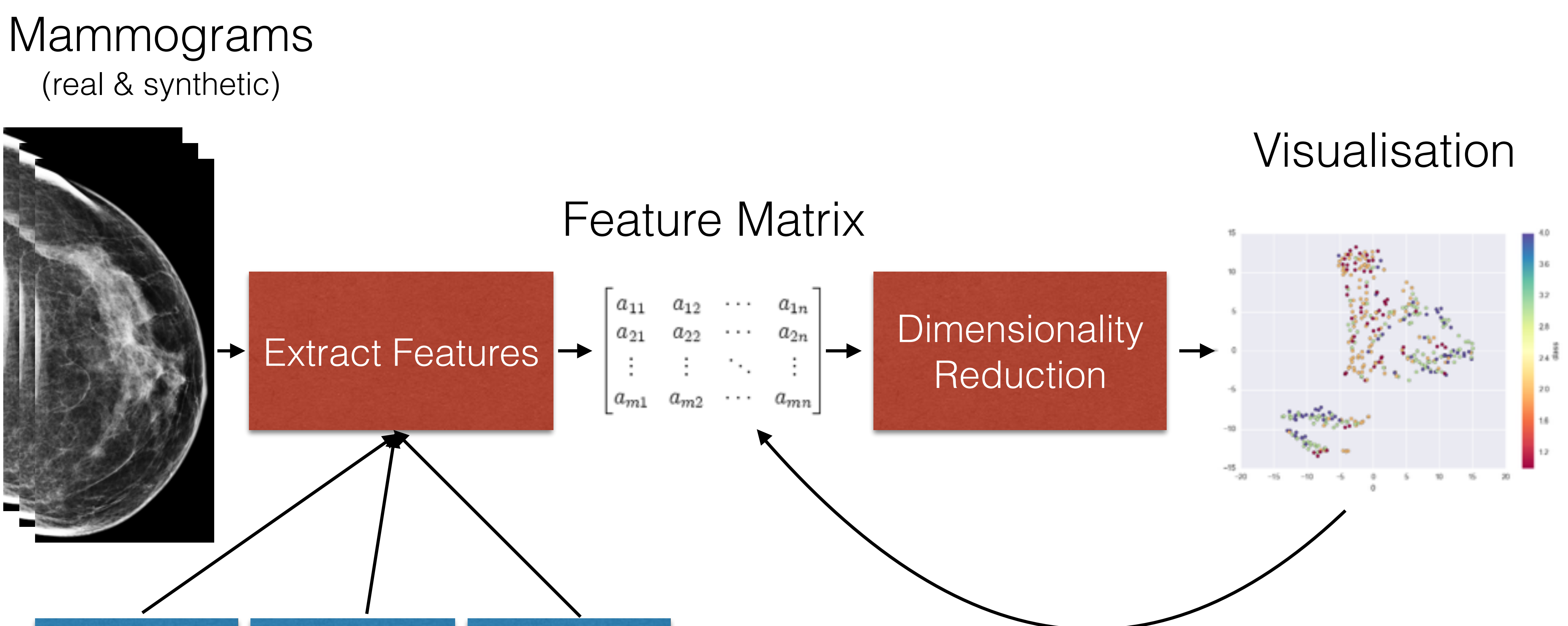
- Typically many features may be detected
- This leads to a high dimensional feature space
- This causes issues:
  - Increased processing time
  - Curse of dimensionality
- Solution: Reduce the dimensionality! (Possibly)

# Project

# Aims

- Examine the effect of dimensionality reduction on the feature space of mammograms
- Compare the resultant mapping with that of synthetic mammogram models
- Examine the properties of the higher dimensional space

# The General Idea



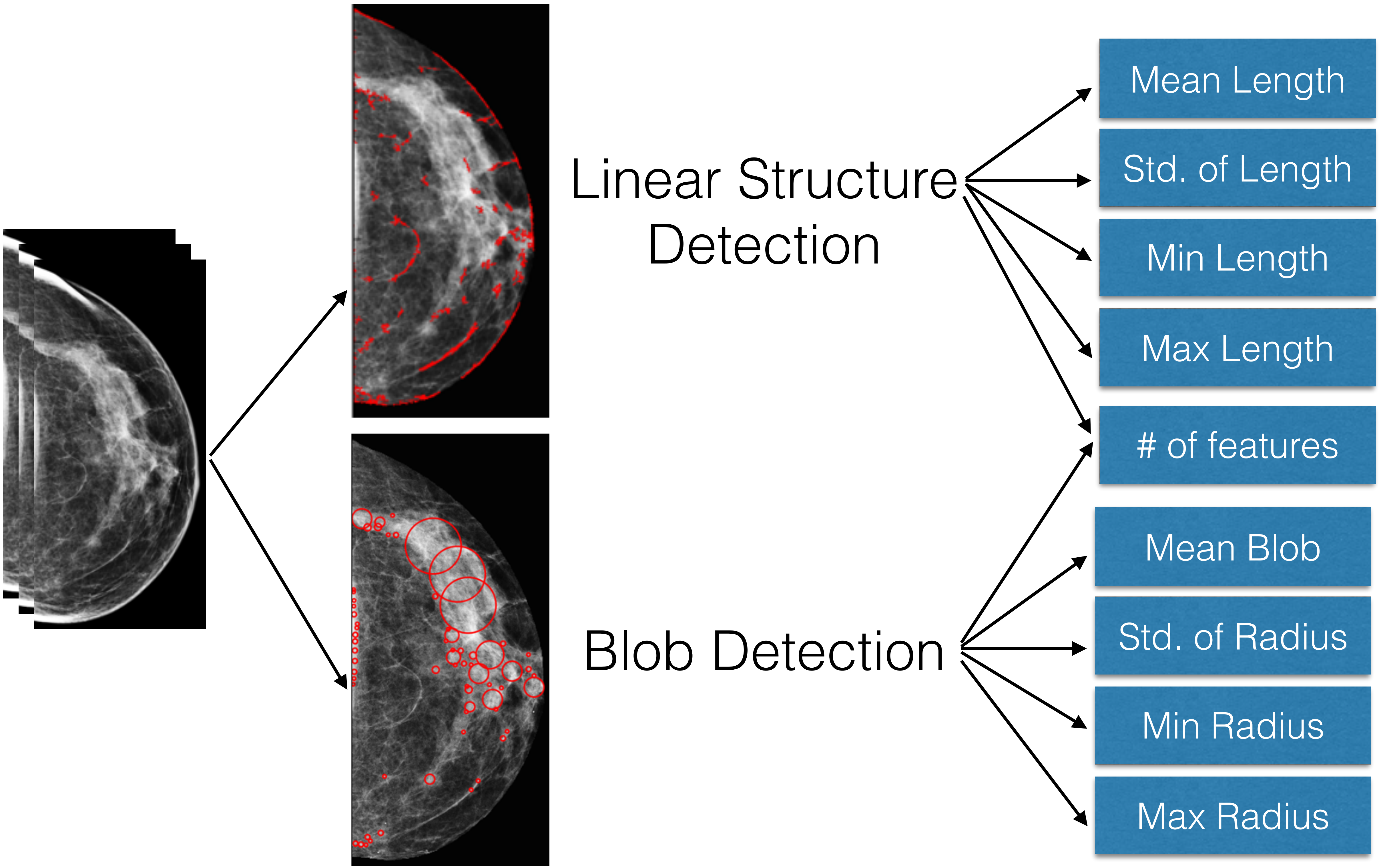

Features

What can we infer about the structure of the feature space?

What can this tell us about mammogram analysis?

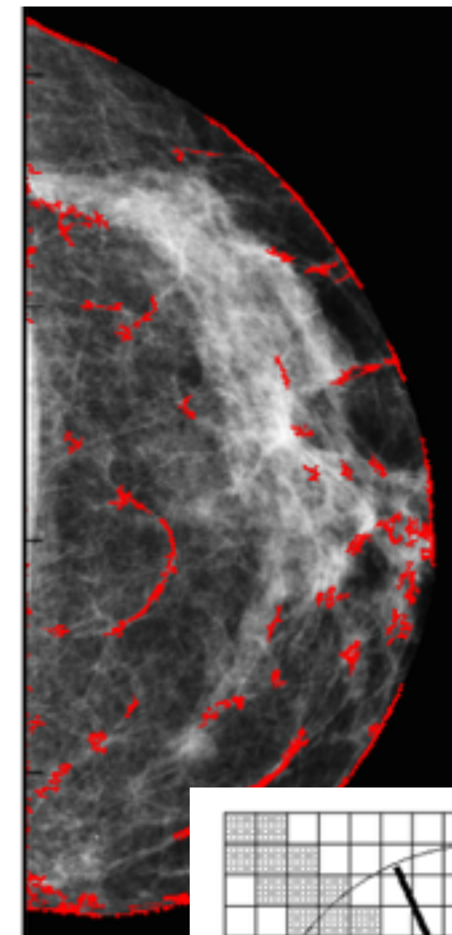
What are the differences between the mappings of real & synthetic mammograms?

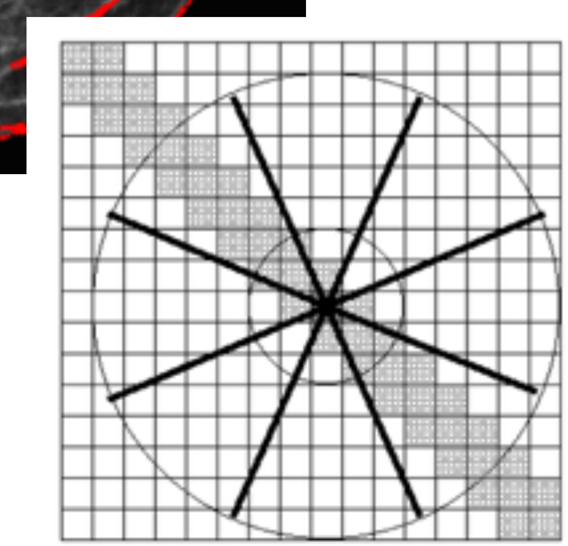
# Feature Extraction



# Linear Structure

- Uses an orientated bins feature[1].
- Filter image with function that computes the difference between two opposing bins and total neighbourhood
- Apply non-maximal suppression and Gaussian filter to strengthen responses
- Apply morphological closing for better connectivity

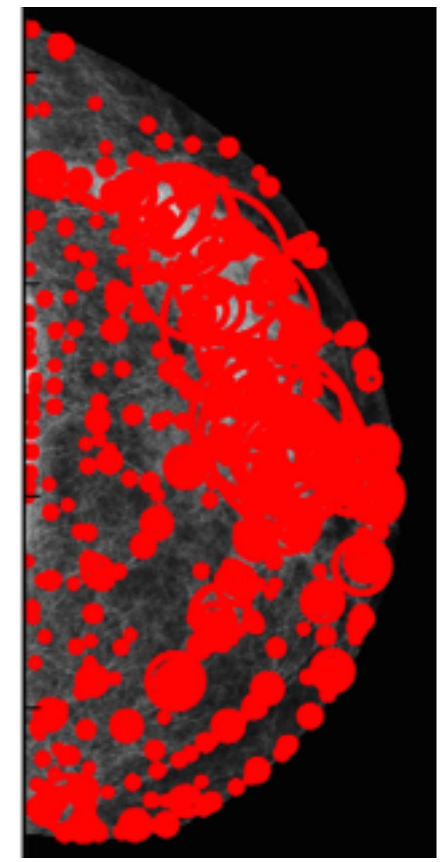


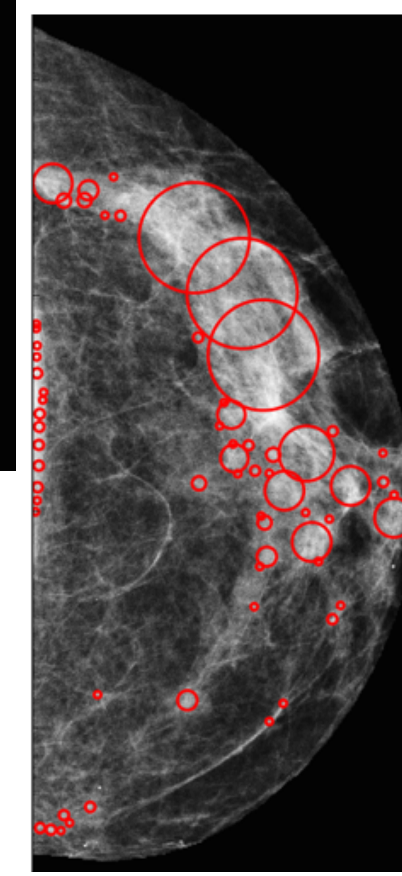


1. Based method suggested in: Zwiggelaar *et al.* "Finding Orientated Line Patterns in Digital Mammographic Images." *BMVC.* 1996.

# Blob Detection

- Uses a LoG pyramid to detect points of high intensity[1].
- Detect areas of high intensity over 10 scales
- Width of Gaussian used as radius estimate. Centre of peak is centre point of blob.
- Blobs merged and thresholded to remove false positives

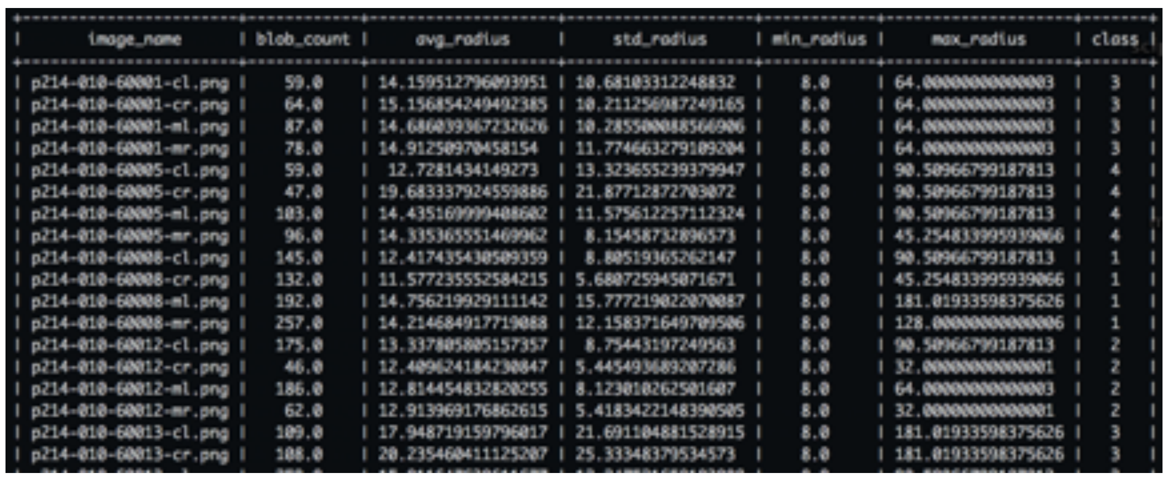




1. Based method suggested in: Chen, Zhili, et al. "A multiscale blob representation of mammographic parenchymal patterns and mammographic risk assessment." *Computer Analysis of Images and Patterns*. Springer Berlin Heidelberg, 2013.

# Dimensionality Reduction

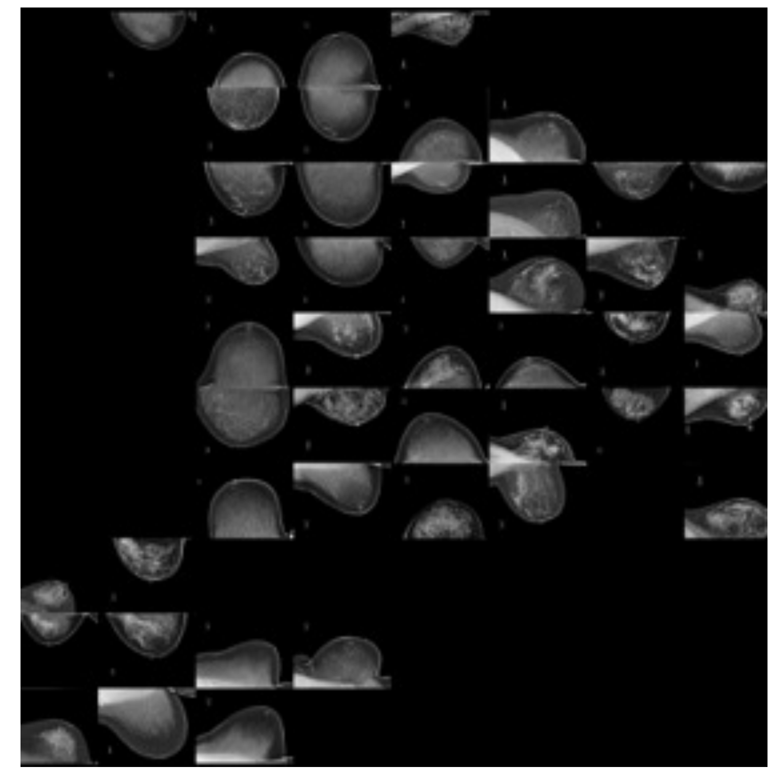
Output of Feature Detection



| image_name | blob_count | avg_radius | std_radius | min_radius | max_radius | class |
|---|---|---|---|---|---|---|
| p214-010-60001-cl.png | 59.0 | 14.159512796093951 | 10.68103312248832 | 8.0 | 64.0000000000003 | 3 |
| p214-010-60001-cr.png | 64.0 | 15.156854249492385 | 10.211256987249165 | 8.0 | 64.0000000000003 | 3 |
| p214-010-60001-ml.png | 87.0 | 14.686039367232626 | 10.285500088566906 | 8.0 | 64.0000000000003 | 3 |
| p214-010-60001-mr.png | 78.0 | 14.91250970458154 | 11.774663279109204 | 8.0 | 64.0000000000003 | 3 |
| p214-010-60005-cl.png | 59.0 | 12.7281434149273 | 13.323655239379947 | 8.0 | 90.5096679918781З | 4 |
| p214-010-60005-cr.png | 47.0 | 19.683337924559886 | 21.87712872703072 | 8.0 | 90.5096679918781З | 4 |
| p214-010-60005-ml.png | 103.0 | 14.435169999408602 | 11.575612257112324 | 8.0 | 90.5096679918781З | 4 |
| p214-010-60005-mr.png | 96.0 | 14.335365551469962 | 8.15458732896573 | 8.0 | 45.25483399593906б | 4 |
| p214-010-60008-cl.png | 145.0 | 12.417435430509359 | 8.80519365262147 | 8.0 | 90.5096679918781З | 1 |
| p214-010-60008-cr.png | 132.0 | 11.577235552584215 | 5.680725945071671 | 8.0 | 45.25483399593906б | 1 |
| p214-010-60008-ml.png | 192.0 | 14.756219929111142 | 15.777219022070087 | 8.0 | 181.01933598375626 | 1 |
| p214-010-60008-mr.png | 257.0 | 14.214684917719088 | 12.158371649709506 | 8.0 | 128.0000000000006 | 1 |
| p214-010-60012-cl.png | 175.0 | 13.337805805157357 | 8.75443197249563 | 8.0 | 90.5096679918781З | 2 |
| p214-010-60012-cr.png | 46.0 | 12.409624184230847 | 5.445493689207286 | 8.0 | 32.0000000000001 | 2 |
| p214-010-60012-ml.png | 186.0 | 12.814454832820255 | 8.123010262501607 | 8.0 | 64.0000000000003 | 2 |
| p214-010-60012-mr.png | 62.0 | 12.913969176862615 | 5.4183422148390505 | 8.0 | 32.0000000000001 | 2 |
| p214-010-60013-cl.png | 109.0 | 17.948719159796017 | 21.691104881528915 | 8.0 | 181.01933598375626 | 3 |
| p214-010-60013-cr.png | 108.0 | 20.235460411125207 | 25.33348379534573 | 8.0 | 181.01933598375626 | 3 |

t-SNE

Reduce to 2 dimensions

Plot of lower dimensional mapping

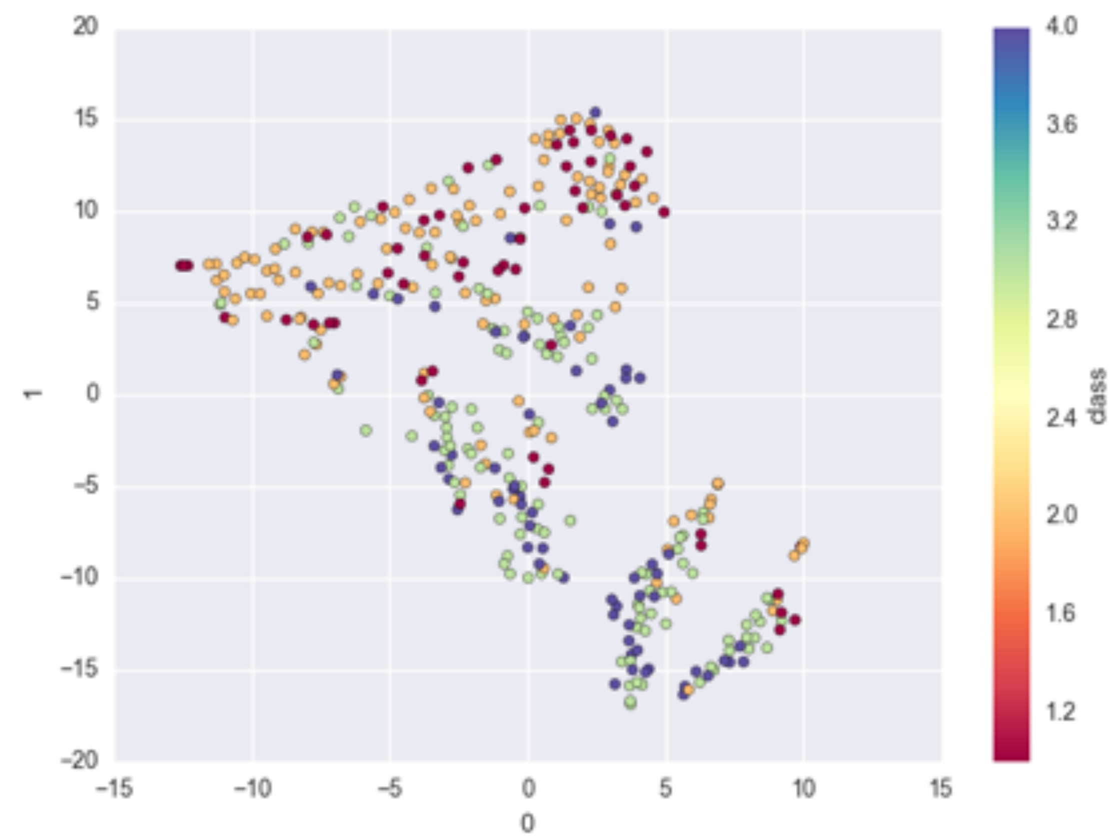
# Issues

# Feature Detection Issues

- Slow! Both shape features take ~30-40s per image. Total of ~3hrs to process the full dataset.
- Blobs: LoG detects more than just blobs: picks up lines too!
- Orientated bins: picks up lots of noise (e.g. blobs and along the breast edge). Also can have a hard time finding structure in low contrast images

# Dimensionality Reduction Issues

- The current mapping isn't fantastic.
- Shows some separation of high and low density classes
- Aiming for a clear(er) separation of risk classes



# Solution to Issues

- Adding multiprocessing reduced the time taken for a full run on the dataset - much more manageable!
- **Blobs**: LoG issues can probably be solved by thresholding out very small detections
- **Linear Structure**: discard shorter detections and perform morphological erosion on the surface of the breast. Contrast enhancement could possibly be used to enhance low response images
- **Dimensionality Reduction**: More statistics, better statistics. E.g. min/max radius don't seem to contribute much to mapping.

# Future Work

# Features

- Different types of feature: Texture and Grey Scale features and there combinations with existing features

- Different types of statistics from existing features: e.g. replace min/max radius with discrete # of “small”, “medium” and “large blobs

# Dimensionality Reduction

- We can try different algorithms and compare the resultant mappings

- If we have more features, try reducing to a larger number of dimensions => less information loss

- We can visualise this through higher dimensional plotting routines (e.g. parallel coordinates, andrews curves, radviz).

# Other Work

- Perform the same reduction on the synthetic mammograms
- Investigate topological structure of higher dimensional feature space. (e.g. Topological Data Analysis)

Any Questions?