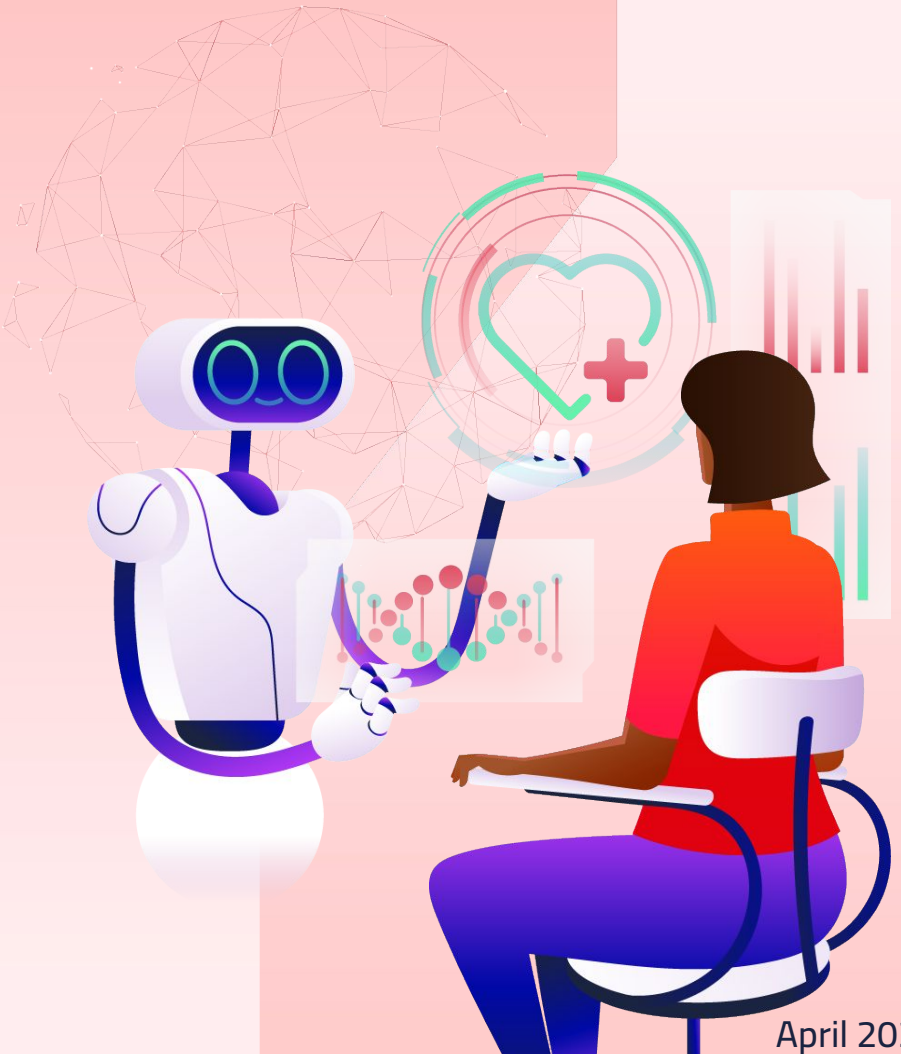
Screening Mammogram Early Breast Cancer Detection

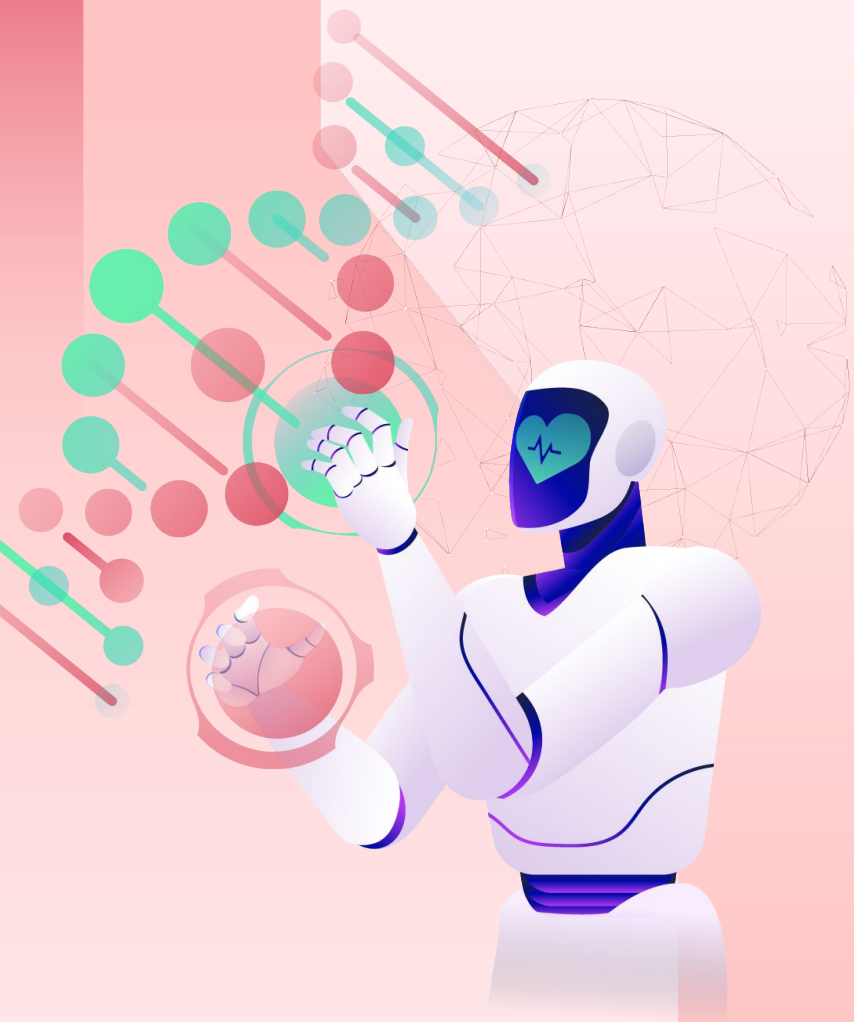


DSB 122
Elaine Chen

April 2024

Presentation Roadmap

- Objectives
- Exploratory Data Analysis
- Image Preview
- Image Folders Restructure and Image Preprocessing
- Modeling and Model Evaluation
- Prediction
- Conclusion
- Future Improvement
- Reference



01 Objectives

What is Mammography?

Mammograms are X-ray images of your breasts designed to detect cancers and other changes in breast tissue.

Regular screening mammograms are essential for most women as they offer the most reliable means of early breast cancer detection.

How common is breast cancer?

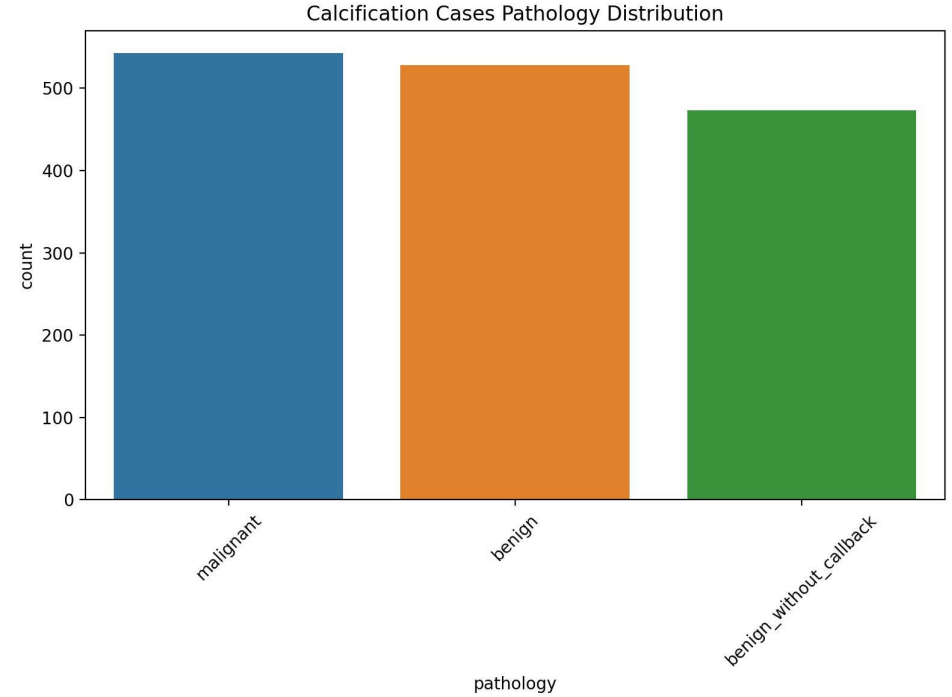
About 30% of all new female cancers each year (Source: American Cancer Society)

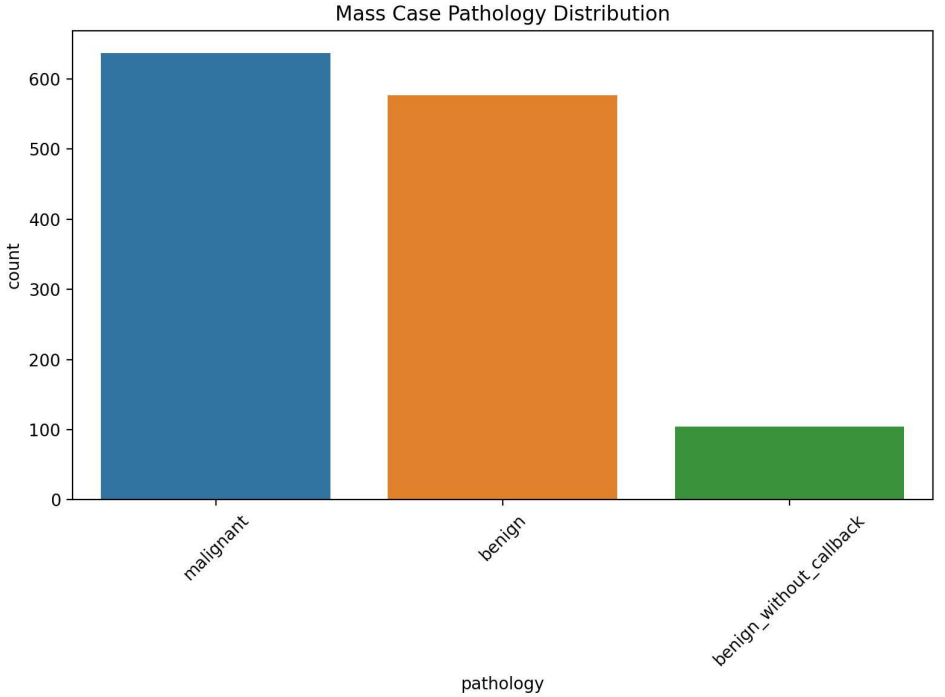
Objectives

Develop a computer-aided detection (CADe) and diagnosis (CADx) system to aid radiologists and physicians in interpreting mammograms and early detection on breast cancer

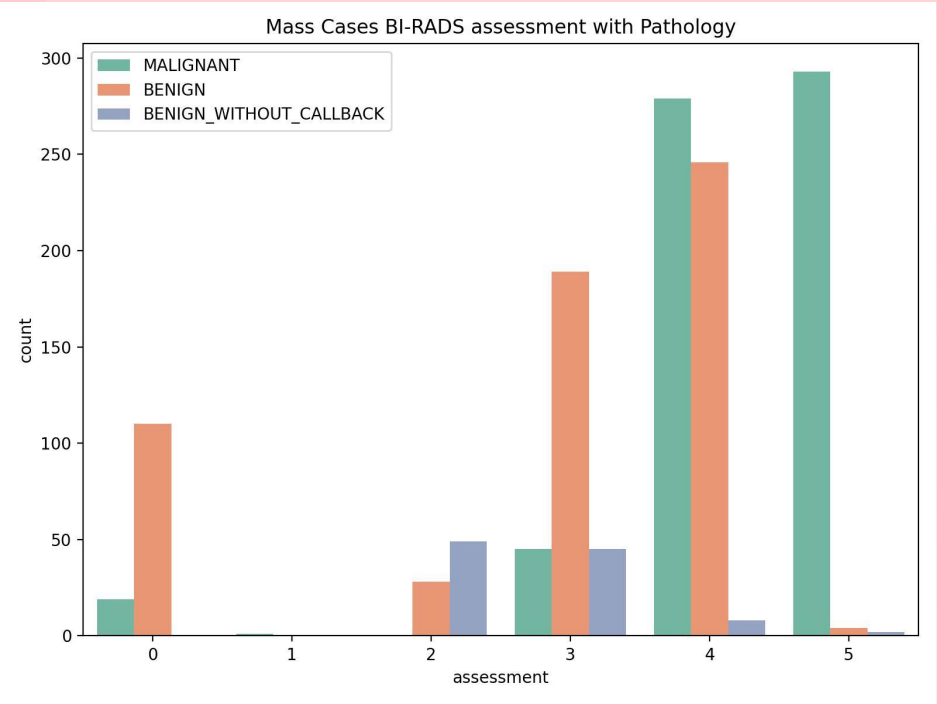
Reduce false-positive rates and support facilitate decision evaluation

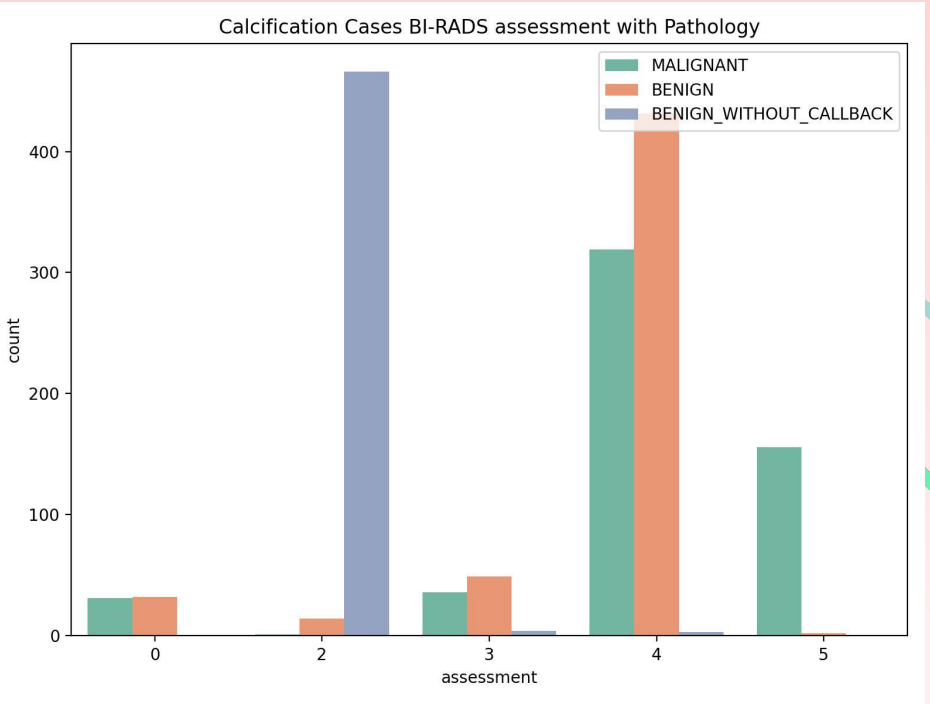
Exploratory Data Analysis



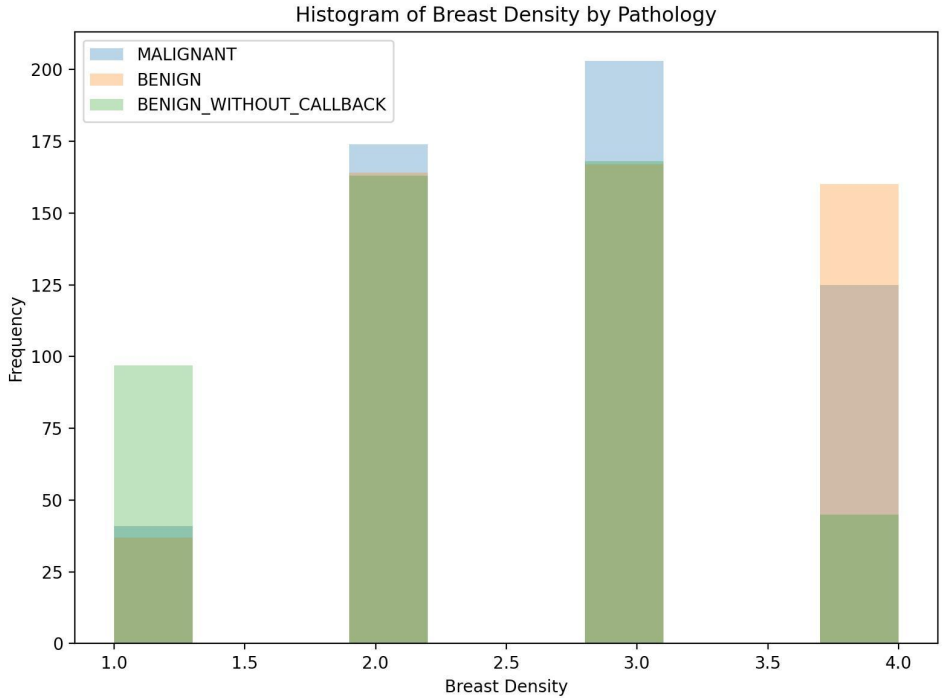


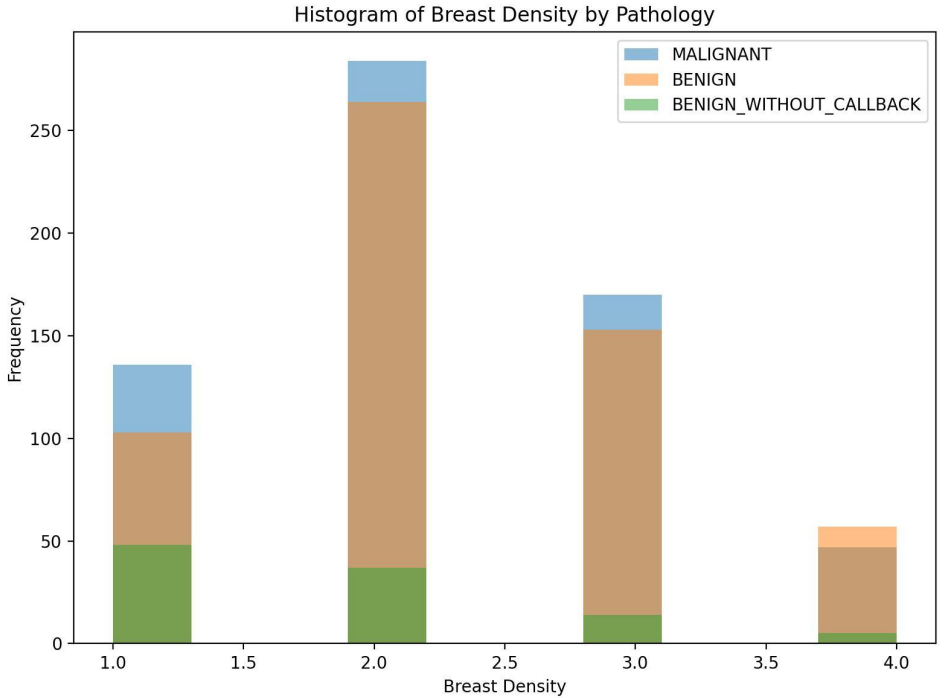
Exploratory Data Analysis



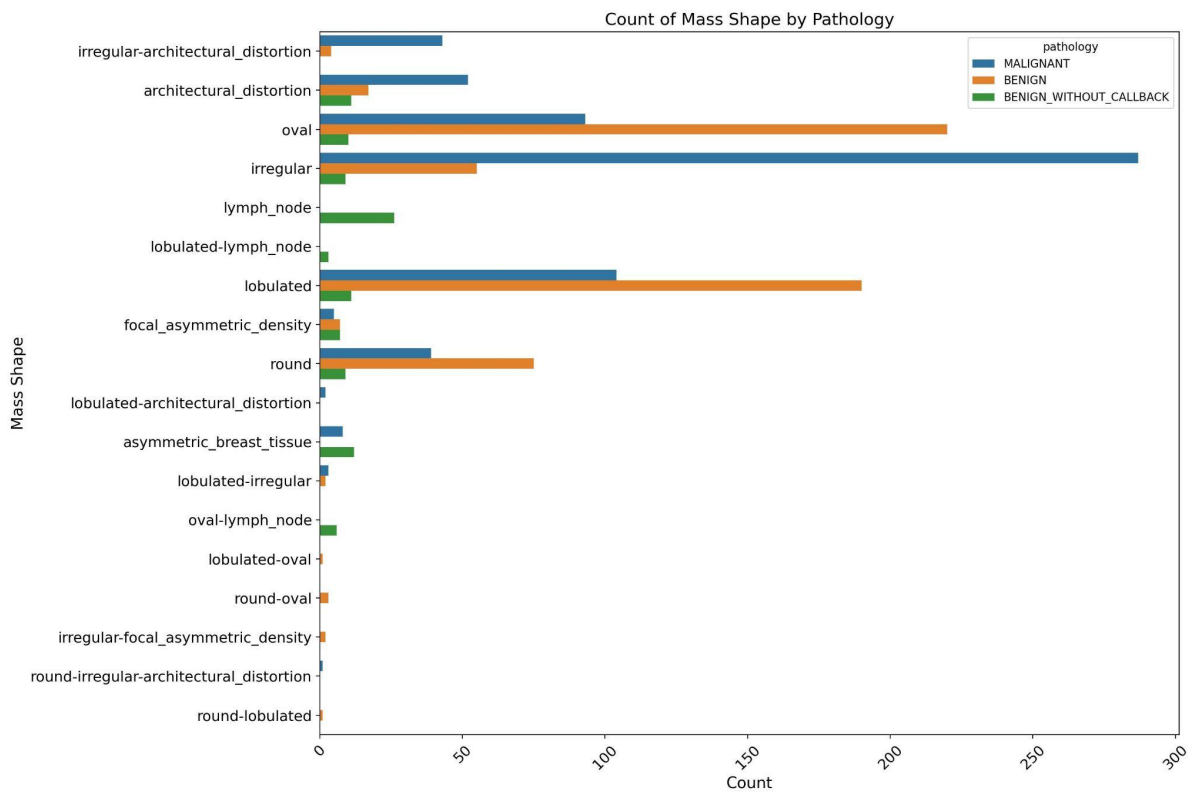


Exploratory Data Analysis

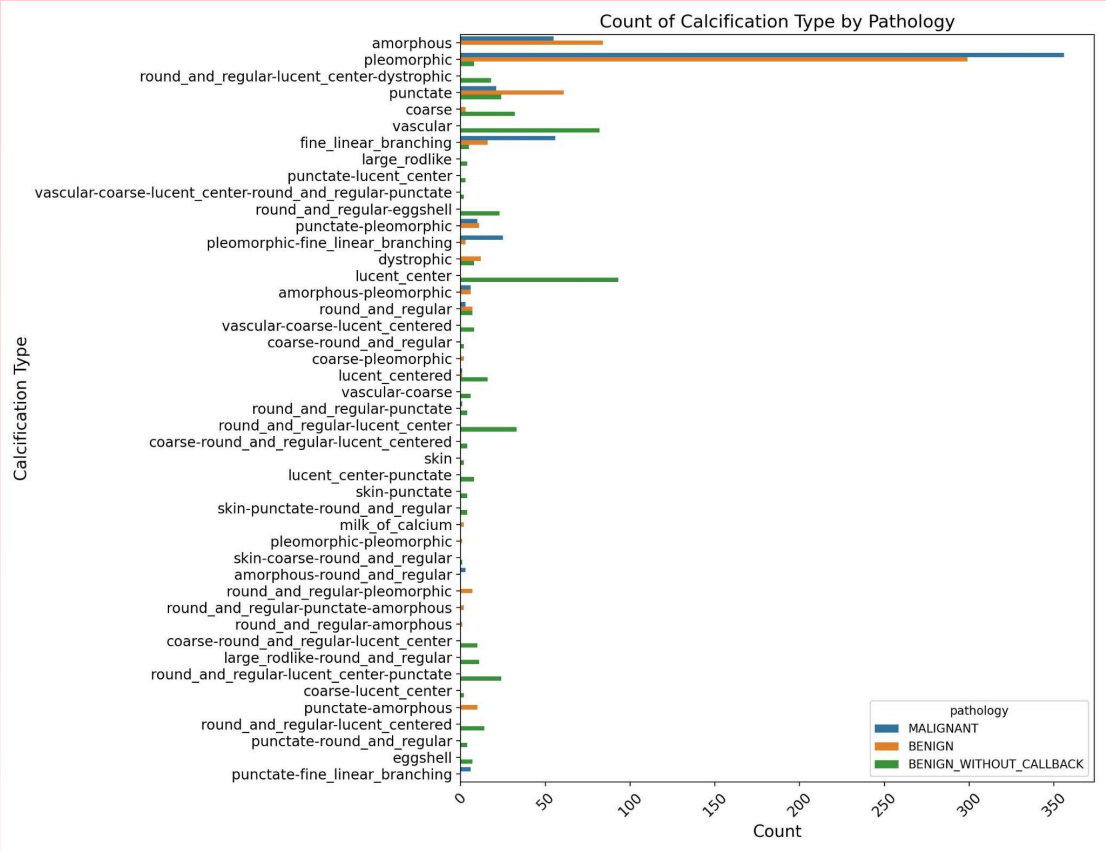




Exploratory Data Analysis



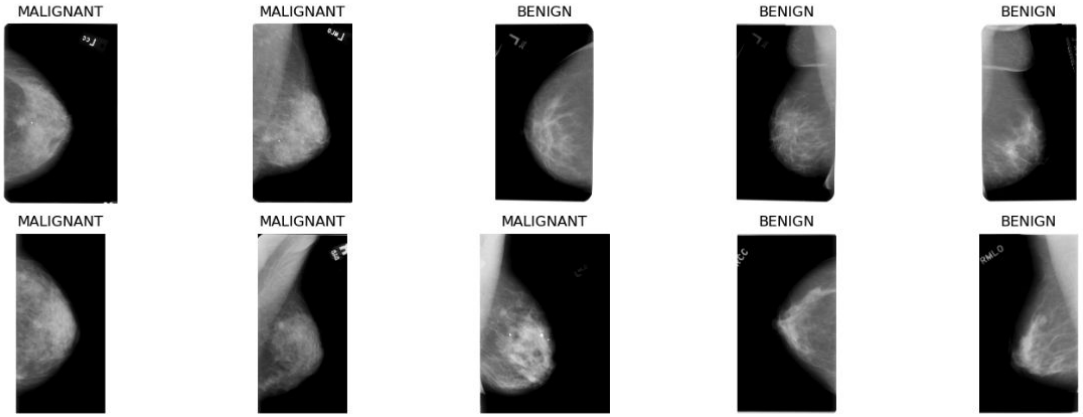
Exploratory Data Analysis



Images Preview

Mass Training Dataset

Full Mammograms:



Calc Training Dataset

Full Mammograms:

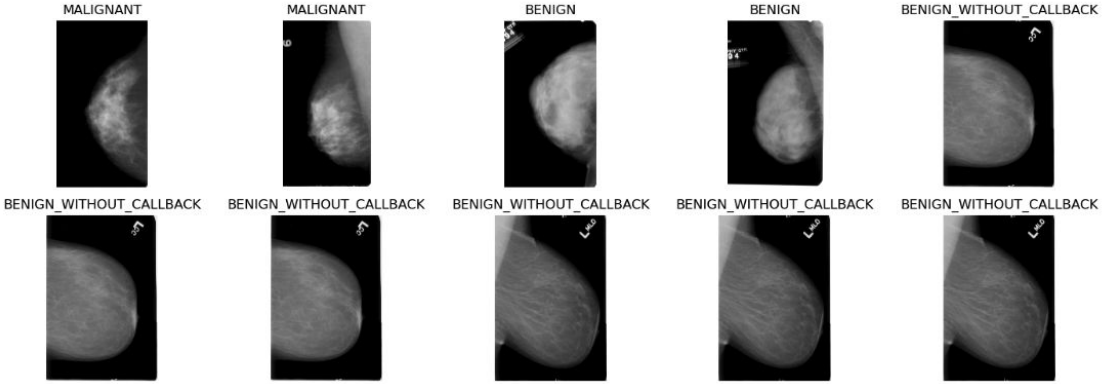
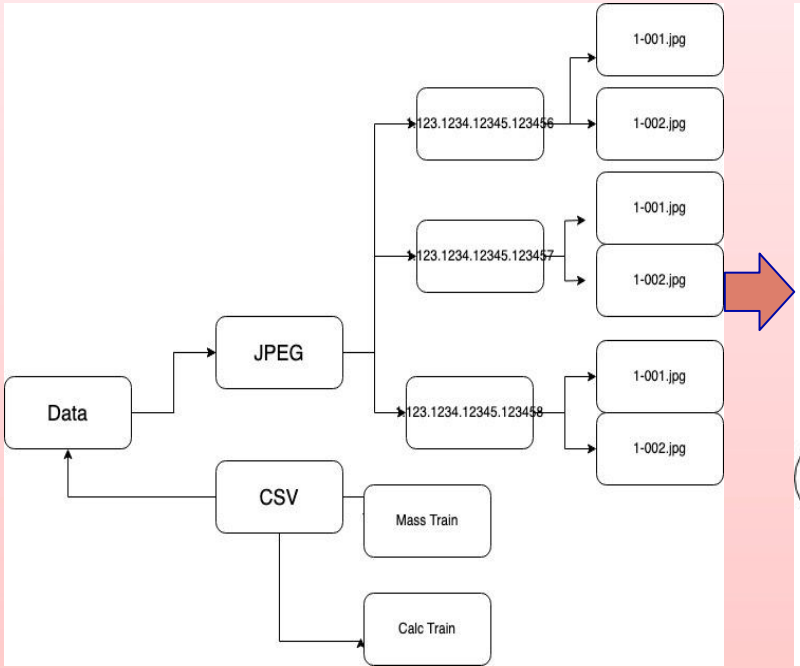


Image Folders Restructure



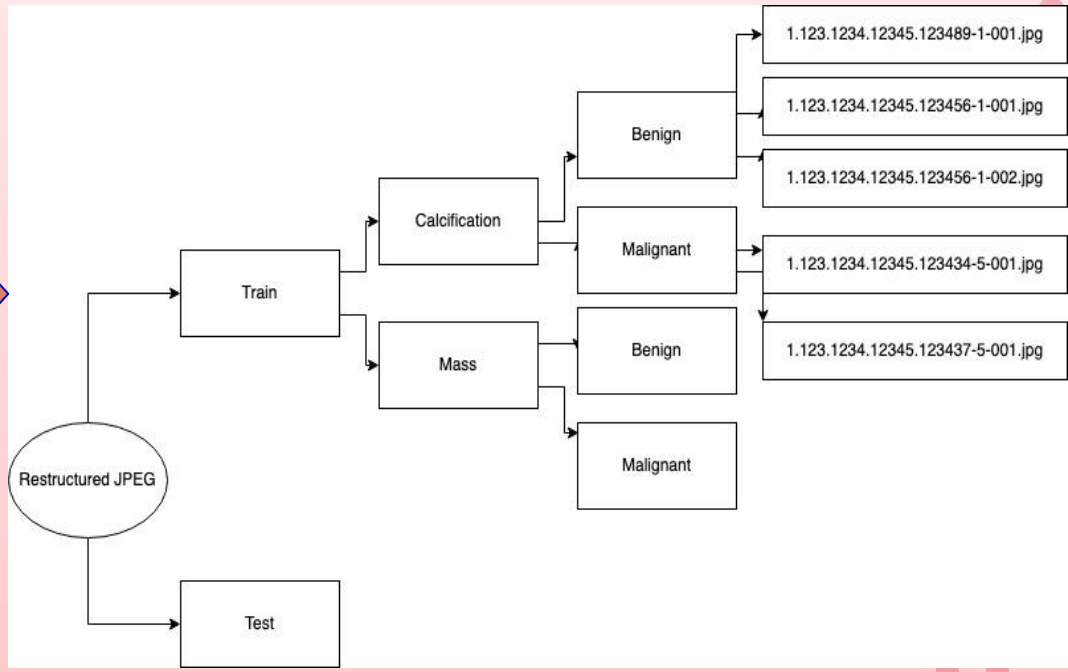
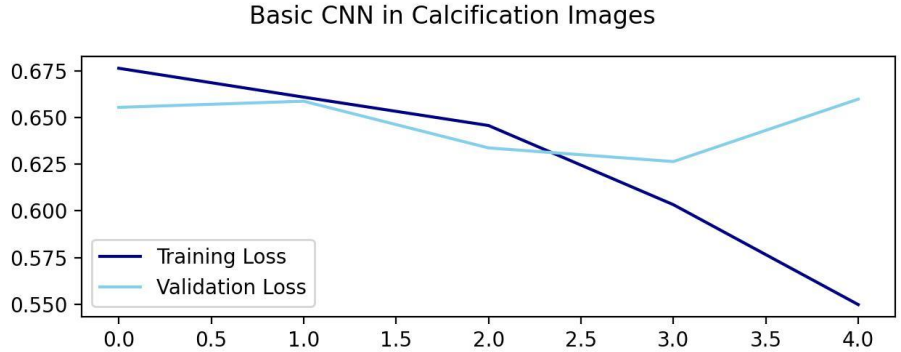


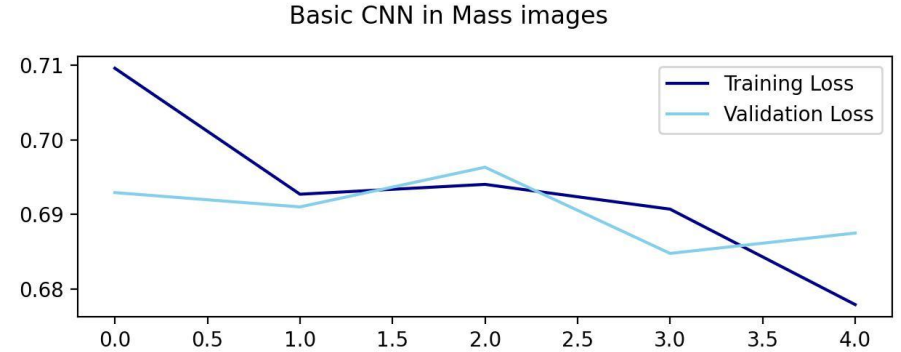
Image Preprocessing

- Resizing: Resizing images from over 3000 pixels by 3000 pixels to 512x512
- Normalization: Scaling pixel values to a standard range

Modeling and Model Evaluation

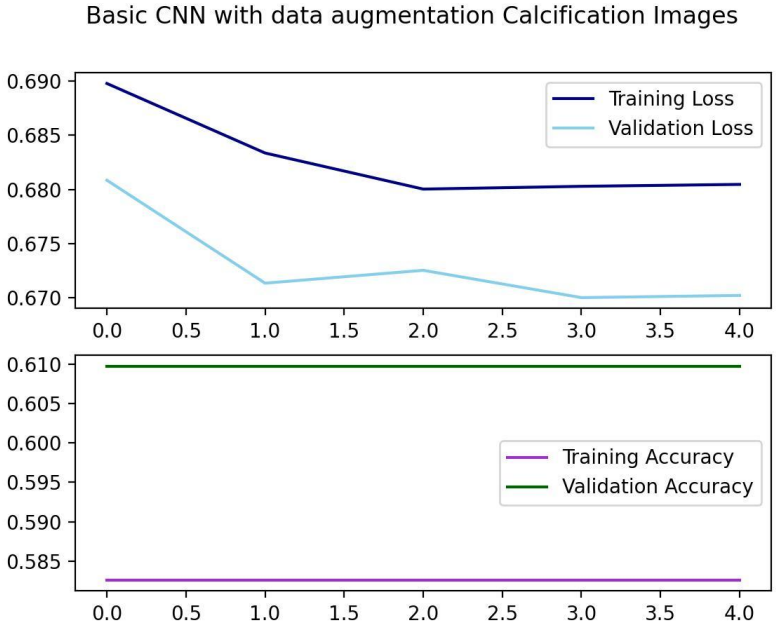
Convolutional Neural Networks

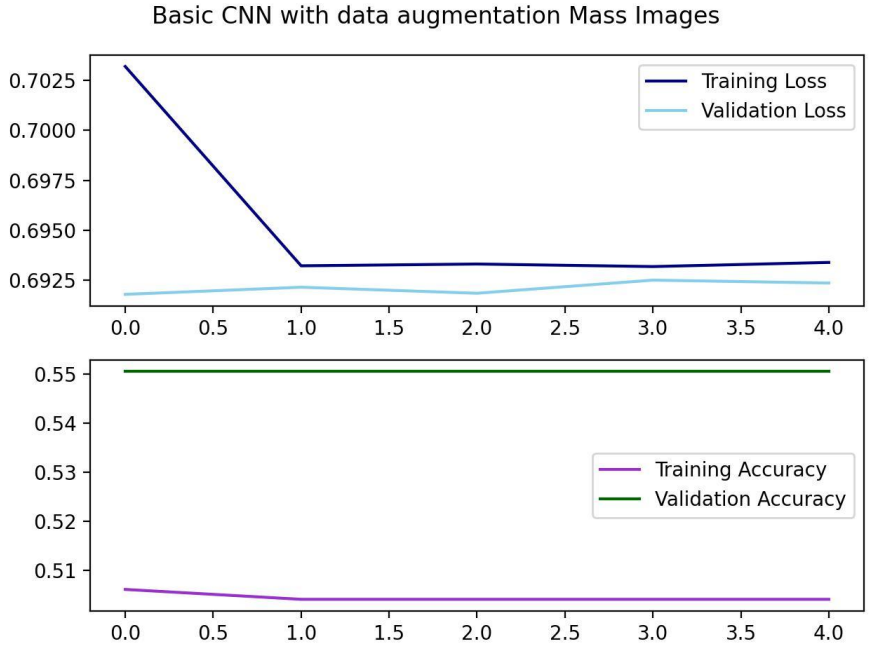




Modeling and Model Evaluation

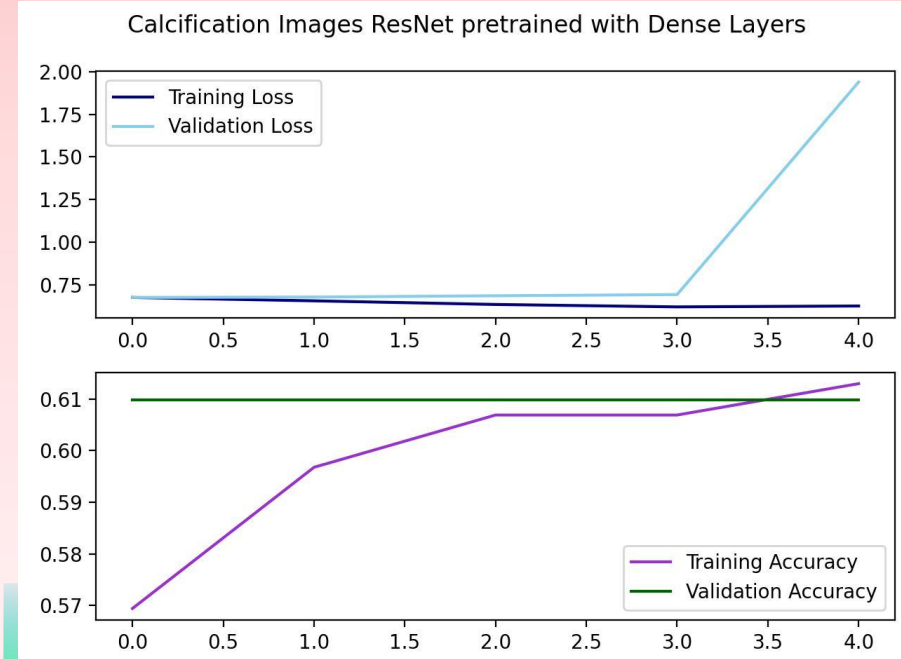
Convolutional Neural Networks with Data Augmentation

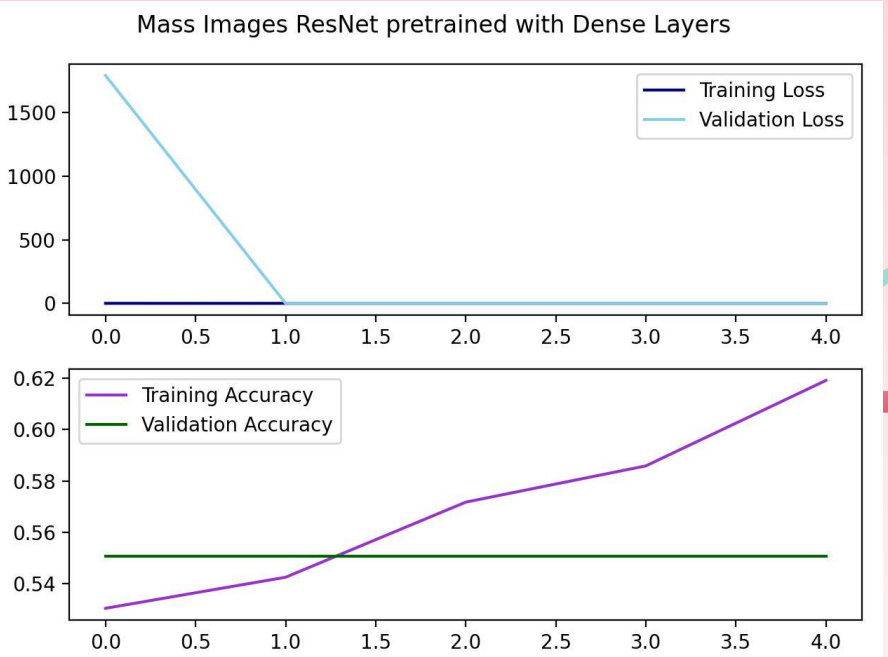




Modeling and Model Evaluation

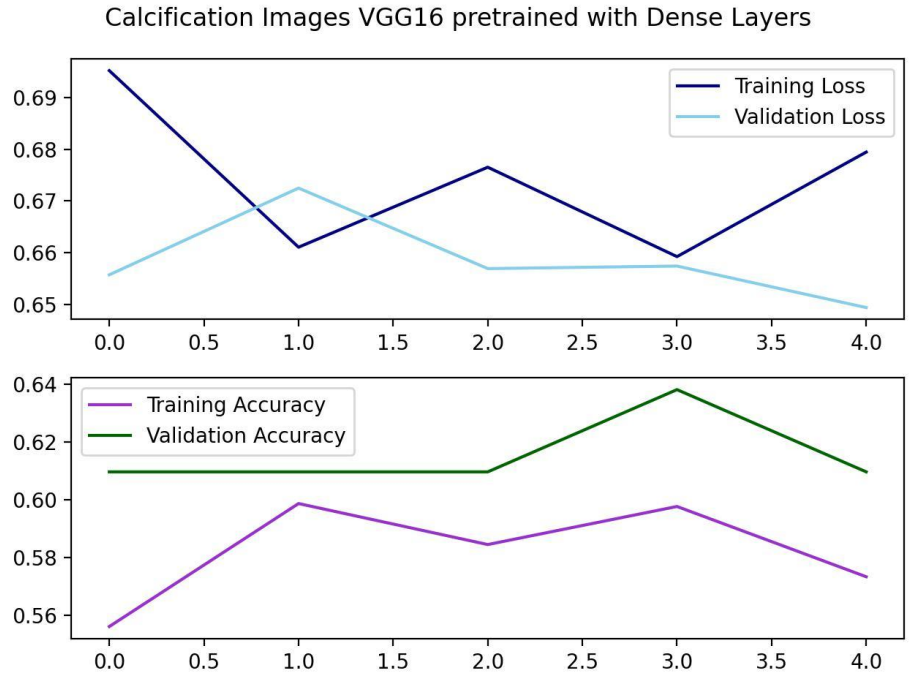
ResNet50 Pretrained Model

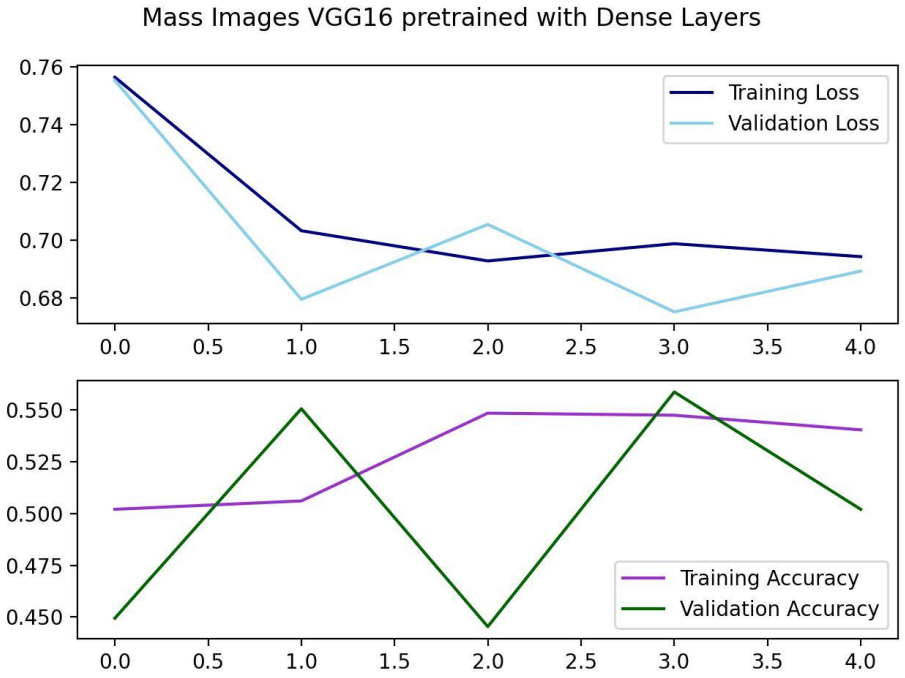




Modeling and Model Evaluation

VGG16 Pretrained Model

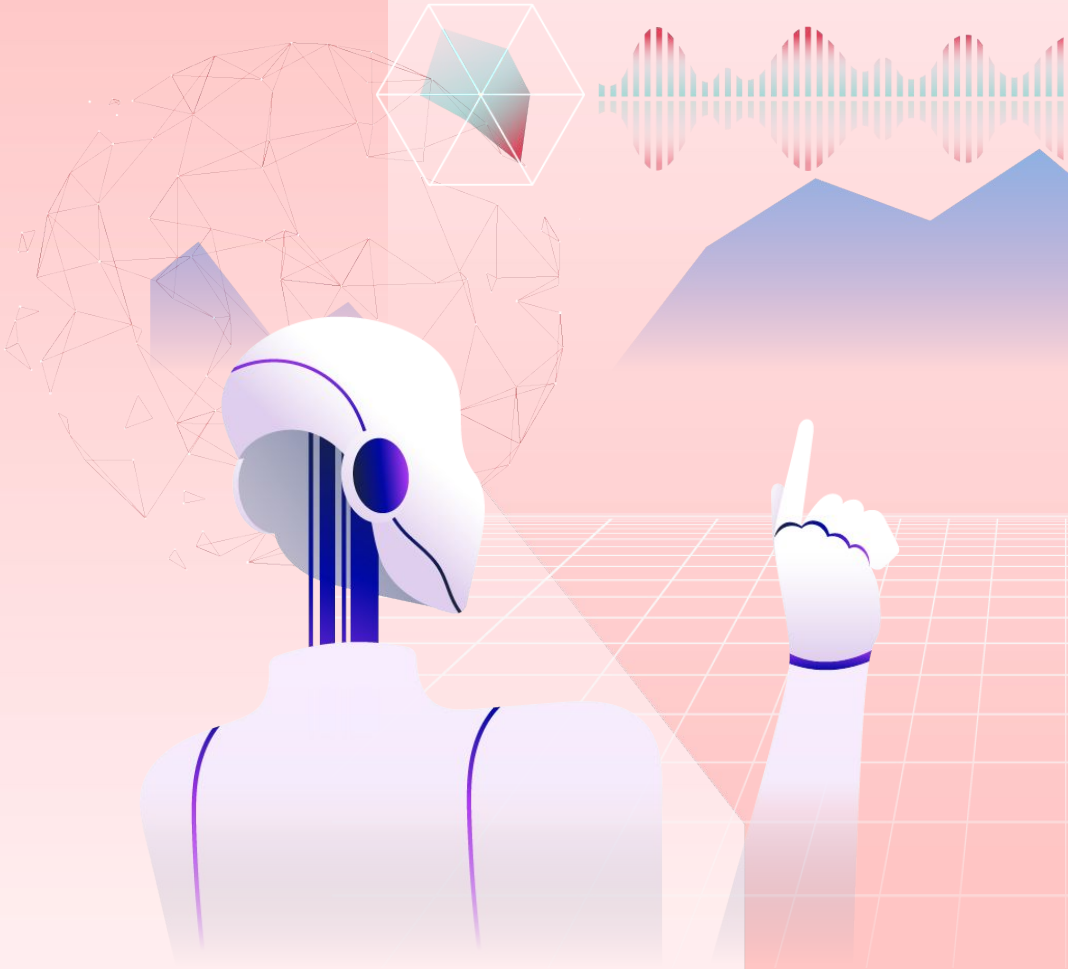




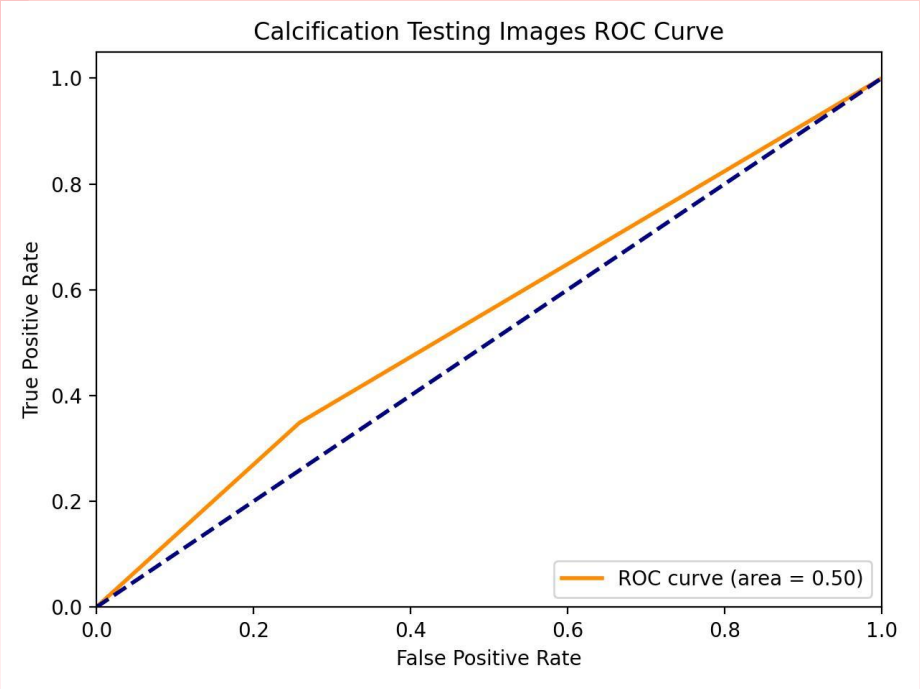
61.0%

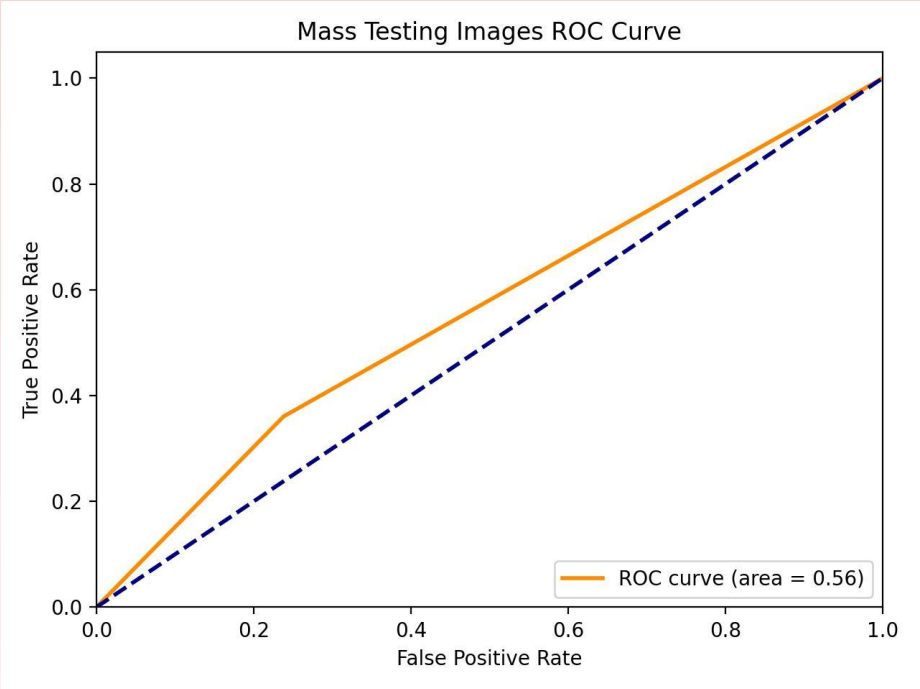
60.6%

Accuracy rates using CNN for mass images and calcification images, respectively.



Prediction


Calcification Testing Images ROC Curve
True Positive Rate
False Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0
ROC curve (area = 0.50)


Mass Testing Images ROC Curve
True Positive Rate
False Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0
ROC curve (area = 0.56)

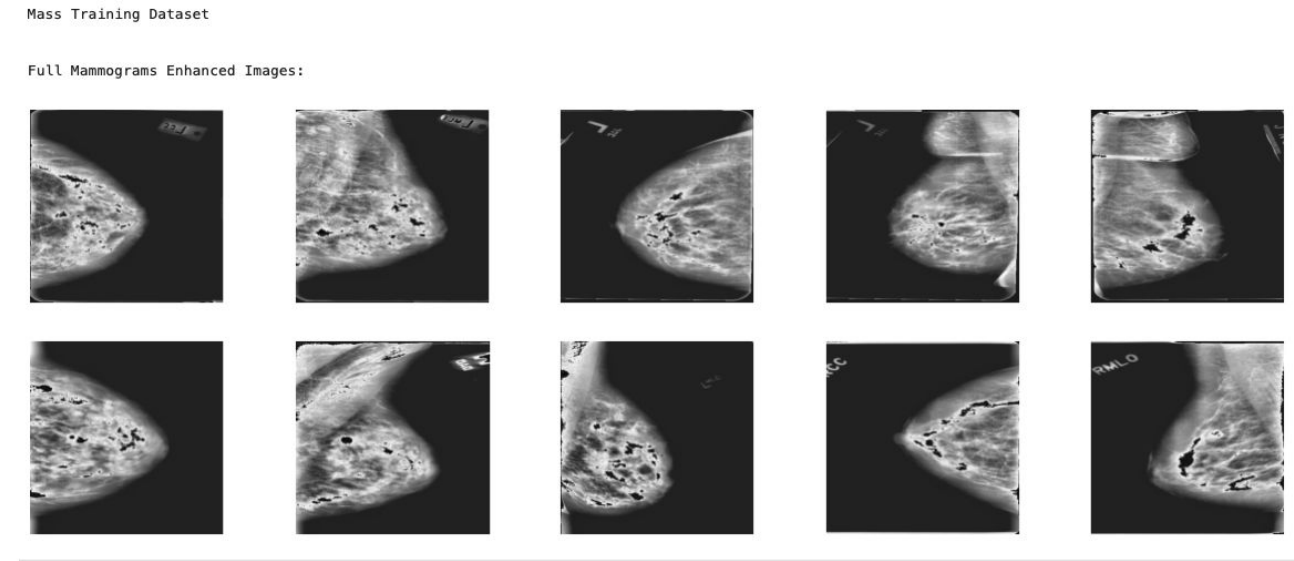
Conclusions

1. CNN model demonstrated moderate accuracy, displaying slightly superior results in overall cancer detection across both calcification and mass images

2. When applied to test images, the model yielded an F1 score of 0.416 for mass cases and 0.4 for calcification cases. However, these outcomes are suboptimal for breast cancer detection.

Future Improvement

Image Enhancement - using Contrast Limited Adaptive Histogram Equalization (CLAHE)



Reference

Data Citation: Sawyer-Lee, R., Gimenez, F., Hoogi, A., & Rubin, D. (2016). Curated Breast Imaging Subset of Digital Database for Screening Mammography (CBIS-DDSM) [Data set]. The Cancer Imaging Archive. https://doi.org/10.7937/K9/TCIA.2016.7O02S9CY

Publication Citation:
Rebecca Sawyer Lee, Francisco Gimenez, Assaf Hoogi, Kanae Kawai Miyake, Mia Gorovoy & Daniel L. Rubin. (2017) **A curated mammography data set for use in computer-aided detection and diagnosis research**. Scientific Data volume 4, Article number: 170177 DOI: https://doi.org/10.1038/sdata.2017.177

TCIA Citation:
Clark K, Vendt B, Smith K, Freymann J, Kirby J, Koppel P, Moore S, Phillips S, Maffitt D, Pringle M, Tarbox L, Prior F. The Cancer Imaging Archive (TCIA): Maintaining and Operating a Public Information Repository, Journal of Digital Imaging, Volume 26, Number 6, December, 2013, pp 1045-1057. DOI: https://doi.org/10.1007/s10278-013-9622-7*

CLAHE Citation
https://www.shs-conferences.org/articles/shsconf/pdf/2022/09/shsconf_etltc2022_03026.pdf

Thanks

Do you have any questions?
https://github.com/Elaine925