Identifying Malignant Mammograms

Using Transfer Learning with CNNs

Impact

- Second most common cancer for U.S. women
- Second most cancer related mortalities
- Fifth leading cause of death worldwide
- 1 in 8 women will develop it during their lifetime

Early detection

- Major reason deaths are down 43% since 2020
- Five year survival rates:
  - 99% - Localized SEER Stage
  - 86% - Regional SEER Stage
  - 31% - Distant SEER Stage
- Early detection = Less radical treatments
- Regular mammograms are key
  - Can detect abnormalities well before they can be felt
  - Skillful interpretation is critical

Problem Statement

- Mammograms aren’t perfect
- Require skilled mammographers
- 1 in 8 cancers go undetected
- High false positive rates
  - Costly (Time and Money)
  - Leads to unnecessary anxiety

Project Goal

- Use transfer learning with CNNs to identify malignant abnormalities in mammogram images
- Use updated and standardized Digital Database for Screening Mammography dataset (CBIS-DDSM)
- Predict malignant abnormalities better than a mammographer (87.5% sensitivity) while having a reasonable specificity

Dataset

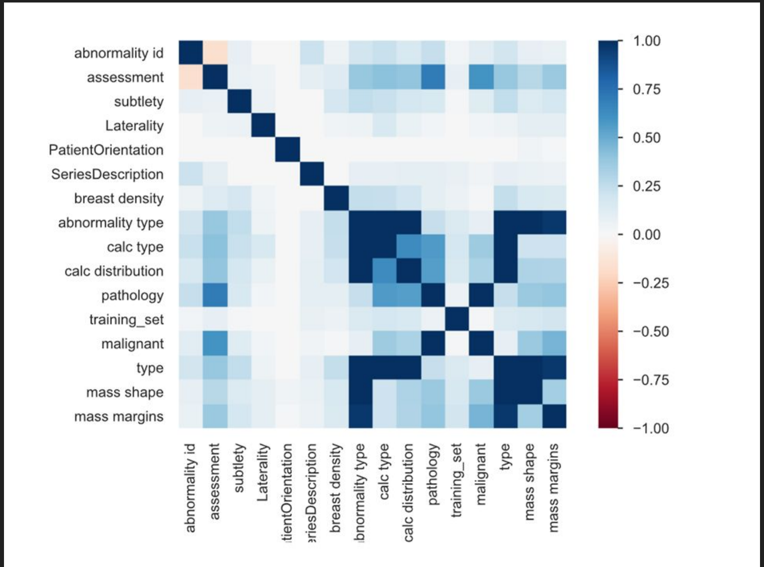
- Digital Database for Screening Mammography (DDSM)
  - 2,857 full mammogram images
  - 3,566 cropped mammogram images
  - 3,463 region of interest (ROI) mask files
- Normal, benign, and malignant cases for masses and calcifications with verified pathology information from four hospitals
  - Massachusetts General Hospital
  - Wake Forest University School of Medicine
  - Sacred Heart Hospital
  - Washington University of St. Louis School of Medicine
- Associated metadata files matched between training and testing sets

CBIS- DDSM Metadata

- Abnormality Type: Mass or Calcification
- Laterality: Side mammogram was performed on
- Patient Orientation: MLO or CC
- Breast density: Category of breast density
- Abnormality Type: Category of abnormality
- Calcification Type (calcifications): Type of calcification
- Calcification Distribution (calcifications): Distribution of the calcifications
- Mass shape (masses): Shape of the mass
- Mass margins (masses): Feature that separates the mass from the adjacent breast
- Assessment: Assessment category
- Pathology: Malignant, benign, or benign without callback
- Subtlety: Subtlety category

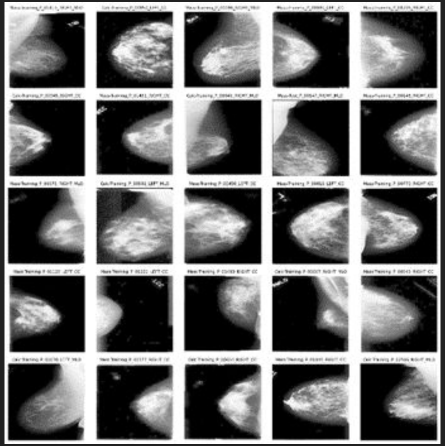
Exploratory Data Analysis

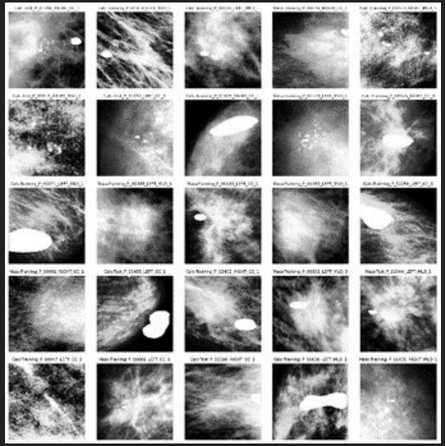
- Project focused on classifying images as (1) “Malignant” and (0) “Not Malignant”
- Differences in distributions between classes:
    - Calcification Distribution
    - Mass Margins
    - Assessment
- Correlations with Malignant:
    - Calcification Distribution
    - Calcification Type
    - Mass Margins
    - Mass Shape
    - Assessment



Model Training - Image Preprocessing

- Cropped images and ROI Masks
- Image Resizing
  - Height: 224 pixels
  - Width: 224 pixels
  - Color Channels: 3
- Histogram equalization



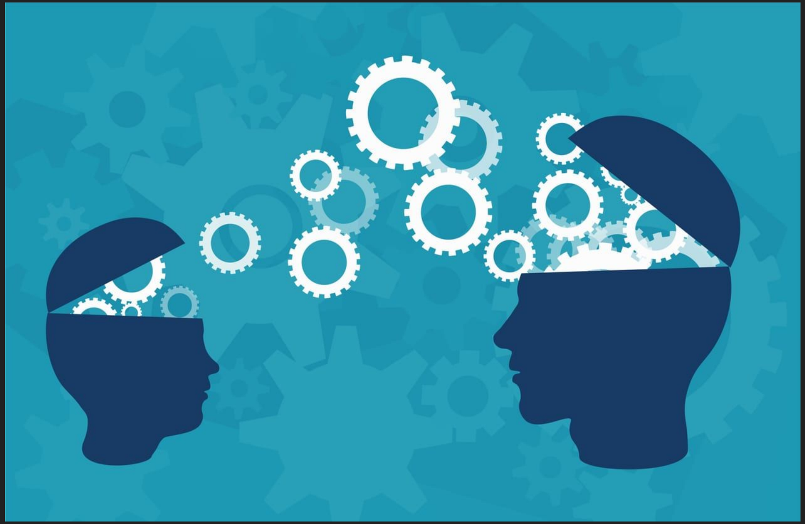


Model Training - Training, Validation, and Testing

- CBIS-DDSM
  - Pre-split into training and testing sets (80%, 20%)
  - Matched on metadata
- Validation Set
  - Created from random sample of testing set
  - 80 / 10 / 10 split

Model Training - Models

- Three different pretrained CNNs were evaluated during training on two sets of labels:
    1. DenseNet169 - Malignant
    2. DenseNet169 - Pathology
    3. Xception - Malignant
    4. Xception - Pathology
    5. EfficientNetV2S - Malignant
    6. EfficientNetV2S - Pathology
- Imagenet weights were used
- Relatively small size models with high accuracy on image classification tasks



Model Training - Model Specifications

- Optimizer: Adam
- Loss Function: Categorical Crossentropy
- Learning Rate: .001
- Batch Size: 32
- Epochs: 30 (with early stopping)

Model: "sequential_7"

Layer (type)	Output Shape	Param #	Trainable
sequential_6 (Sequential)	(None, 224, 224, 3)	0	Y
xception (Functional)	(None, 2048)	2086148 0	Y
dropout_3 (Dropout)	(None, 2048)	0	Y
dense_3 (Dense)	(None, 3)	6147	Y

Total params: 20867627 (79.60 MB)
Trainable params: 4754947 (18.14 MB)
Non-trainable params: 16112680 (61.46 MB)

Model Selection

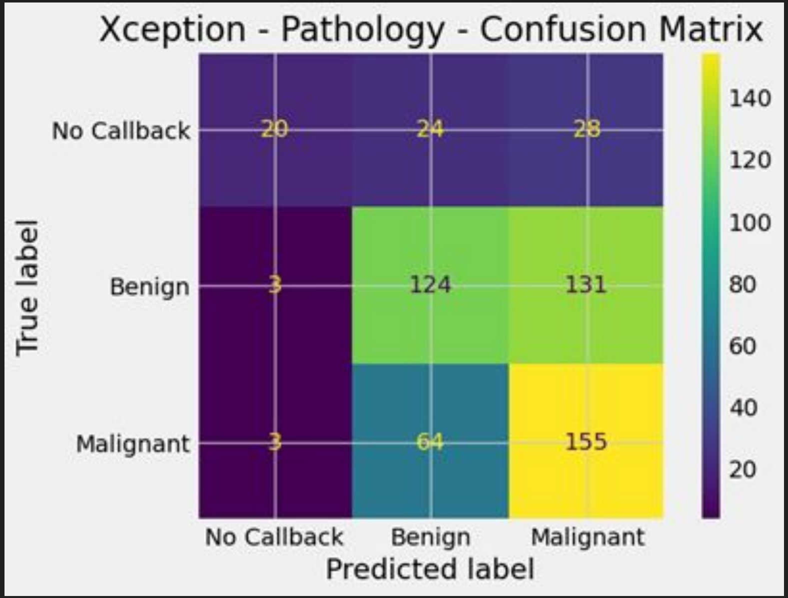
- Xception
    1. Malignant labels
        - Sensitivity = 0.85
        - Specificity = 0.29
    2. **Pathology Labels**
        - Sensitivity = 0.70
        - Specificity = 0.52
- Highest sensitivity on the testing set with specificity higher than Dummy Classifier

	precision	recall	f1-score	support
Benign	0.72	0.52	0.60	330
Malignant	0.49	0.70	0.58	222
accuracy			0.59	552
macro avg	0.61	0.61	0.59	552
weighted avg	0.63	0.59	0.59	552

F1 Score: 0.59
Sensitivity: 0.70
Specificity: 0.52

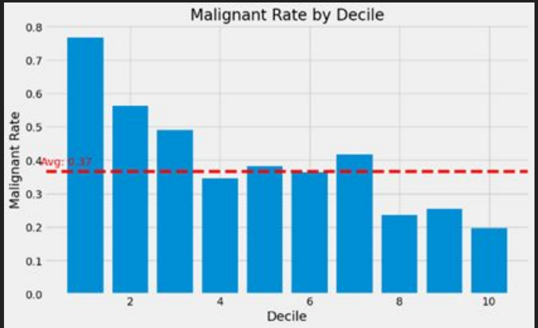
Testing Set - Further Analysis

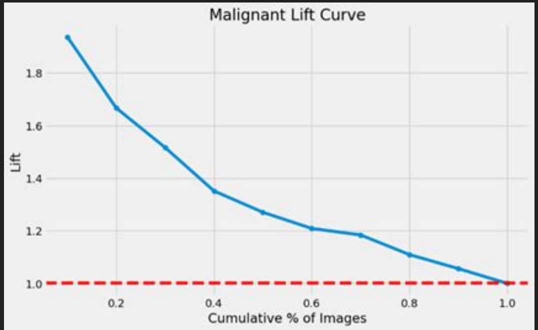
- Most errors confused malignant with benign images (and vice versa)
- Very few predictions for No Callback (Normal)
- Model seemed to naturally group images into two categories



Testing Set - Decile Analysis

- More true malignant cases as predicted probabilities increased
- Effect was most noticeable at the extremes
- Implications for practical applications





Final Thoughts and Future Research

- Model was not able to meet the initial goal of classifying malignant mammograms with > 87.5% sensitivity
- Importance of false positives in final model selection
- Ideas for future research / practical applications:
  1. Train model on more images
  2. Use multiple images from same patient to diagnose cases
  3. Use predicted probabilities to reduce number of reviewed mammograms

Thank you.