DEEP LEARNING BASED BREAST CANCER PREDICTION SYSTEM

STATS & FACTS

1. Stephen Gathai.
2. Tony Munene.
3. Loise Mbago.
4. Nazra Nyangwara.
5. Cleophas Opati.
6. Immanuel Omondi.
7. Bahati Ndwiga.

TABLE OF CONTENT

- Business Overview
- Problem statement
- Objectives
- Success Criteria
- Data Understanding
- Data Preparation
- Modelling
- Conclusions
- Recommendations

Business Overview

Breast cancer is a leading global cancer and the second most common cause of cancer-related deaths, particularly among women aged 45-55.
Early detection is crucial for improving breast cancer outcomes, with proper diagnosis and treatment significantly enhancing survival rates.

Mammography and ultrasound are primary tools for early breast cancer diagnosis, but they have limitations. They often lead to false positives, requiring follow-up tests like aspiration or biopsy after a breast ultrasound. Some tumours also go undetected by ultrasound, and annual mammography can miss early tumours presenting as calcifications. Advanced tools are needed for more effective early breast cancer detection.

Problem Statement

The breast cancer research foundation (BCRF) has partnered with us to develop a breast cancer prediction system(BCPS) aimed at using deep learning technology to enhance early breast cancer detection.

This system will analyse medical imaging data, such as mammograms and ultrasounds, to detect breast cancer presence with a high degree of accuracy. Through the deep learning-(BCPS), BCRF seeks to transform breast cancer research, improving the precision and efficiency of diagnosis, providing a valuable tool for healthcare professionals to enable timely intervention, enhance patient outcomes, and support decision-making.

Objectives

To develop a deep learning model capable of efficient segmentation of breast masses in ultrasound images.

To identify critical parameters for breast cancer detection

To implement a user-friendly interface for healthcare professionals to upload medical images and receive predictions.

To develop a model with at least 90% specificity and 90% sensitivity for accurate predictions.

Success Criteria

- ❖ Achieve a classification accuracy of at least 85% on the test dataset
- ❖ Achieve a sensitivity of at least 90% and a specificity of at least 90%
- ❖ Have a high F1-score to ensure a balance between precision and recall
- ❖ Demonstrate good generalization by performing well on unseen data

Data Understanding

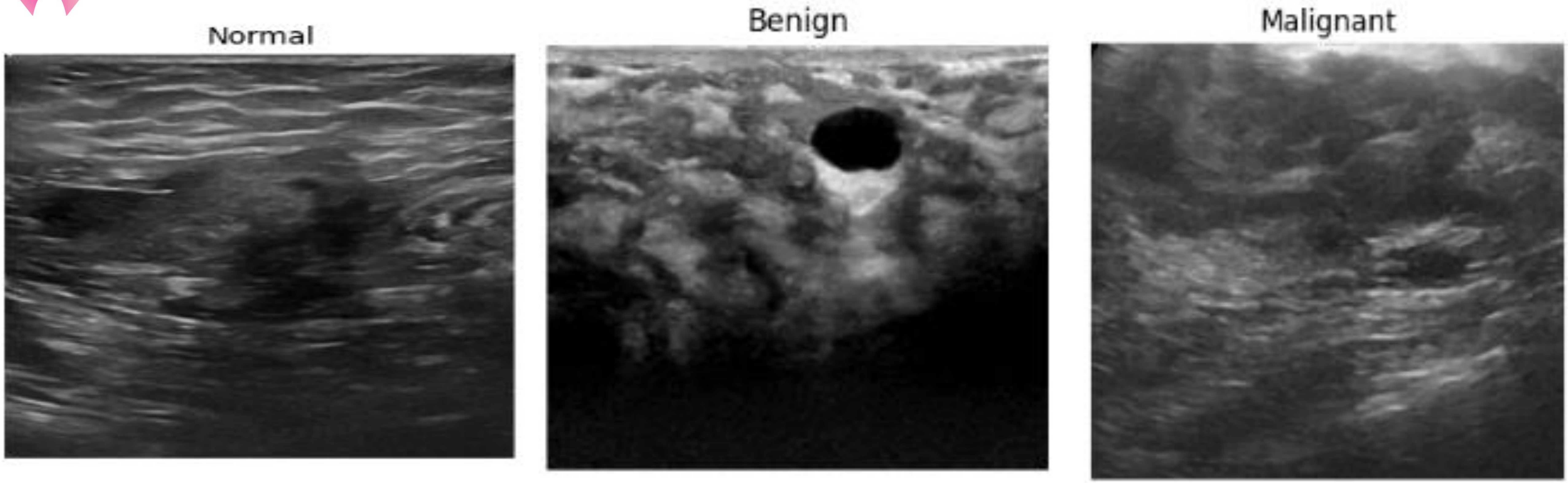
The project used one dataset, consisting of ultrasounds from the IN-breast database.

The dataset has 780 breast ultrasound images from 600 female patients aged between 25 and 75 years.

The data is categorized into three sets:

- Benign-has 437 images.
- Malignant-has 210 images.
- Normal-has133 images.

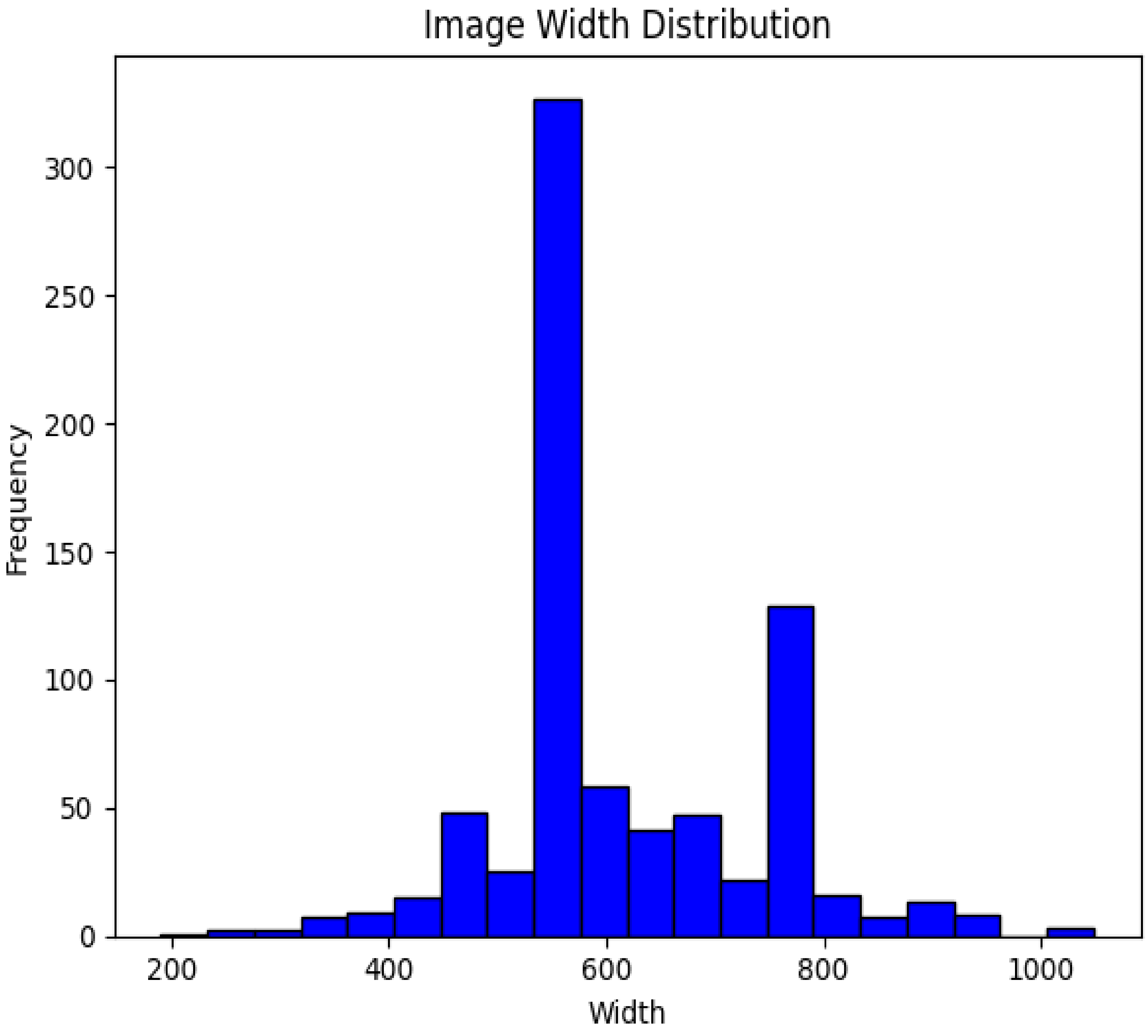
Sample Images



- ❖ A normal image is mostly grey, with some white areas showing healthy dense tissue.
- ❖ A Benign mass is usually circumscribed oval and round, it has a regular shape.
- ❖ An irregular shape suggests a greater likelihood of malignancy.

Image Sizes

The Images vary in width and height.



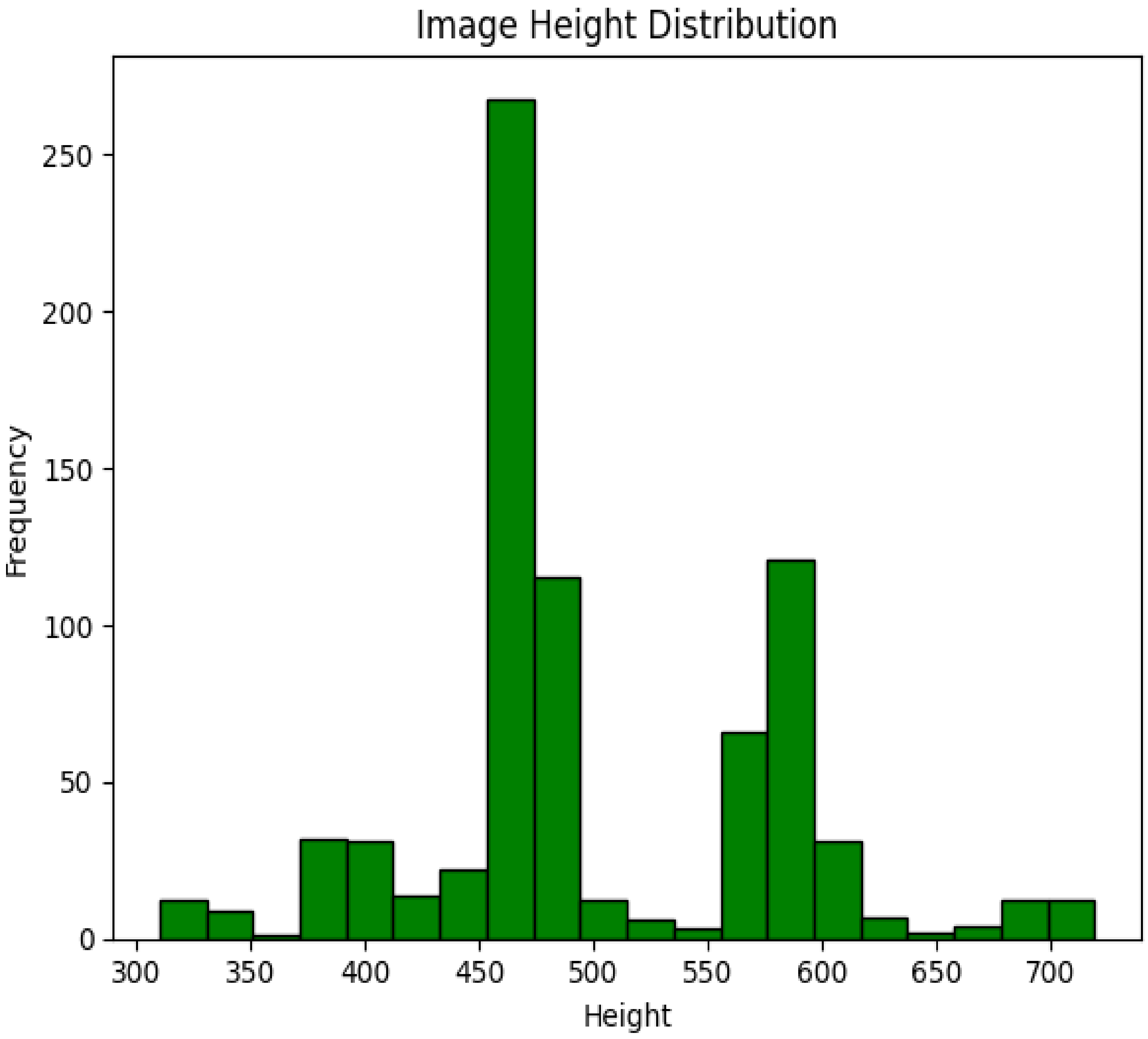
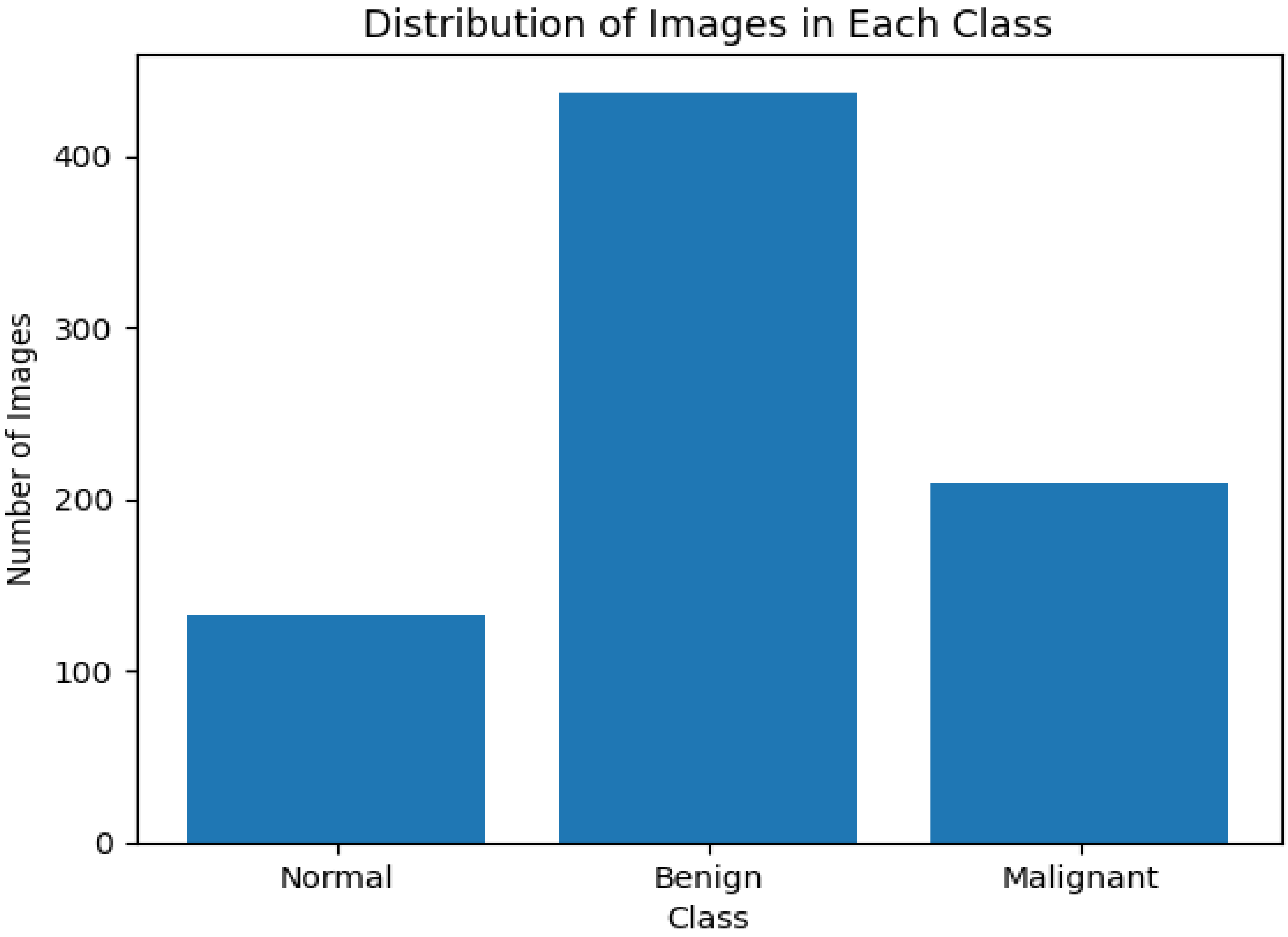


Image distribution per class



- The output reveals a noticeable class imbalance in the dataset

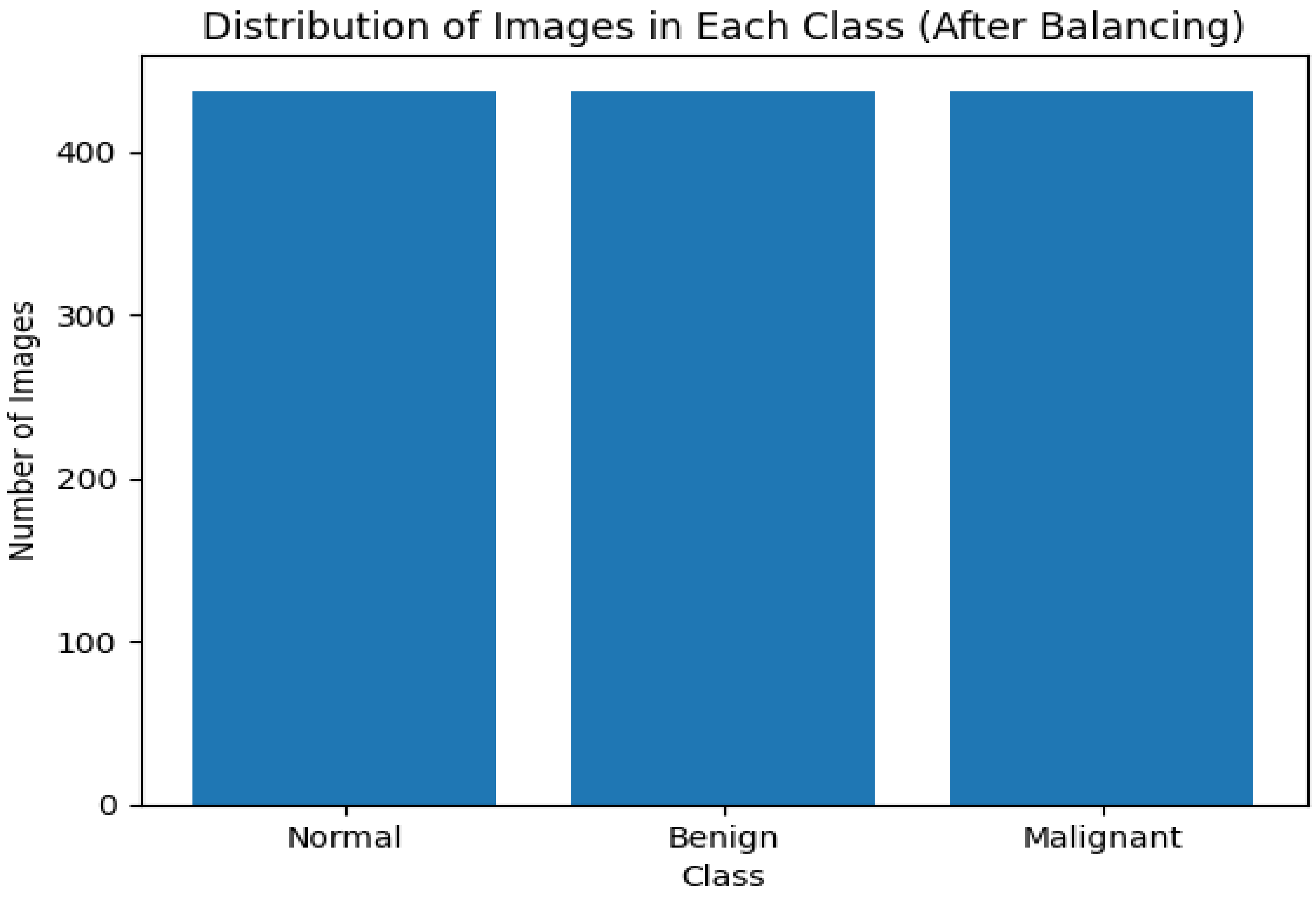
- The benign category is significantly over represented compared to the other categories.

Data Preparation

This step involved:

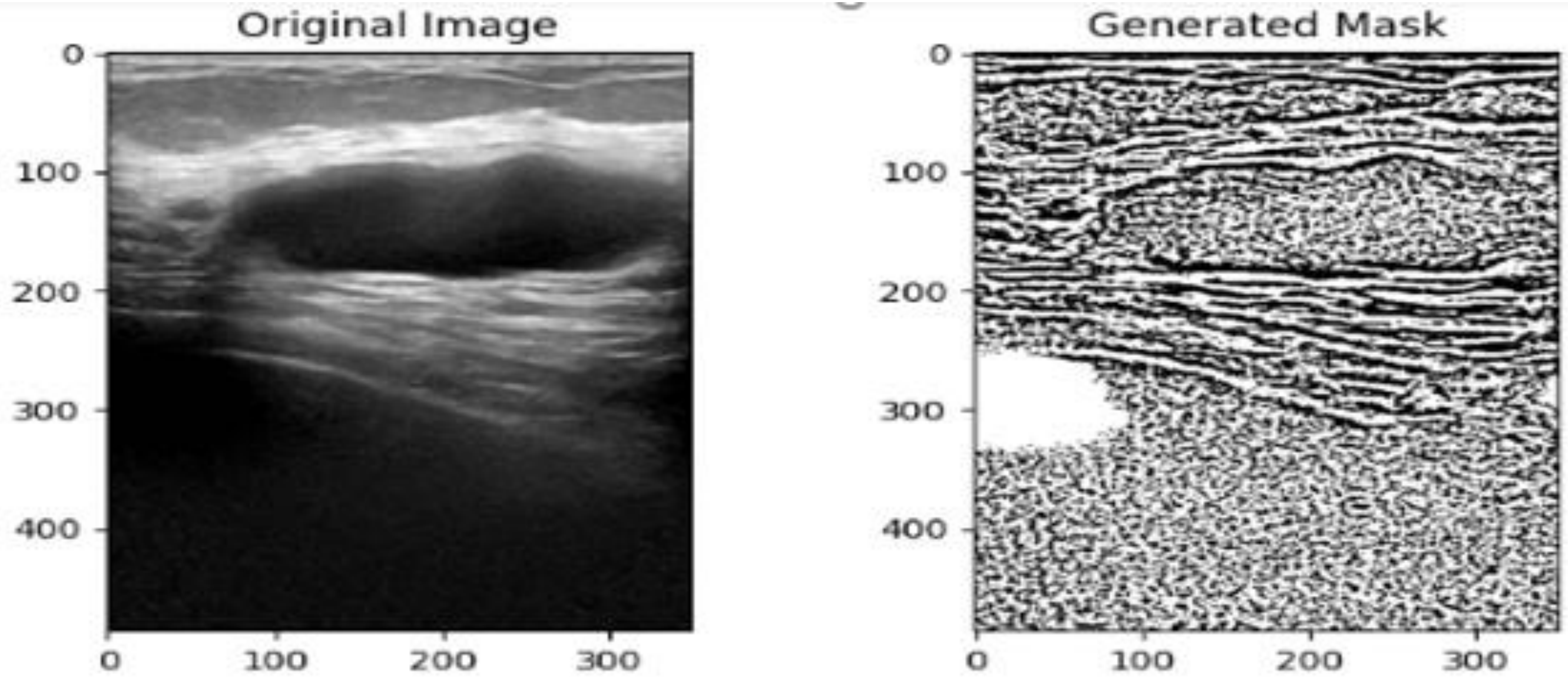
- Addressing class imbalance
- Resizing the images to a consistent size
- Normalizing the pixel values to a range between 0 and 1
- Creating labels for each class
- Applying data augmentation techniques to increase the variability and size of the training dataset

Image distribution after handling class Imbalance



- The Class Imbalance was completely addressed.
- Each category has 437 images.

Mask Application

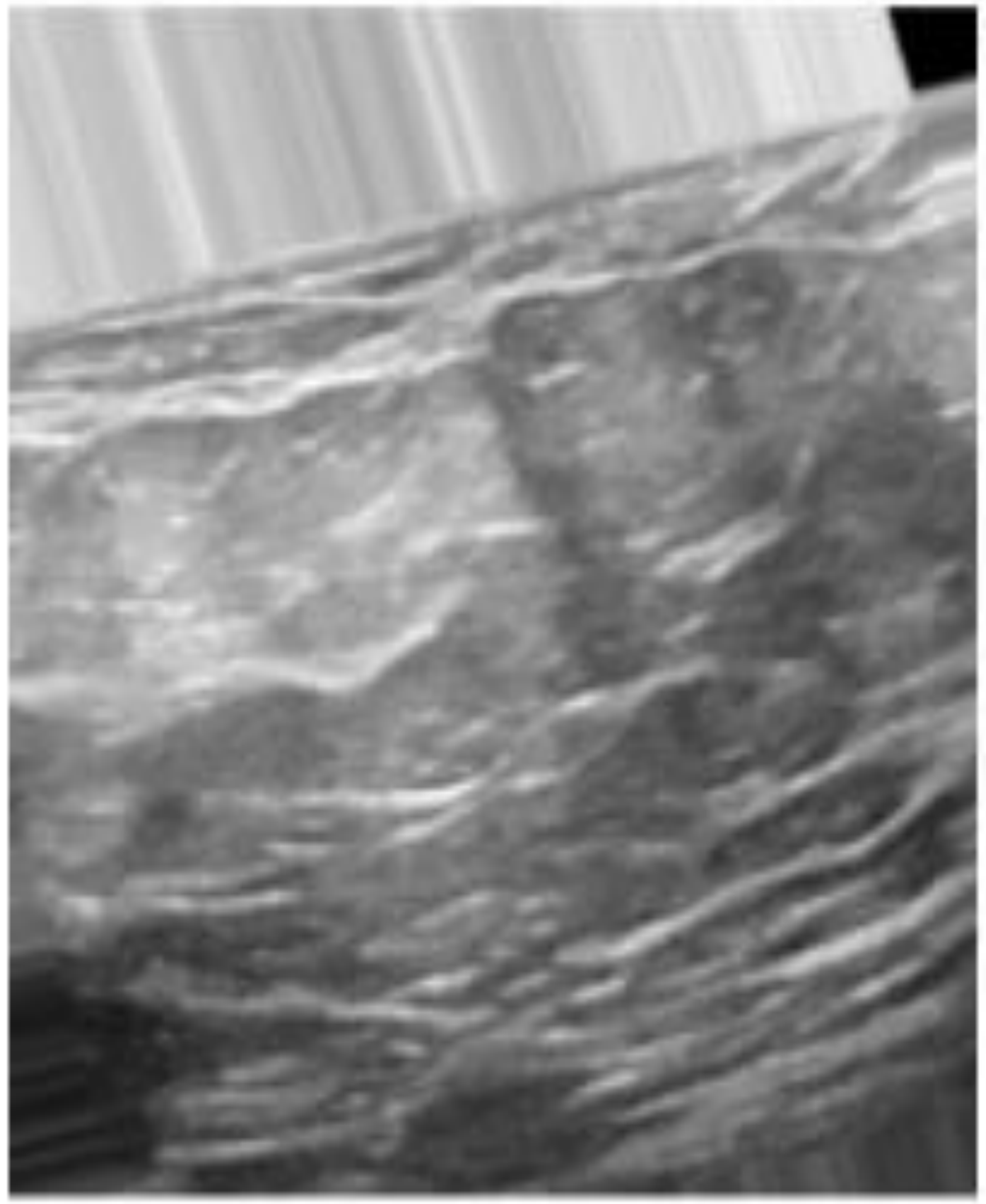


❖ Masking is crucial for possibly highlighting areas with cancerous cells as denoted by the white patch.

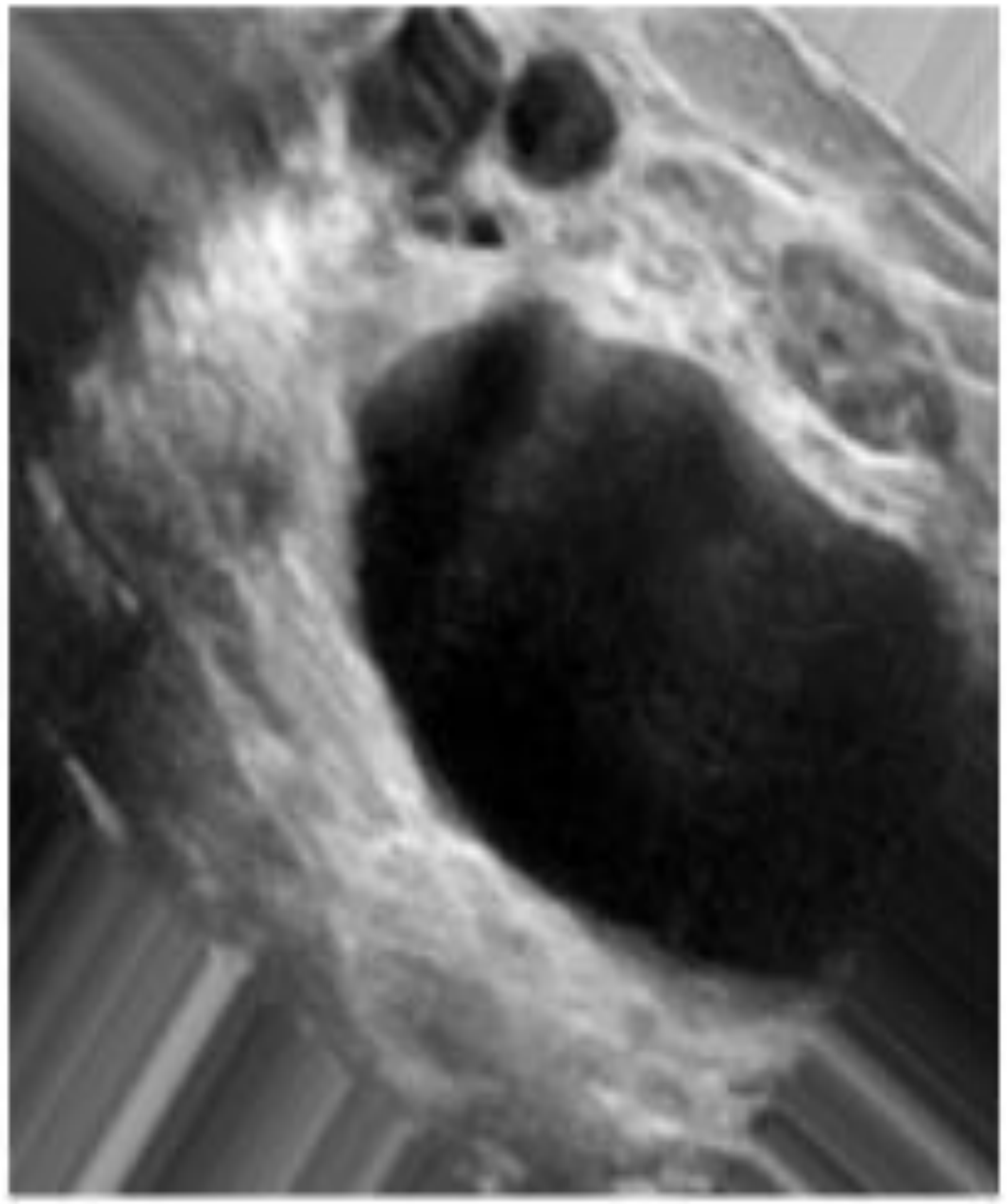
Image Augmentation

Data Augmentation increased the images form 1311 to 13110

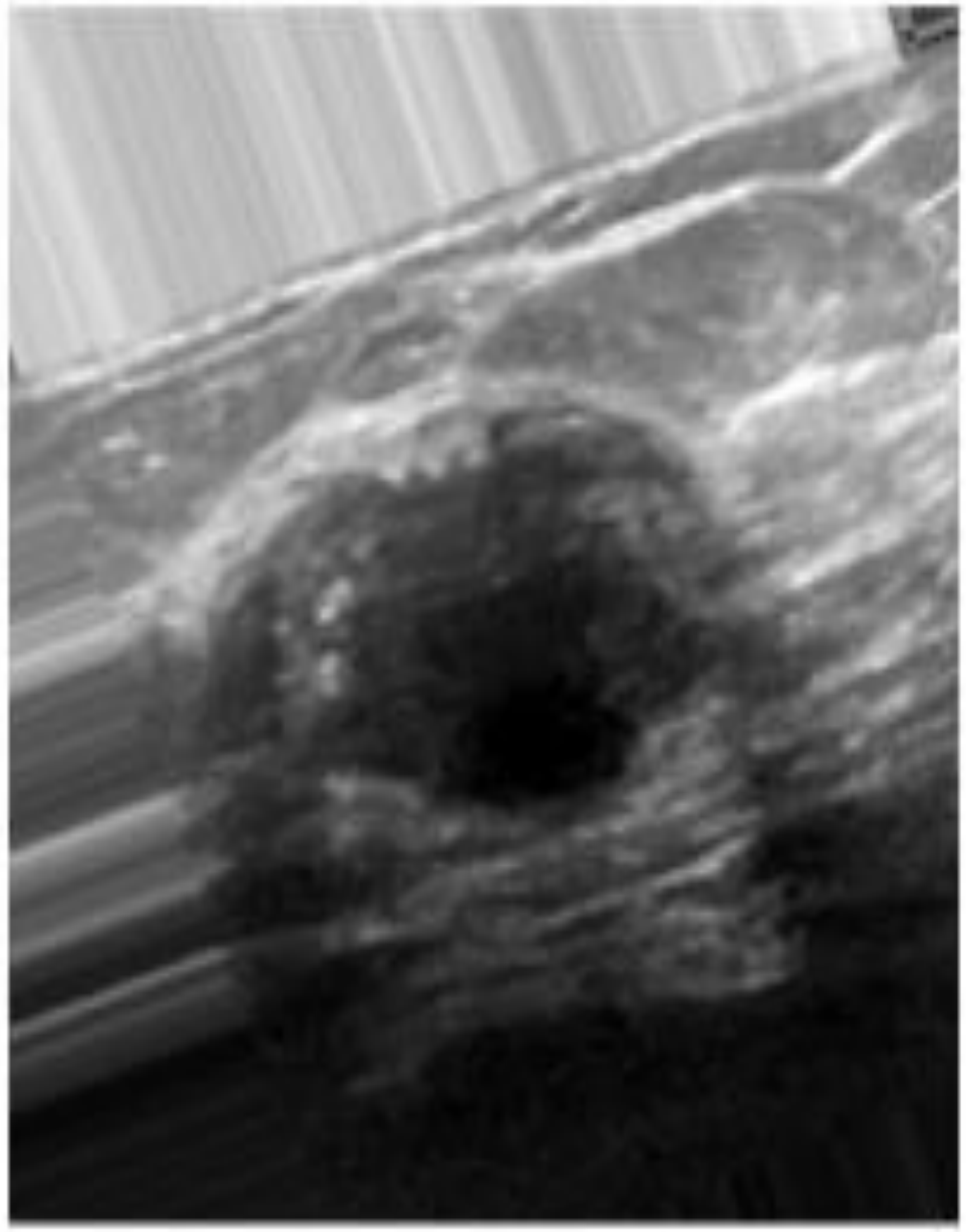
Augmented Normal image



Augmented Benign image



Augmented Malignant image



Modeling

The Following models were used;

- Baseline CNN model
- Baseline CNN model with additional layers
- Pretrained VGG16 model

Modelling results and Evaluation

Models	Accuracy	Recall Normal	Recall Benign	Recall Malignant
Baseline Model CNN	0.81	0.88	0.71	0.84
Model 2_CNN Model with added layer	0.34	0.04	0.00	0.99
Model 3_VGG16	0.53	0.54	0.49	0.56
Tuned Model	0.84	0.91	0.75	0.87

- The baseline CNN moc achieved an accuracy of 81% which was the best
- High accuracy ensures our test is reliable
- It also demonstrated a balanced performance in terms of recall with values approximately around 0.88 for recall in the "Normal" class, 0.71 for recall in the "Benign" class, and 0.84 for recall in the "Malignant" class
- Recall ensures our test captures all the cancer cells

Conclusions

1.With specificity and sensitivity over 81%, The model shows promise in positively impacting early breast cancer detection and healthcare decisions

2. The model's training progress reveals steady improvement in breast cancer image classification, reflected by decreasing loss, indicating effective learning and convergence.

3.Overall, our deep learning based breast cancer prediction system holds significant promise for improving detection and intervention in breast cancer cases

Recommendations

- ❖ Experiment with different model architectures and pretrained models like ResNet, Inception, or Efficient Net to enhance model performance.
- ❖ Establish a feedback loop with clinicians for model refinement.
- ❖ Expand the dataset with more diverse samples for increased model robustness.
- ❖ Incorporate patient metadata and clinical parameters for improved accuracy

Thank you!

Q&A?