Screening Mammography Breast Cancer Detection

Domain Background:

Healthcare (AI in medical imaging).

Recently, deep learning has earned a great momentum in medical imaging applications. Deep learning can empower radiologists and physicians in their workflow. Three major areas in that DL excels are anomaly detection, characterization (including segmentation and diagnosis), and progression monitoring.

I am interested in this particular project since my master's thesis is about breast cancer risk prediction from mammograms using deep learning. I am excited to apply the AI techniques to the detection problem stated below.

Problem Statement:

Breast cancer is one of the most prevalent cancers worldwide. According to the WHO, in 2020 alone, there were 2.3 million new breast cancer diagnoses and 685,000 deaths. Early detection of breast cancer is crucial for better treatment plans and survival outcomes. Early detection now relies on the radiologist's expertise and judgement. Therefore, screening mammography is expensive. Also, it is subjected to large number of false positives.

Dataset:

Kaggle's Competition: RSNA Screening Mammography Breast Cancer Detection

https://www.kaggle.com/competitions/rsna-breast-cancer-detection/data

Inputs: screening mammogram images

Solution Statement:

Deep learning, specifically convolutional neural networks (CNNs) can aid in breast cancer detection. A CNN can be used to detect the presence of cancer and therefore reduce the time/effort needed by the radiologist. I propose using a Siamese neural network which has two CNNs with shared weights. This can leverage the anatomical symmetry between the two breasts and therefore produce more reliable results and reduce false positive detection.

Benchmark Model:

A normal CNN can be used as a benchmark.

Evaluation Metrics

- Accuracy

- Probabilistic F-score: This extension of the traditional F score accepts probabilities instead of binary classifications.

Project Design
- Data download and EDA
- Data preprocessing
- Build the siamese model
- Train on the dataset
- Evaluate the results using the metrics