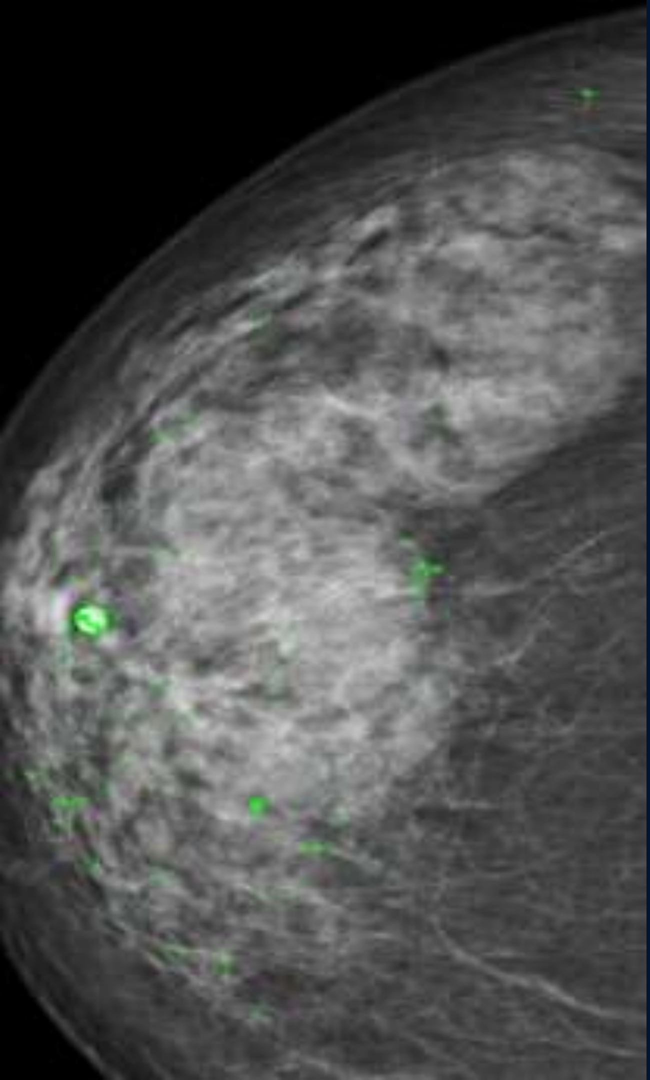
Finding informative regions in grayscale mammogram images.

Student : Arzu Fatullayeva
Course#name: CSCI6917: Guided Research Methods
Date: 08.08.2023



Project Objective:

The problem addressed in this research is the identification of informative regions in grayscale mammogram images and then classification of these as negative or positive ones .

Heilmeier Questions

1. What are you trying to do?
Developing an approach to identify keypoints in mammogram images that contain relevant information for breast cancer detection and diagnosis.
2. How is it done today, and what are the limits of current practice?
Automated methods exist but often lack the ability to accurately locate informative regions
mammogram analysis is primarily performed manually by radiologists
3. What's new in your approach and why will it be successful?
Building custom approach for feature extraction and training
4. Who cares? If you are successful, what difference will it make?
The work will benefit both radiologists and patients.
5. What are the risks and the payoffs?
Effectiveness of algorithms may be affected by the variability in image characteristics, leading to possible false positives or false negatives.
6. How much will it cost? How long will it take?
Team of 5 researchers.
No budget is required.
End of 2023.
7. What are the midterm and final exams to check for success?
Feature evaluation for midterm and classification for final.
8. Why now?
High demand, and there are no alternatives yet.

2.3 million women

Were diagnosed with breast cancer in 2020

Early detection may prevent the lethal cases

Strategy

QUALITATIVE AND QUANTITAVE STRATEGY is applied for this research.

- **Theory/hypothesis**
- **Data-Driven Approach**
- **Measurable Performance Metrics**
- **Statistical Analysis**
- **Quantitative Validation**

Qualitative Analysis – Theory and Hypothesis

The images we are dealing with:

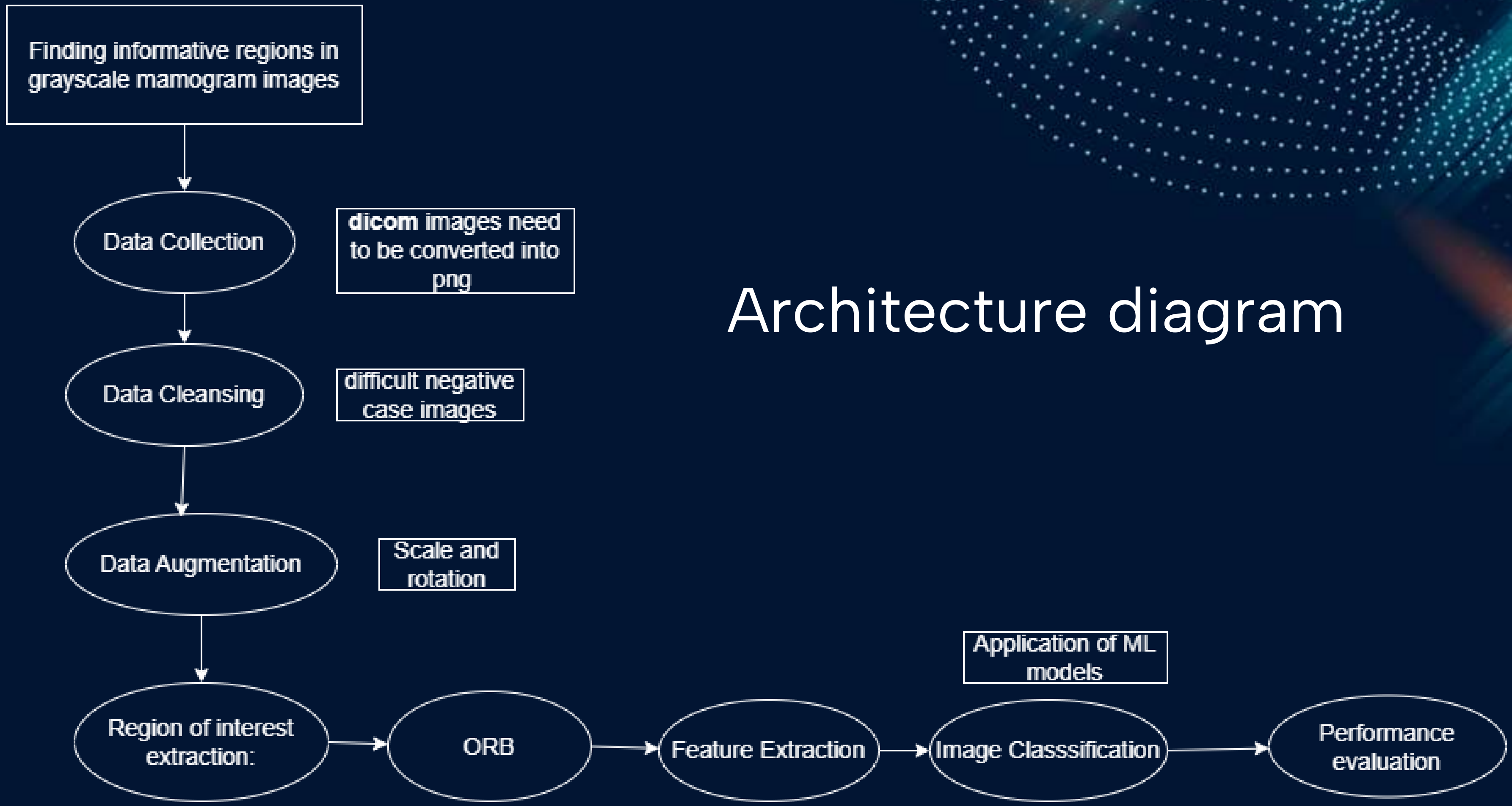
- ❑ are grayscale
- ❑ have bigger size
- ❑ have unnecessary dark background

Therefore, the traditional approaches with pretrained models may not deliver good results, since:

1. Pre-trained models expect 3 channel color images (RGB)
2. Images should be downsized to smaller size (224x224)
3. Classification models designed for multiclass classification

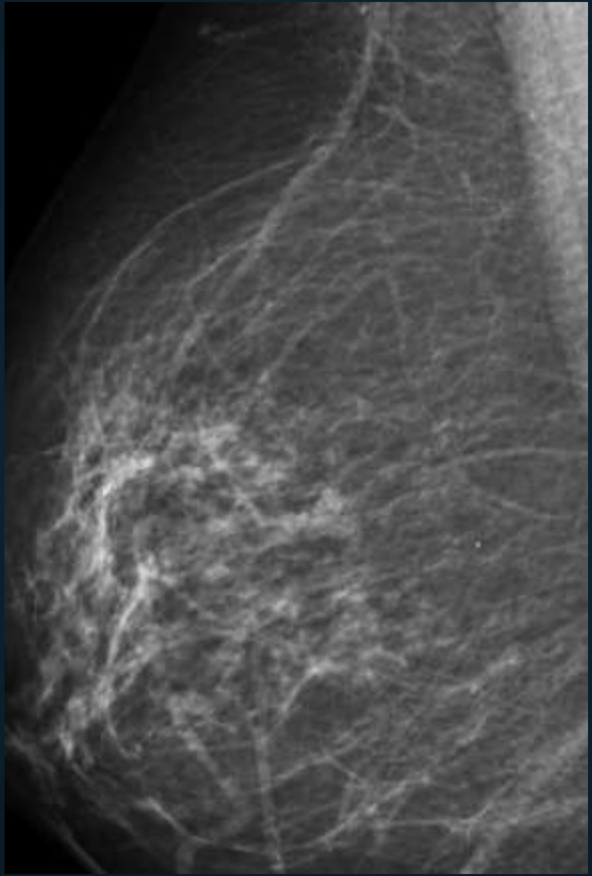
Consider this, we decide to build a custom approach for feature extraction and training.

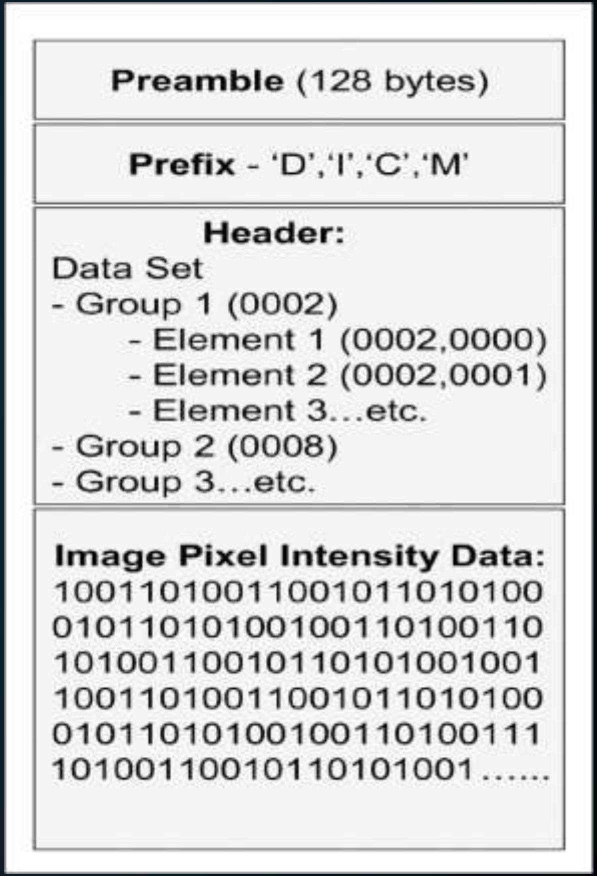
Architecture diagram



Challenges with image processing

Unlike standard image formats, DICOM files contain additional information such as patient data, imaging parameters, and acquisition details.





Data Collection

SOURCE of dataset: → RSNA Screening Mammography Breast Cancer

Malignant cancer

- **Positive** (1158)
- **Negative** (53548)
  - ~~**Difficult negative case**~~ (7705)
  - **Normal**

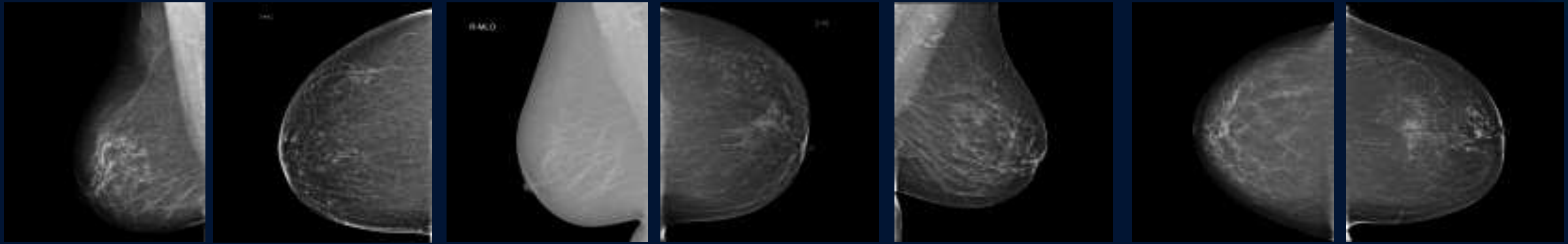
Data Selection

	Positive	Negative
Train	500	500
Test	518	518

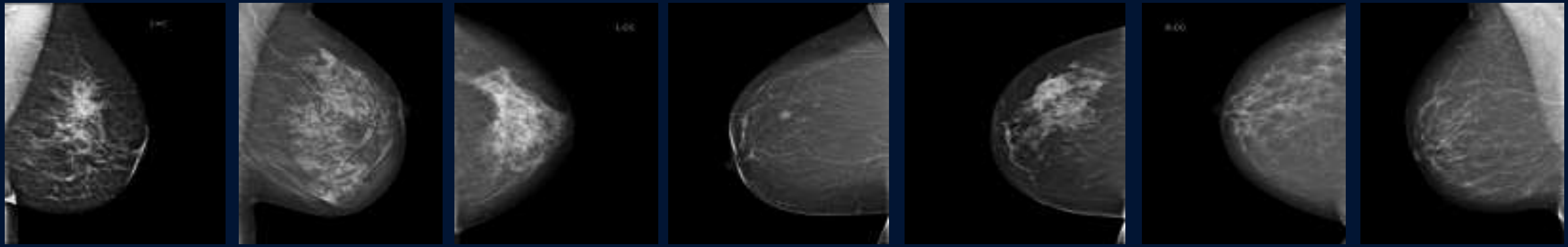
While data cleaning difficult negative cases were excluded. By excluding difficult negative cases, the goal is to create a more balanced and representative dataset that allows the model to focus on learning from the clear, well-defined negative instances

500/500	train
518/518	test

Positive cases



Negative cases

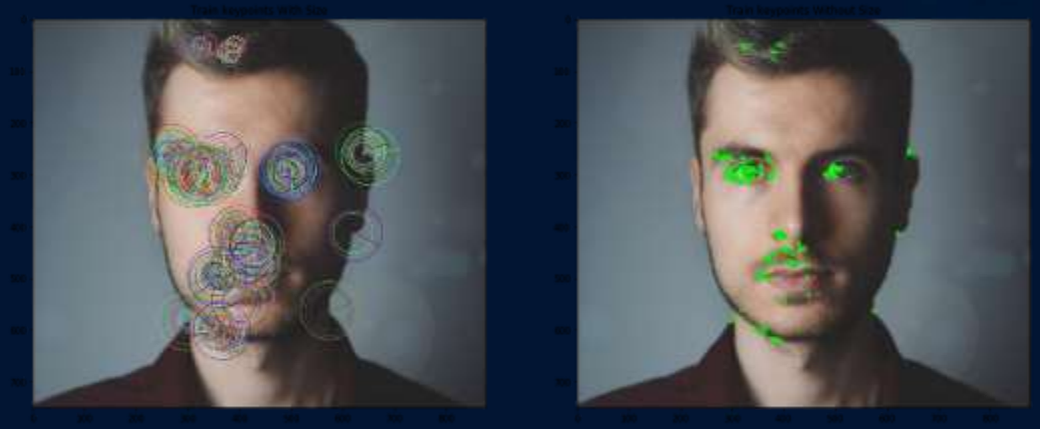


Feature extraction (ORB)

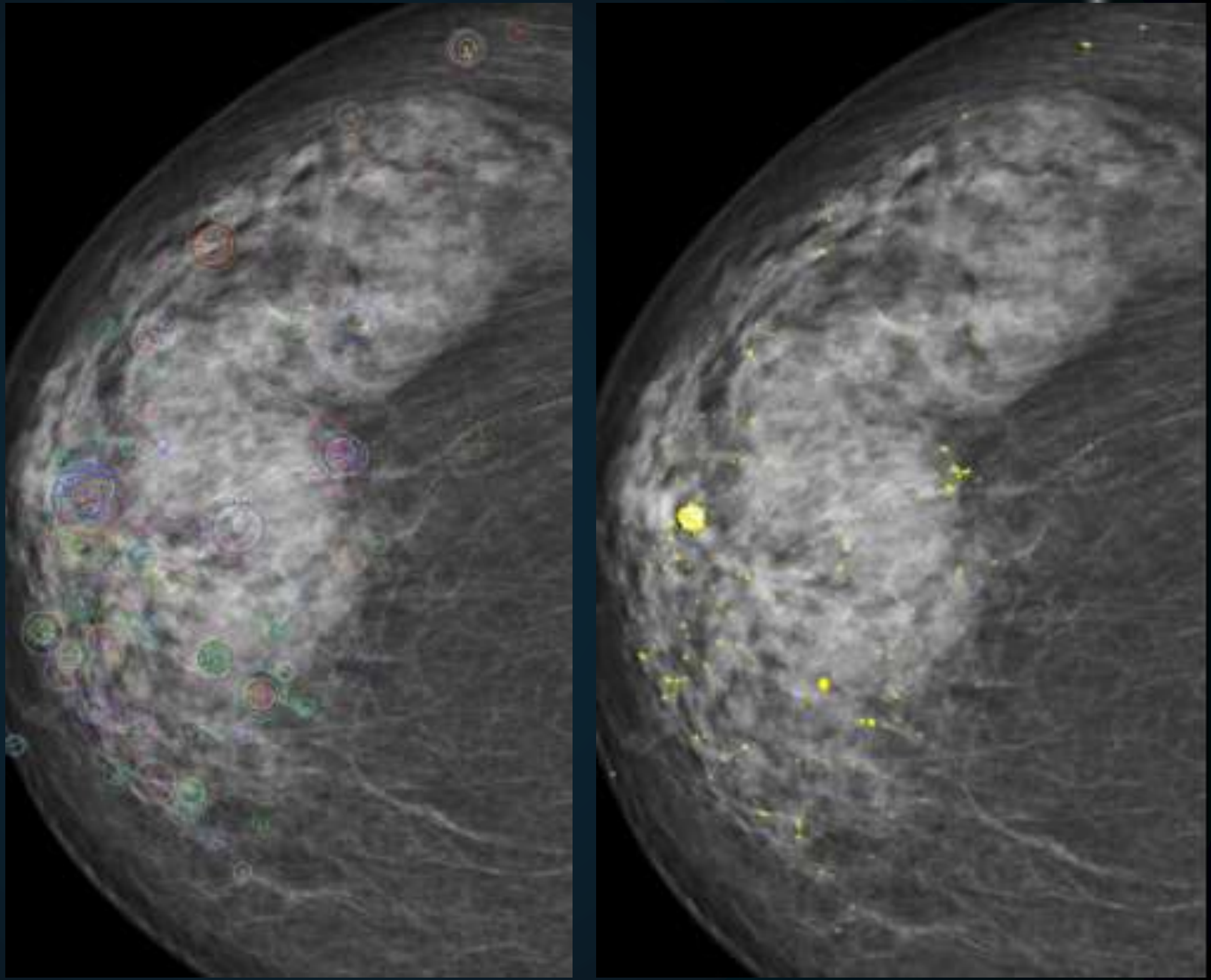
Oriented FAST and Rotated BRIEF (ORB) was developed at OpenCV labs in 2011, as an efficient and viable alternative to SIFT and SURF.

ORB performs as well as SIFT on the task of feature detection (and is better than SURF) while being almost two orders of magnitude faster.

ORB builds on the well-known FAST keypoint detector and the BRIEF descriptor.



ORB applied to breast cancer



Results

	Accuracy	Recall	Precision	F1 Score
Random Forest				
SVM				
KNN				
Logistic Regression				

Future work

Questions?