Computer Aided Detection and Diagnosis of Breast Cancer

Sanja Stanisic, matr. 800409

Contents

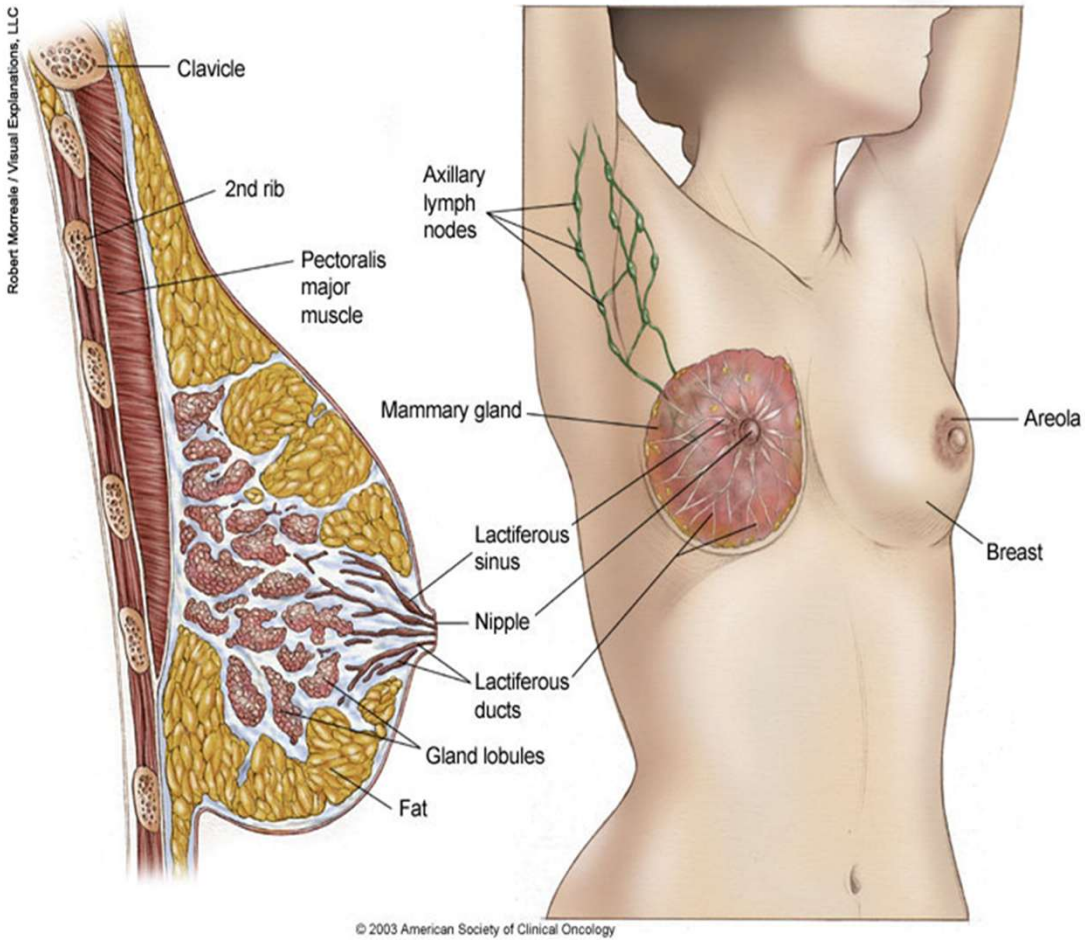
WHO - Breast Cancer Data

2020

There were **2,3** million women diagnosed with breast cancer and **685 000** deaths globally. As of the end of **2020**, there were **7.8** million women alive who were diagnosed with breast cancer in the past **5** years, making it the world's most prevalent cancer.

The strongest risk factor

Female gender is the strongest risk factor Only **0.5-1%** of breast cancers occur in men.



BREAST CANCER

- Breast cancer arises in the lining cells (epithelium) of the ducts (85%) or lobules (15%) in the glandular tissue of the breast.
- Initially, the cancerous growth is confined to the duct or lobule ("in situ") where it generally causes no symptoms and has minimal potential for spread (metastasis).
- Over time, these in situ (stage 0) cancers may progress and invade the surrounding breast tissue (invasive breast cancer) then spread to the nearby lymph nodes (regional metastasis) or to other organs in the body (distant metastasis). If a woman dies from breast cancer, it is because of widespread metastasis.

BREAST CANCER

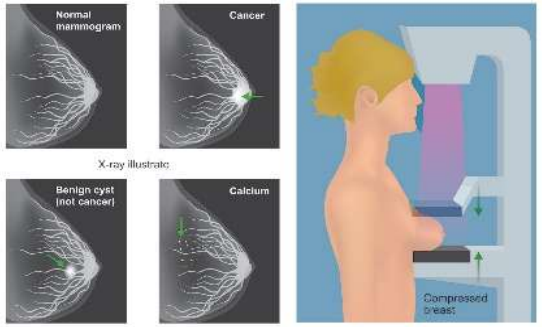
- The cancer's stage is calculated based on three clinical characteristics T, N, and M:
  - **T** - the size of the tumor and whether or not it has grown into nearby tissue
  - **N** – number of lymph nodes positive for malignancy
  - **M** – presence of distant metastasis

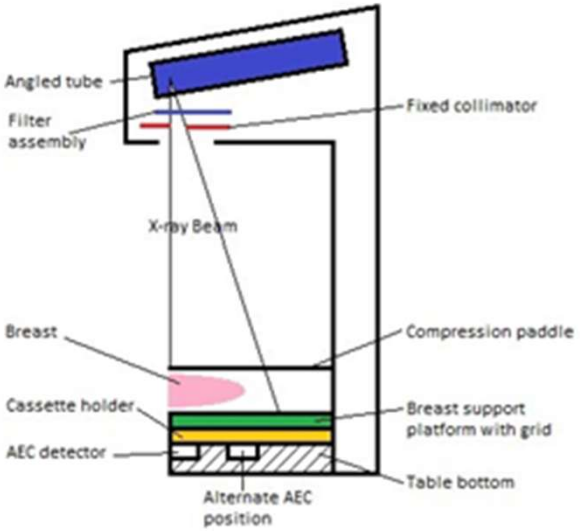
- **Treatment**: surgery, radiotherapy and medication (chemotherapy, hormonal and/or targeted biological therapy)

- **Early detection** and state-of-the-art cancer treatment are the most important strategies to prevent deaths from breast cancer

- Getting regular **screening tests** is the most reliable way to detect breast cancer early.

Computer Aided Detection and Diagnosis of Breast Cancer





- **Mammography** is the most efficient tool to help detect breast cancer, especially at its earliest stage. It is a process of using **low-energy X-rays** to examine the human breast and identifying abnormalities, typically through detection of characteristic masses or microcalcifications.

- In order to standardize mammography reports, American College of Radiology (ACR) introduced **Breast Imaging-Reporting and Data System, BI-RADS**. BI-RADS is a quality control system, but in day-to-day usage this term refers to the mammography assessment categories. It is used by medical professionals to evaluate the patient's risk of developing breast cancer.

MAMMOGRAPHY

BI-RADS Assessment Categories

0: Incomplete

1: Negative

2: Benign

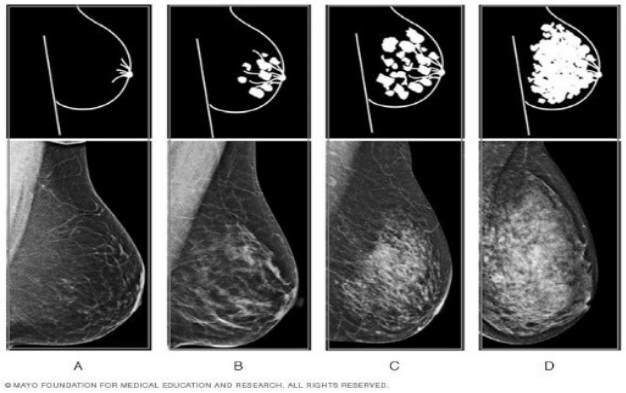
3: Probably benign

4: Suspicious

5: Highly suggestive of malignancy

6: Known biopsy – proven malignancy

BI-RADS Breast Density Levels

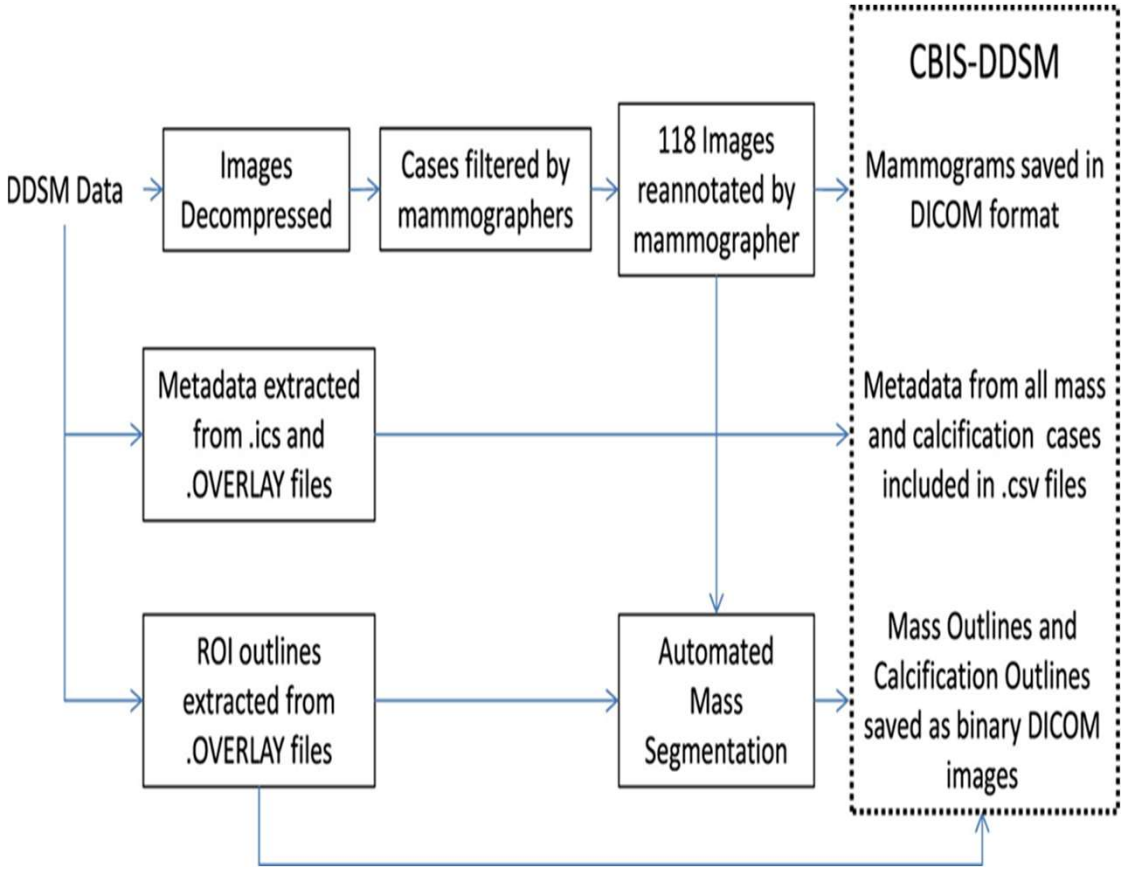


A. Mostly fatty
B. Scattered fibroglandular density
C. Heterogeneously dense
D. Extremely dense

MATERIALS: DATASET

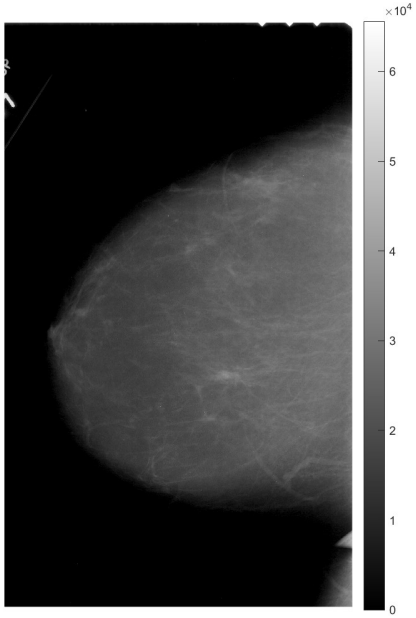
- The dataset used in this project is CBIS-DDSM (Curated Breast Imaging Subset of DDSM) available at https://wiki.cancerimagingarchive.net/display/Public/CBIS-DDSM
- It is an updated and standardized version of the Digital Database for Screening Mammography (DDSM), a database of **2.620** **scanned film mammography studies.**
- It contains normal, benign, and malignant cases with verified pathology information.
- This collection is structured such that each participant has multiple patient IDs.
- The data set contains **753** calcification cases and **891** mass cases..

UPGRADING THE ORIGINAL DATABASE

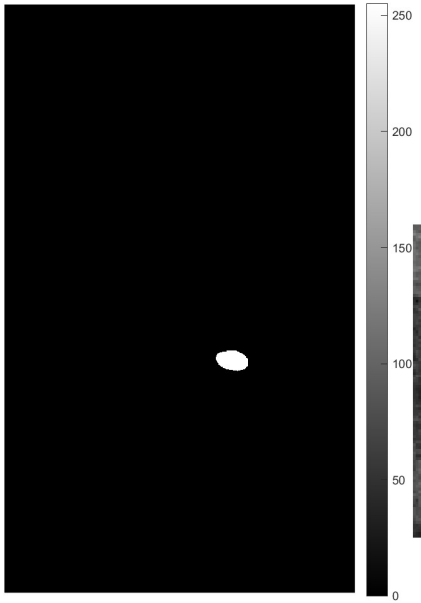


DATASET: IMAGES & .CSV METADATA FILES

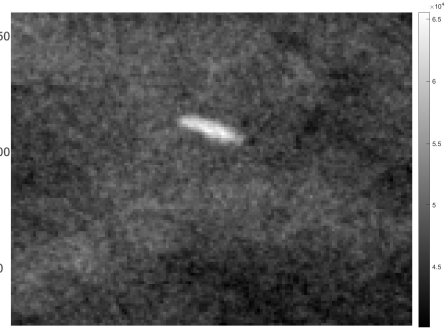
Full mammography



ROI



Cropped Image



.csv

- Patient ID: the first 7 characters of images in the case file
- Density category
- Breast: Left or Right
- View: craniocaudal (CC) or mediolateral-oblique (MLO)
- Number of abnormalities for the image (This is necessary as there are some cases containing multiple abnormalities.
- Mass shape (when applicable)
- Mass margin (when applicable)
- Calcification type (when applicable)
- Calcification distribution (when applicable)
- BI-RADS assessment
- Pathology: Benign, Benign without call-back, or Malignant
- Subtlety rating: Radiologists' rating of difficulty in viewing the abnormality in the image
- Path to image files

METHODS

Dataset pre-processing

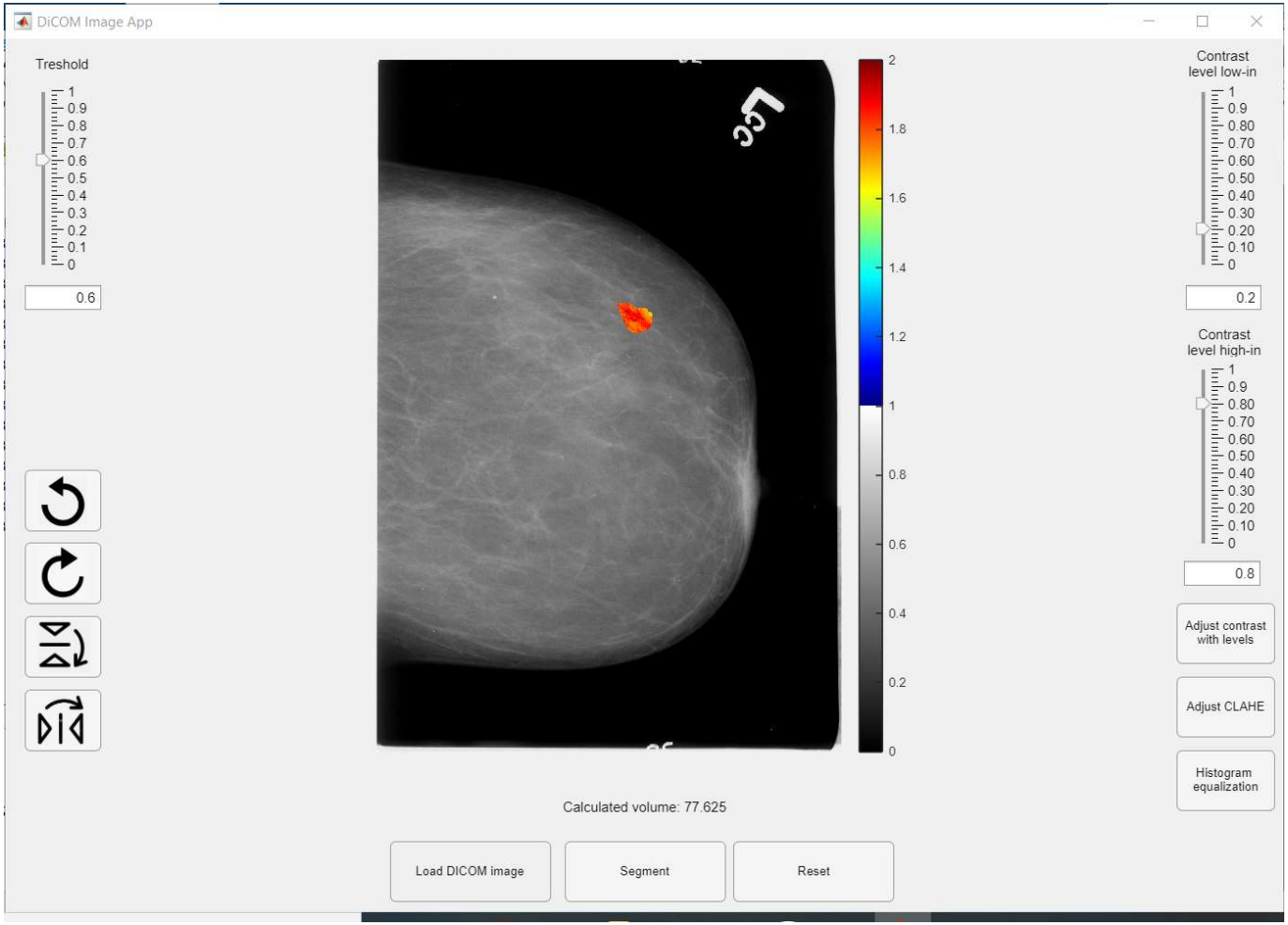
Given the size of the dataset, in this project only the data related to **calcifications** were analysed. There were **753** calcification cases. As requested the data relating to cases classified as "benign with call-back" were removed, downsizing the dataset to **1311** instances, i.e. 1311 images where the dimensions of the ROI mask corresponded to the dimensions of the full images.

Pixel size

The images in this datasets are in DICOM format, but they are digitized analog images and not digital images. Therefore the information about pixel size, pixel spacing and slice thickness were not provided. However, this database was analyzed in many papers, so as suggested in scientific literature[1] the **pixel size** was set at **1 pixel = 50 µm**, while **slice thickness** was set at **1 mm** as in digital tomosynthesis.

[1] Ragab DA, Sharkas M, Marshall S, Ren J. 2019. Breast cancer detection using deep convolutional neural networks and support vector machines. PeerJ 7:e6201

GUI APPLICATION



MACHINE LEARNING TECHNIQUES IN BREAST CANCER DETECTION AND DIAGNOSIS

- Detection of subclinical breast cancer on screening mammography is challenging as **an image classification** task because the tumors themselves occupy only a small portion of the image of the entire breast. For example, a full-field digital mammography image is typically 4000×3000 pixels while a potentially cancerous region of interest (ROI) can be as small as 100×100 pixels.
- Challenge: automated segmentation of ROI. This dataset contains automatically segmented ROI masks.
- These mask were overlaid over full mammography images to extract potential lesions. Following morphological measurements of the lesions were calculated:

  Lesion Volume, **Lesion Area**, **Spherical Disproportion**, **Sphericity**, and **Surface to Volume Ratio**

CLASSIFICATION MODELS

Binary Classification

This is a binary classification problem with classes being **BENIGN** and **MALIGNANT.**

Feature Selection

Classification based only on morphological measurements is compared to the classification which includes both morphological measurements and features obtained from metadata: **breast density**, **BI RADS assessment**, **calcification type** and **calcification distribution.**

Applied Methods

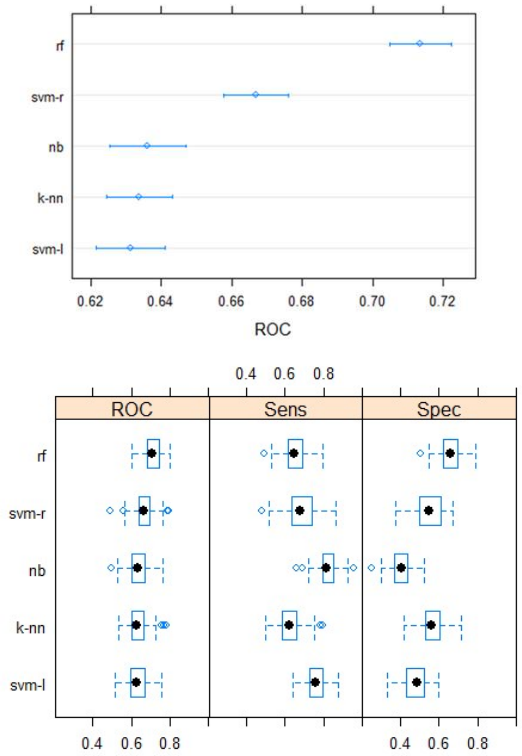
The following classification methods were applied: **kNN**, **Naïve Bayes**, **SVM with linear kernel**, **SVM with radial kernel** and **Random Forest.**

Validation

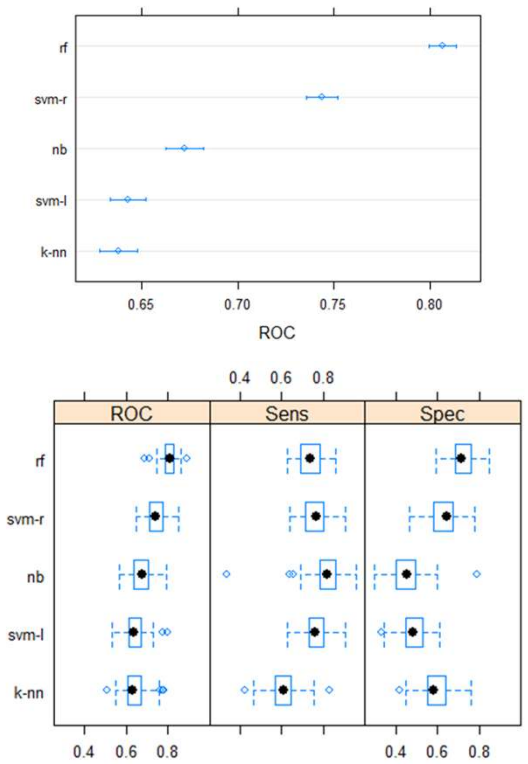
10 - fold cross-validation was applied to evaluate the performance of the classifiers.

RESULTS

Only morphological features

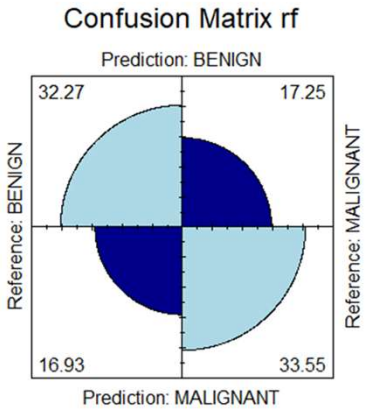


All features



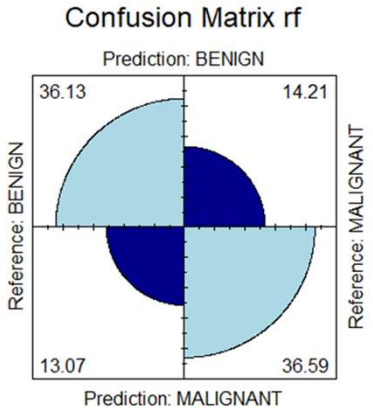
RESULTS – RANDOM FOREST

Only morphological features



Specificity	0.6603604
Sensitivity	0.655814
Accuracy	0.6581236

All features



Specificity	0.7202703
Sensitivity	0.7344186
Accuracy	0.7272311

CONCLUSIONS

1. Classification is significantly better when the features from metadata are included in the process.
2. This classification is not good enough to be used for clinical purposes, not even for screening mammographies.
3. In terms of pathology classification, scientific papers treating the same dataset didn't obtain significantly better results, apart from specificity. According to the open literature the best classification model is convolutional neural networks.

BIBLIOGRAPHY

Lee, R., Gimenez, F., Hoogi, A. et al. A curated mammography data set for use in computer-aided detection and diagnosis research. Sci Data 4, 170177 (2017).

Ragab DA, Sharkas M, Marshall S, Ren J. 2019. Breast cancer detection using deep convolutional neural networks and support vector machines. PeerJ 7:e6201

Kupersmith, J., Breast Cancer Detection Algorithm
https://jordankupersmith.github.io/its_not_a_tumor/projectdetails/

https://www.who.int/news-room/fact-sheets/detail/breast-cancer

https://www.cancer.net/cancer-types/breast-cancer

https://wiki.cancerimagingarchive.net/display/Public/CBIS-DDSM

https://senologiadiagnostica.it/mammografia/