

PrediCare

AI Breast Cancer Diagnosis

Capstone Project Presentation
22.11.2023



OUR TEAM



Helene Schönewolf

Entrepreneur |
Data Scientist



María José Zapata

Physiotherapist |
Data Scientist



Yen Ramisch

PhD Analytical Chemist |
Data Scientist

TABLE OF CONTENTS

01. BREAST CANCER & DIAGNOSIS

02. DATA & BASELINE MODEL

03. IMAGE SEGMENTATION & CLASSIFICATION

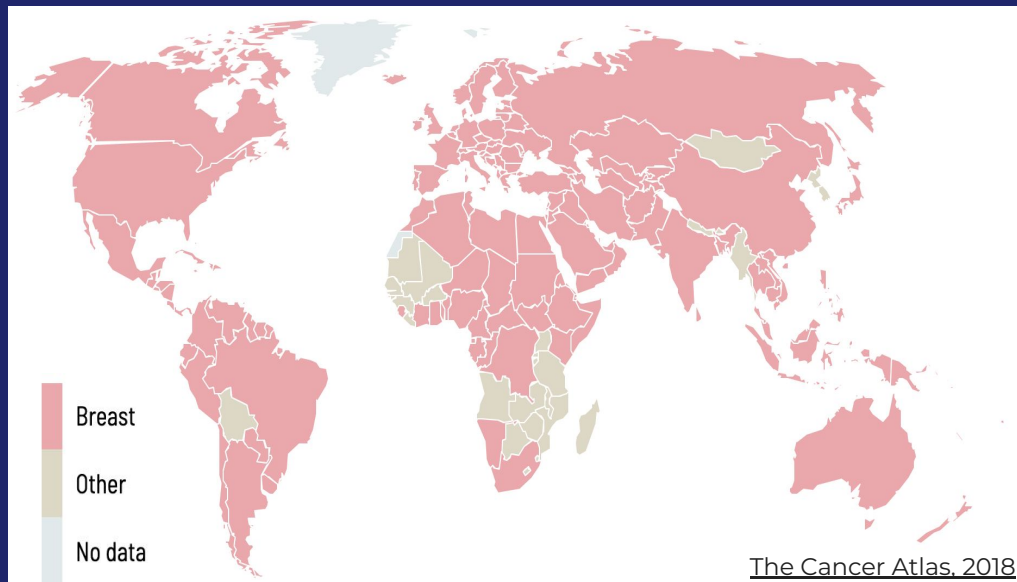
04. ACCESSIBILITY (WEB APP)



01. BREAST CANCER & DIAGNOSIS



Breast Cancer – Most common Cancer in Women worldwide



2.3 M women diagnosed globally with breast cancer (2020)

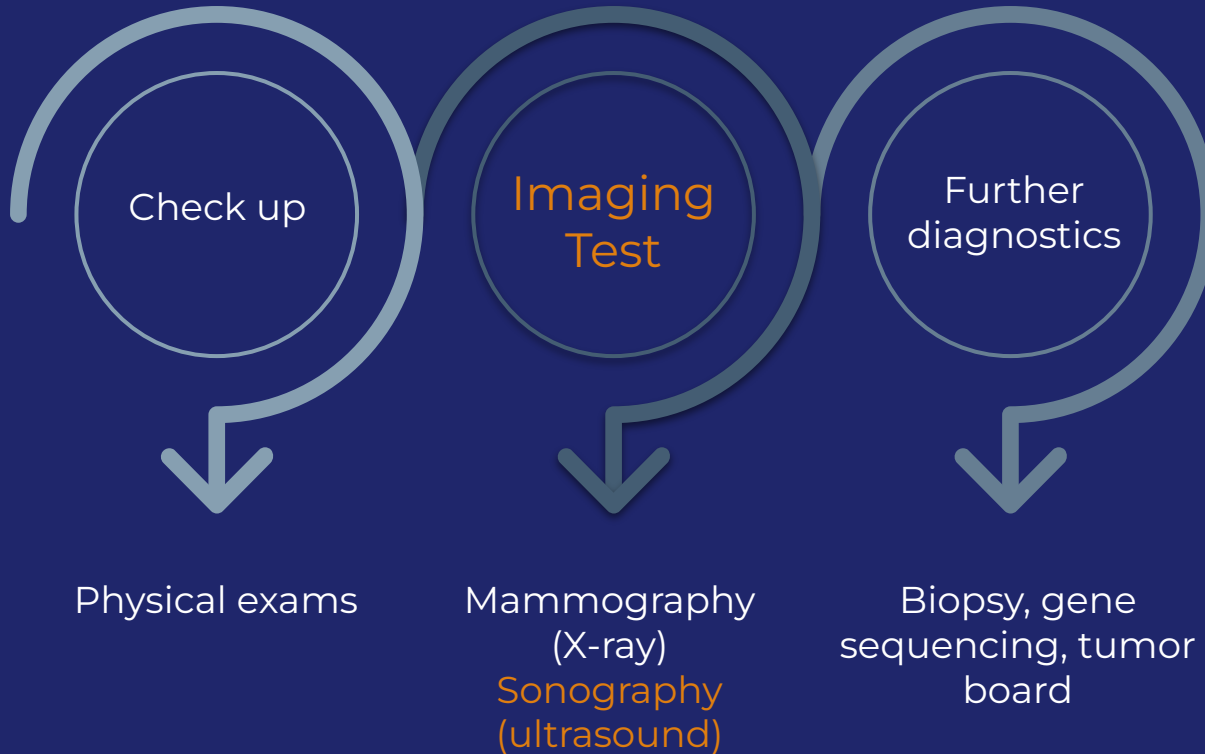
20 % missed breast cancer diagnoses during check ups

30 % global death rate

Reduce mortality by:

- **health promotion** for early detection
- **early diagnosis**

Breast Cancer - Early Detection



X-Ray vs. Ultrasound Imaging

	X-ray	Ultrasound
Advantages	Standard procedure	Differentiation of cystic/fluid and solid lesions Non ionizing radiation, safe fore repeated examinations No breast compression
Disadvantages	In many countries no accessibility Harmful radiation	Need much more practice Difficult to interpret

Ultrasound Image Analysis: normal, benign, malignant

	normal	benign	malignant
Shape & Margins	consistent, symmetrical appearance	regular, well-defined borders of mass	irregular, poorly-defined, spiculated mass
Echogenicity	uniform echo	homogeneous, similar to surrounding tissue	heterogeneous
Shadowing	not present	more pronounced	less pronounced

1. Identification of anomalies (node/tumor)
2. Inspecting the shape of node/tumor
3. Inspecting the texture of the node/tumor (in comparison to the surrounding tissue)

Objective

Identifying anomalies and predicting malignancy of breast tumors based on breast ultrasound images



Supporting doctors in early detection of breast cancer

Providing patients with a tool to help them detect breast cancer earlier



Improve survival rate



02. DATA & BASELINE MODEL



Original Data Set - BUSI dataset

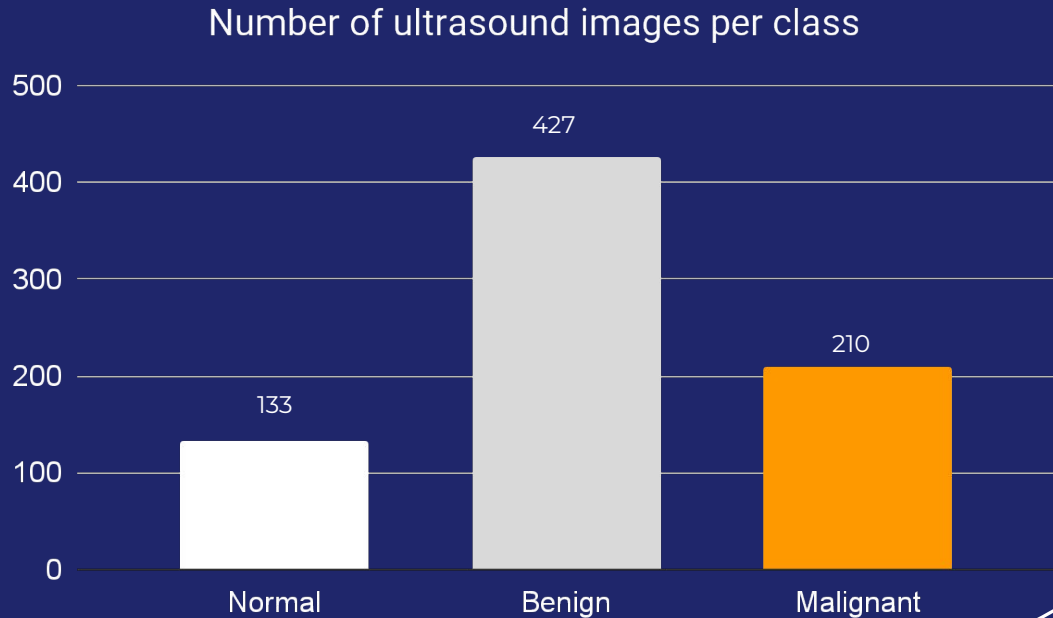
Ultrasound images of women's breast collected in 2018

600 female patients (age 25-75)

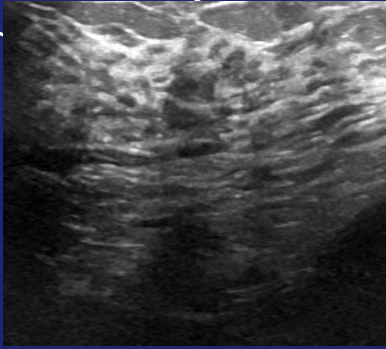
780 images (500 x 500, png)

3 categories:

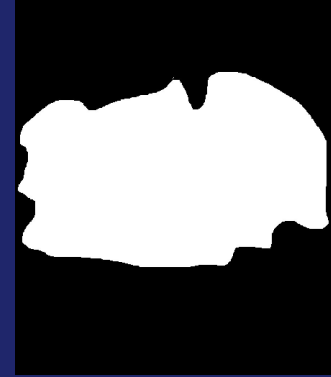
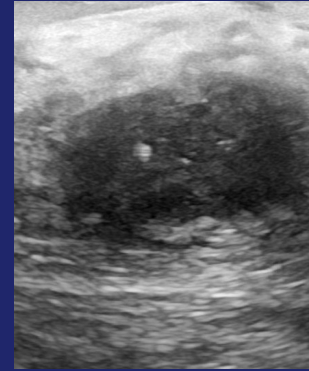
- Normal: 133 (17,05 %)
- Benign: 427 (54,74 %)
- Malignant: 210 (26,92 %)



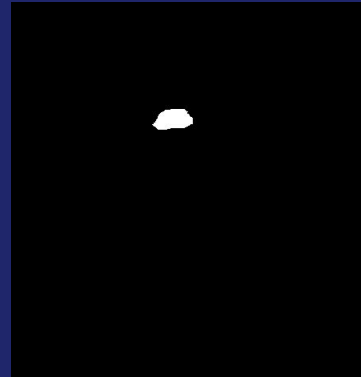
What the data set looks like



Normal



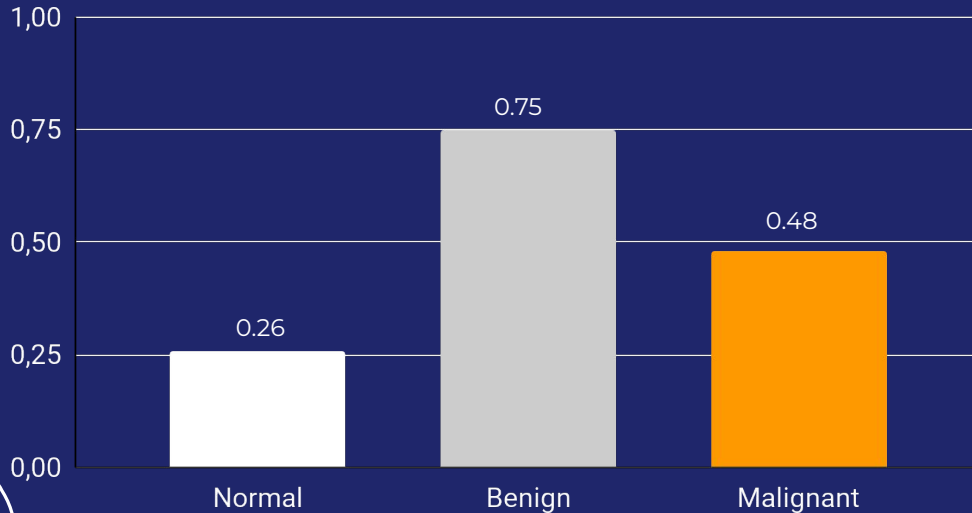
Malignant



Benign

Baseline Model: Support Vector Machine (SVM)

F-1 score SVM baseline model



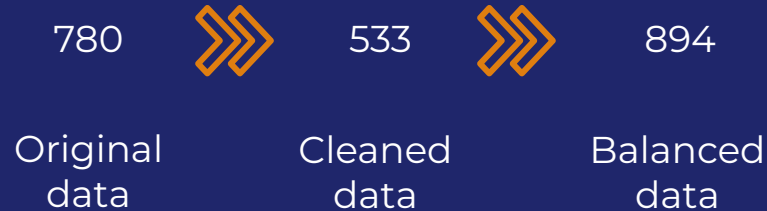
780 images original data set

Accuracy: 0.69

- Normal
- Benign
- Malignant

Data cleaning

1. Deleting misclassified images, duplicates and images of axilla
2. Creating unique masks on those images that have more than one mask
3. Data augmentation (rotation, translation, shear, zoom) to create balanced data set



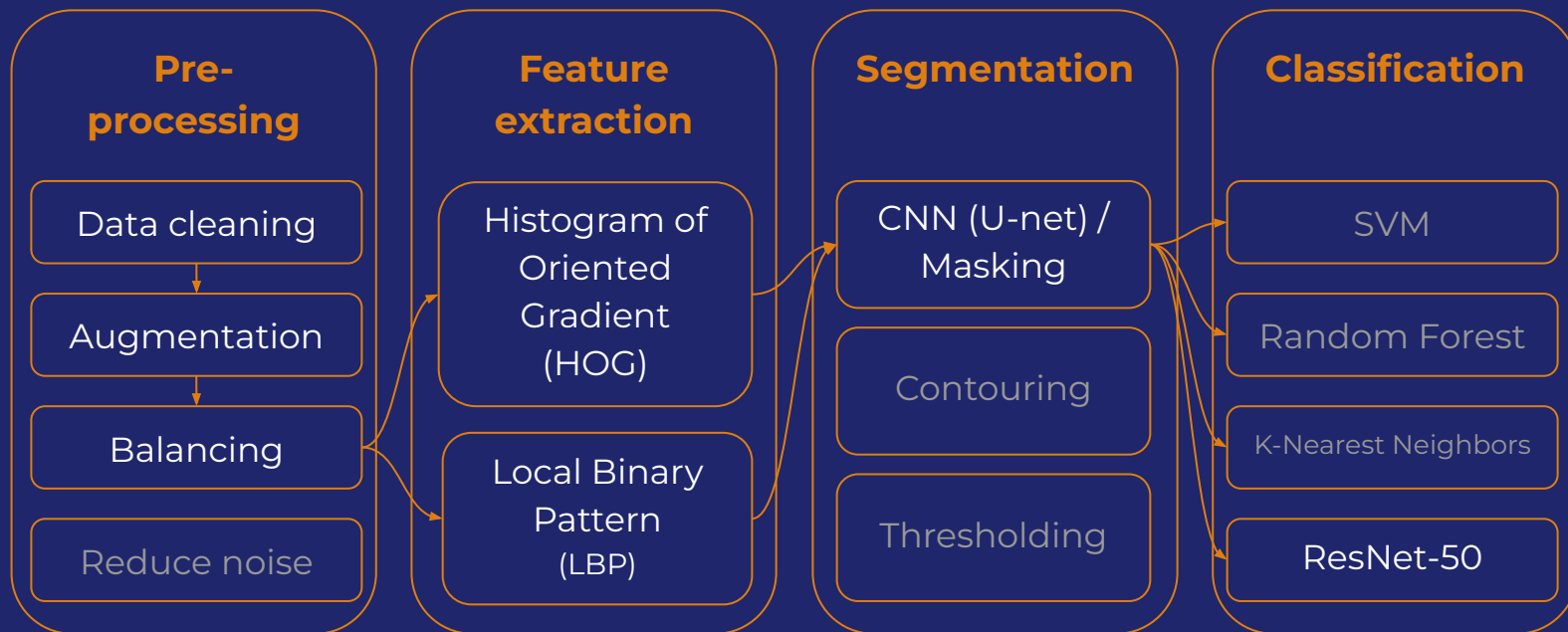


IMAGE

03. SEGMENTATION & CLASSIFICATION



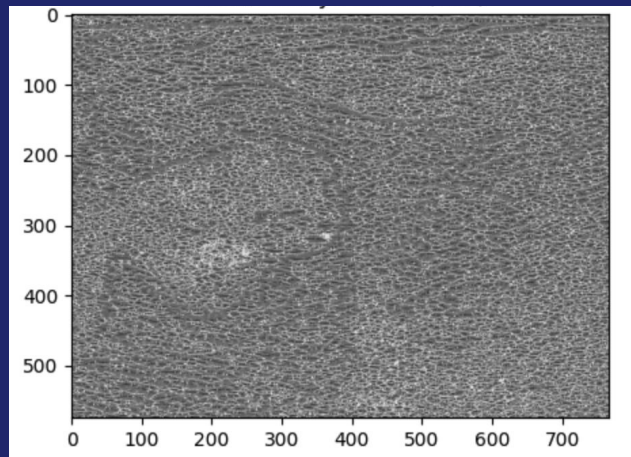
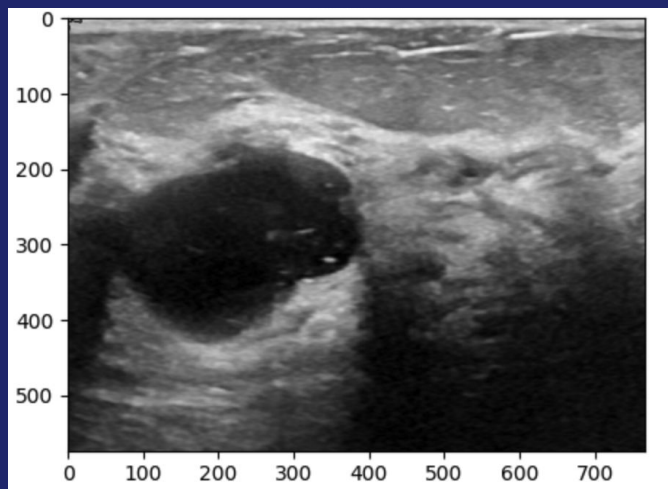
Image Recognition Process



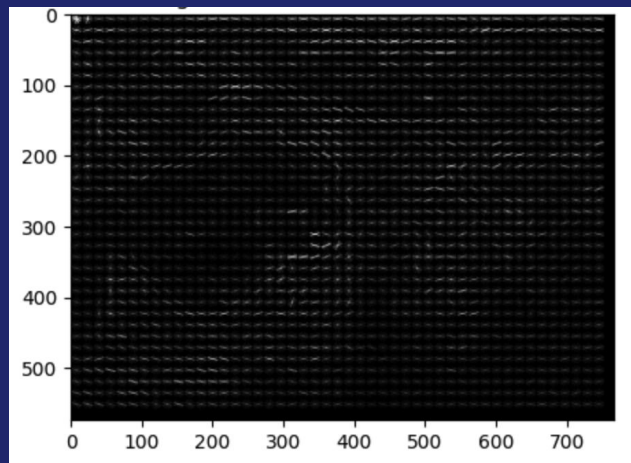
Feature Extraction

Image texture/characteristics

Original image



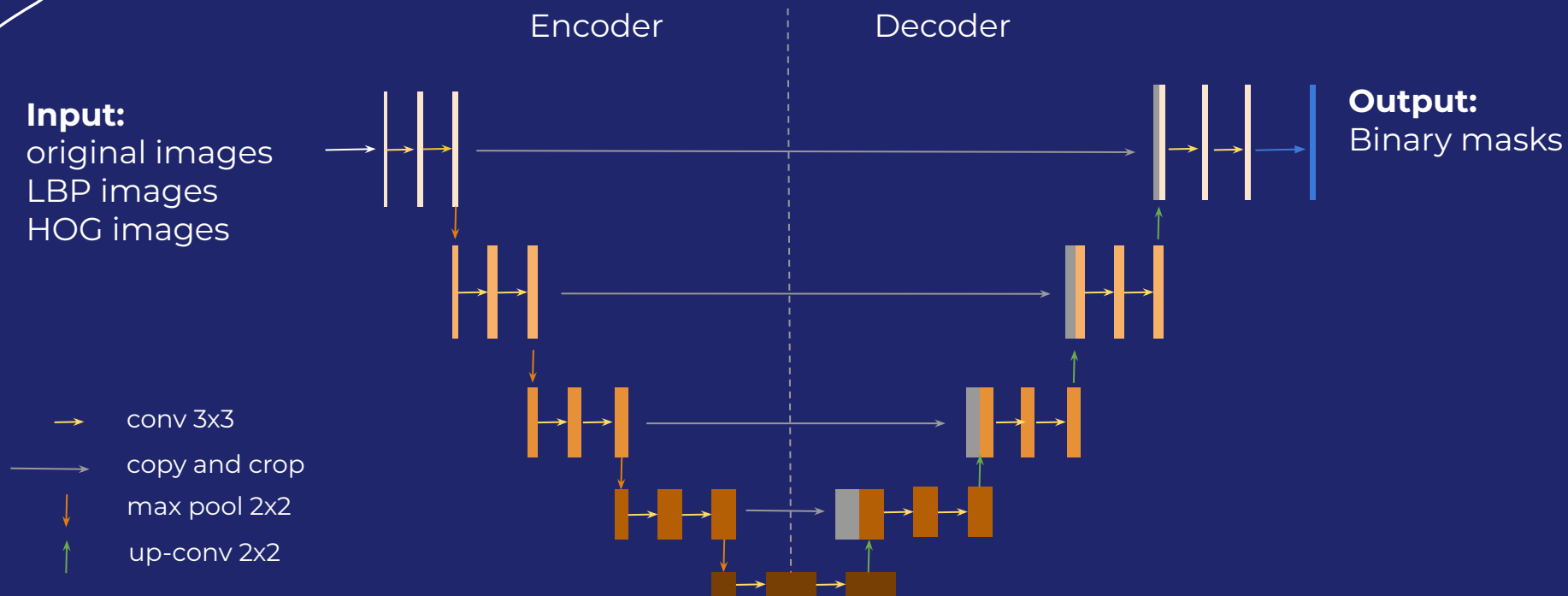
Local Binary Pattern (LBP)

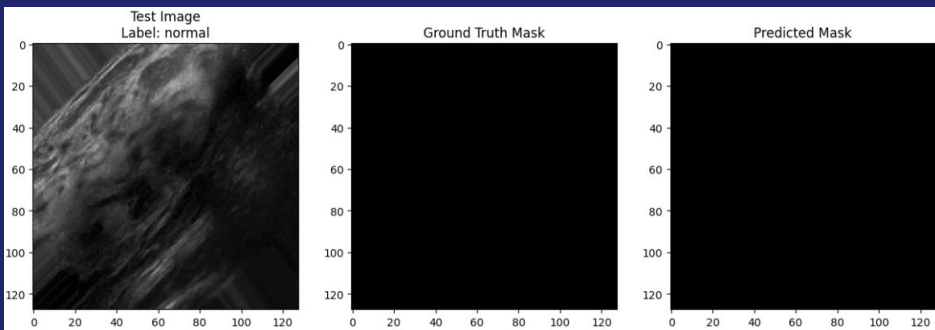
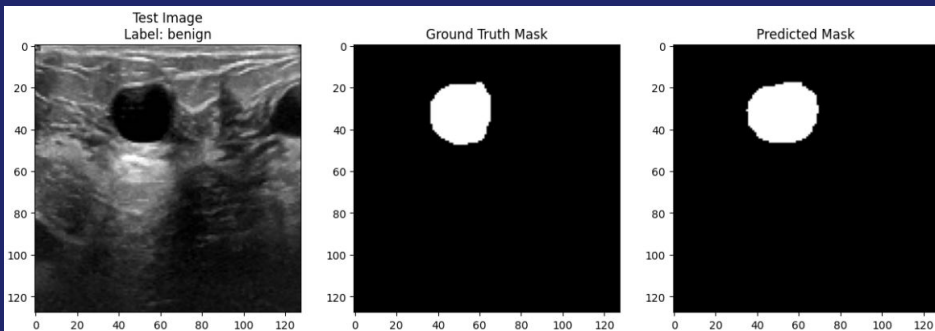
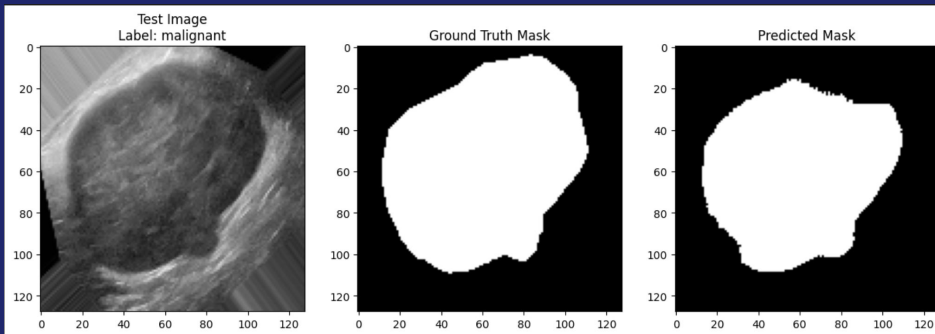


Histogram of Oriented Gradients (HOG)

Image segmentation - U-net

Identify anomalies and determine their shapes





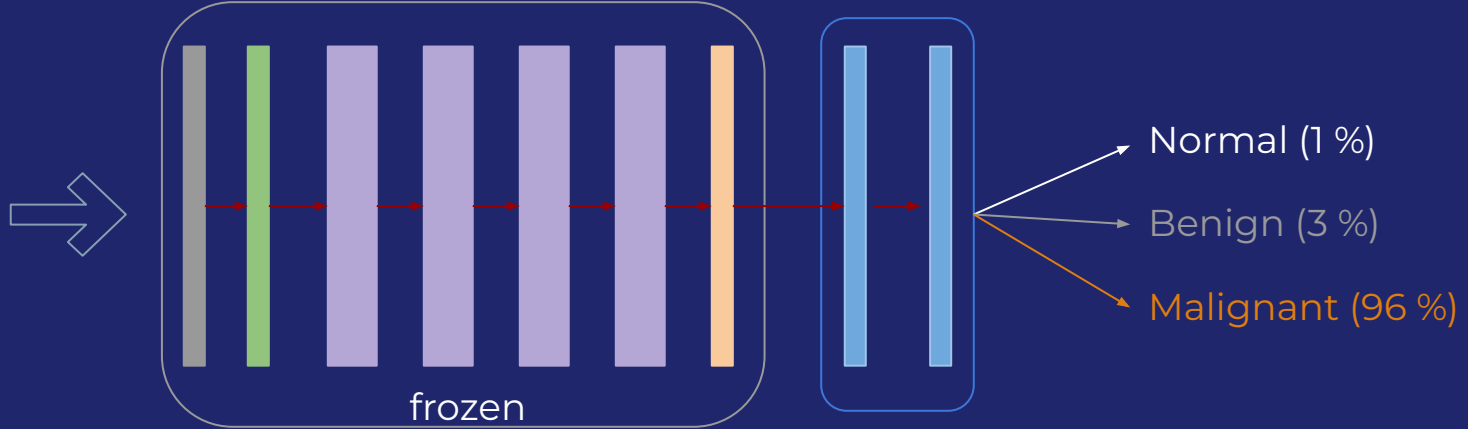
Intersection over
Union (IoU):

0.77

Image classification: Transfer Learning

Input:

original images
LBP images
HOG images
Binary masks



- Conv
- Max Pool
- Residual block
- Average Pool
- Dense & Dropout

ResNet50
(Residual Neural Network
with 50 layers)

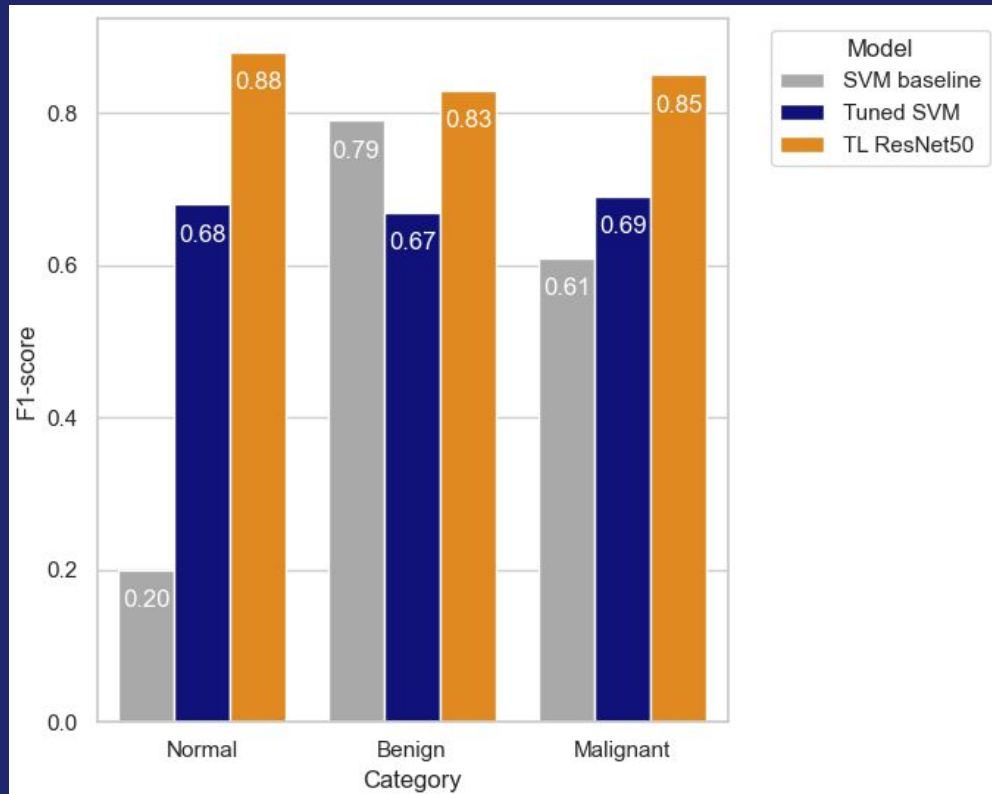
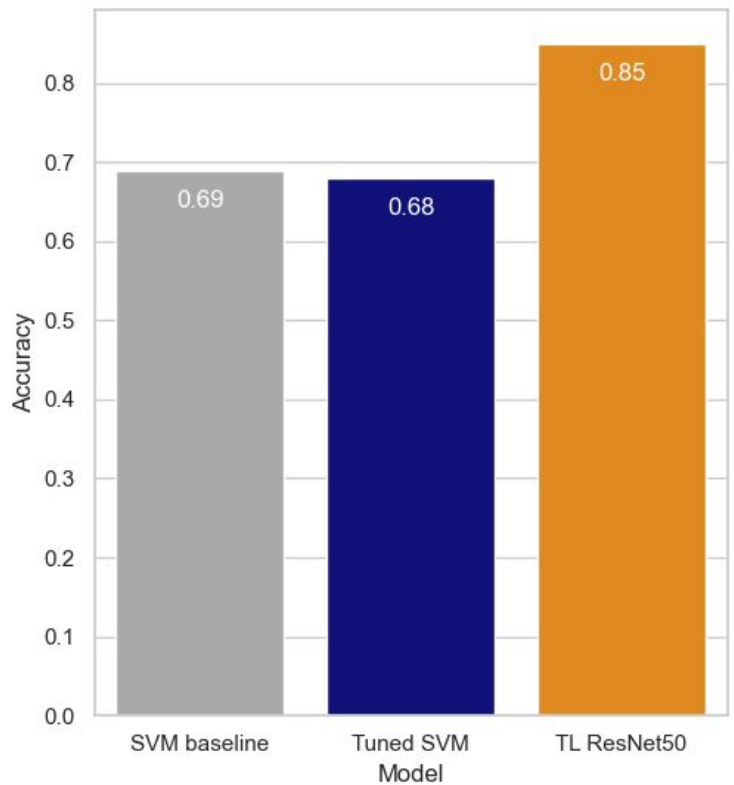
New dense layers

Normal (1 %)

Benign (3 %)

Malignant (96 %)

Evaluation of the models





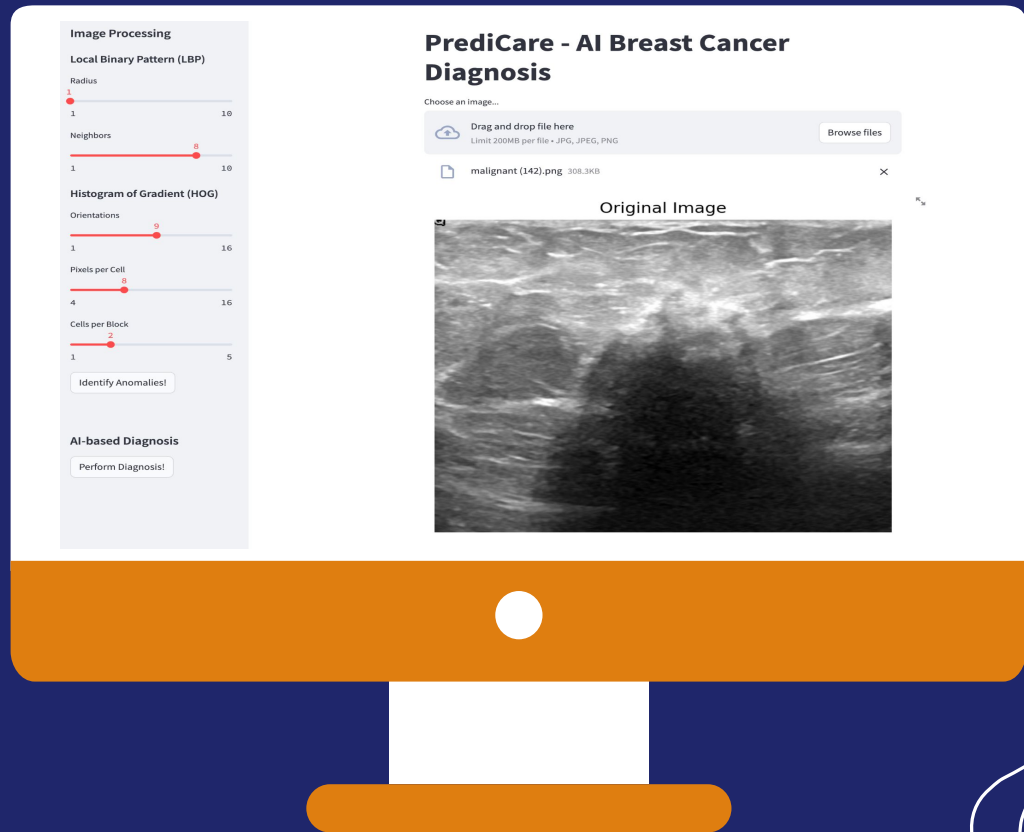
04. ACCESSIBILITY



Web app

PrediCare

AI Breast Cancer Diagnosis



THANK YOU FOR THE ATTENTION!

Do you have any questions?

