Artificial Intelligence in Medicine

An explainable unified approach for mass detection and classification in DBT

Andrea Berti

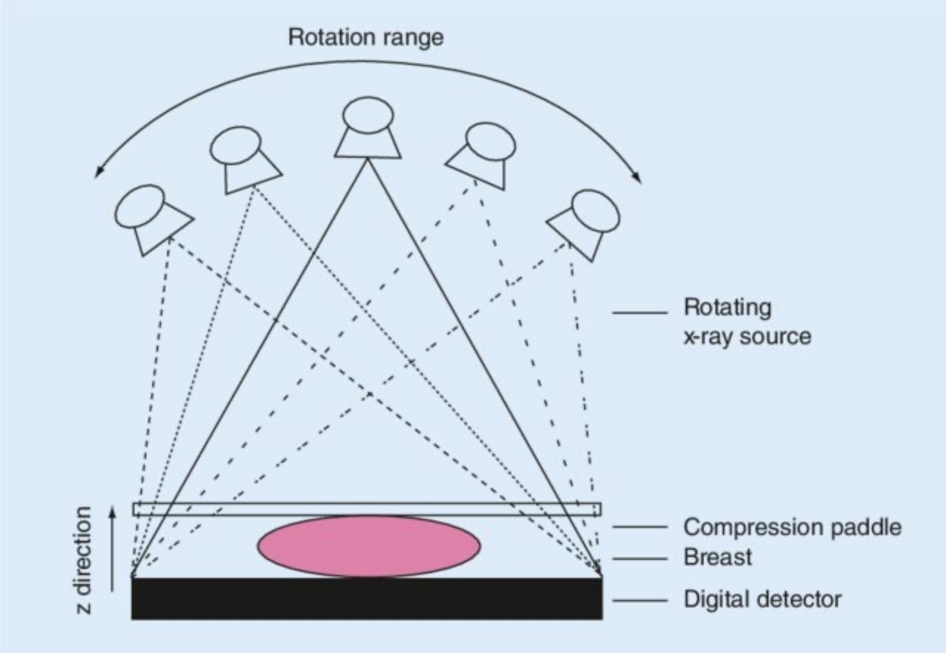
INFN

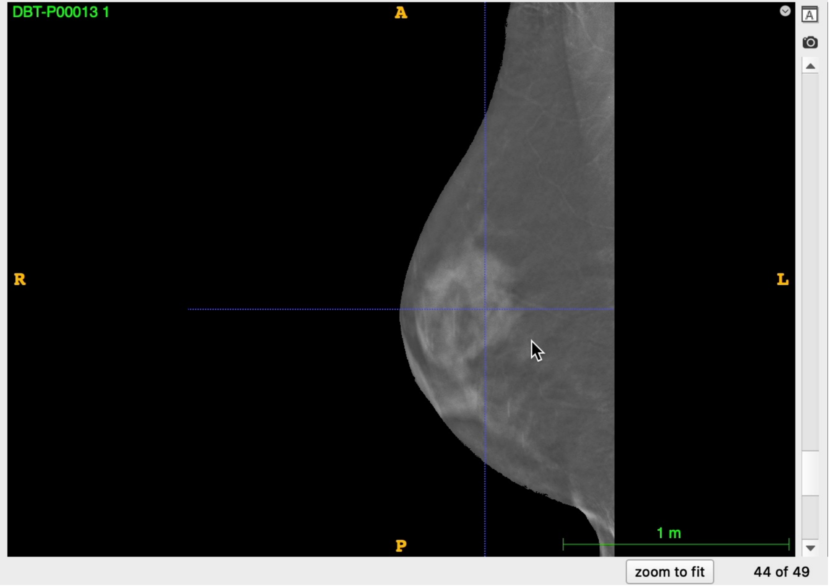
Istituto di Scienza e Tecnologie dell'Informazione "A. Faedo"
Consiglio Nazionale delle Ricerche

Overview

- Digital Breast Tomosynthesis
- The datasets: BCS-DBT & EMBED
- Proposed framework
  - Deep Learning
  - Radiomics
- Explainability in classification

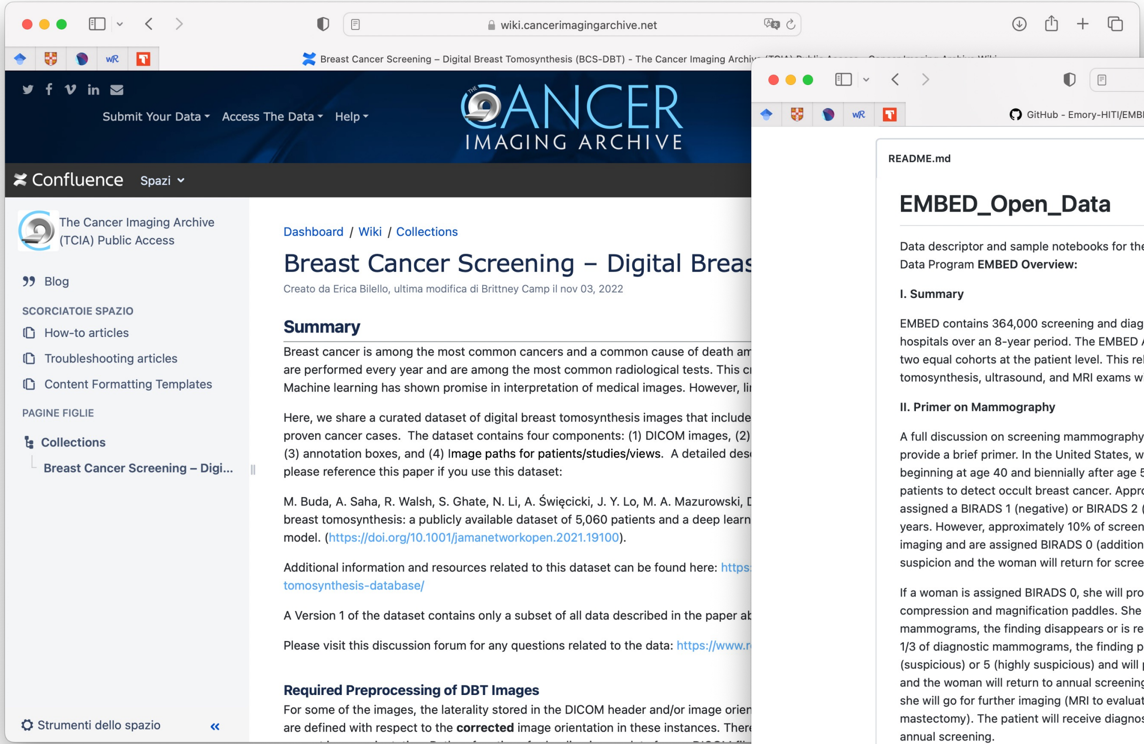
Digital Breast Tomosynthesis





Digital Breast Tomosynthesis

https://wiki.cancerimagingarchive.net/pages/viewpage.action?pageId=64685580



wiki.cancerimagingarchive.net

Breast Cancer Screening – Digital Breast Tomosynthesis (BCS-DBT) - The Cancer Imaging Archiv...

Submit Your Data ▾ Access The Data ▾ Help ▾

CANCER IMAGING ARCHIVE

Confluence Spazi ▾

The Cancer Imaging Archive (TCIA) Public Access

Blog

SCORCIATOIE SPAZIO

How-to articles

Troubleshooting articles

Content Formatting Templates

PAGINE FIGLIE

Collections

Breast Cancer Screening – Digi...

Strumenti dello spazio

Dashboard / Wiki / Collections

Breast Cancer Screening – Digital Breas

Creato da Erica Bilello, ultima modifica di Brittney Camp il nov 03, 2022

Summary

Breast cancer is among the most common cancers and a common cause of death am
are performed every year and are among the most common radiological tests. This c
Machine learning has shown promise in interpretation of medical images. However, li

Here, we share a curated dataset of digital breast tomosynthesis images that include
proven cancer cases. The dataset contains four components: (1) DICOM images, (2)
(3) annotation boxes, and (4) Image paths for patients/studies/views. A detailed des
please reference this paper if you use this dataset:

M. Buda, A. Saha, R. Walsh, S. Ghate, N. Li, A. Święcicki, J. Y. Lo, M. A. Mazurowski, D
breast tomosynthesis: a publicly available dataset of 5,060 patients and a deep learn
model. (https://doi.org/10.1001/jamanetworkopen.2021.19100).

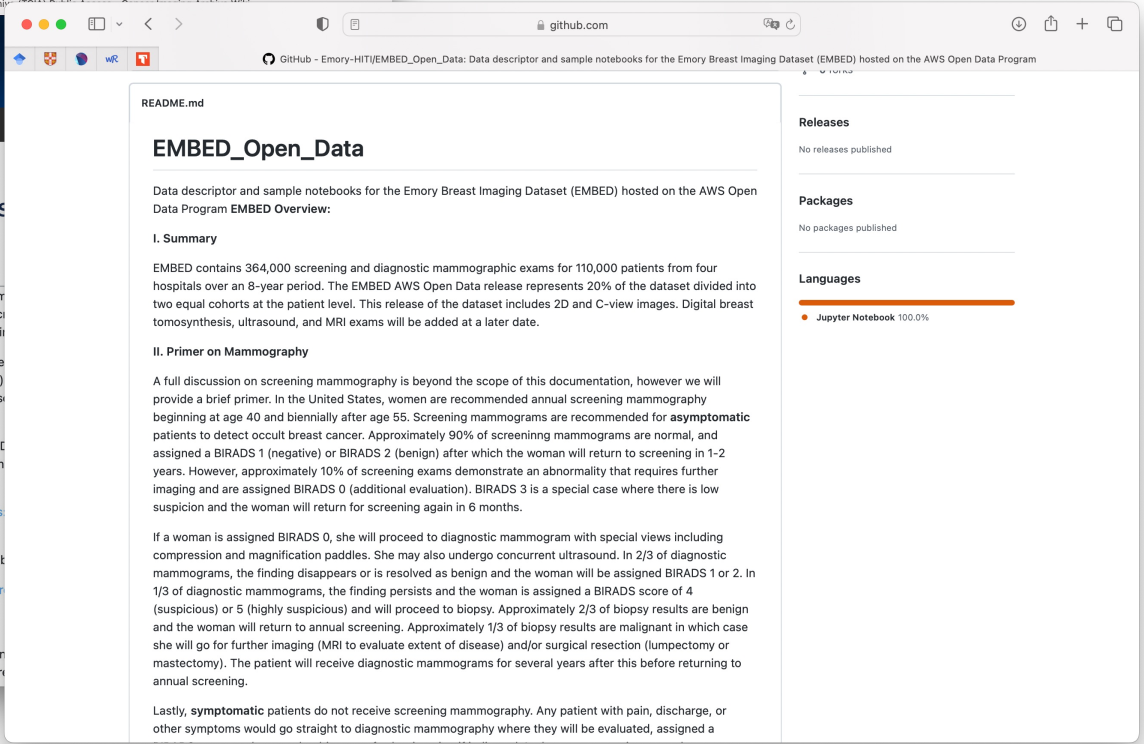
Additional information and resources related to this dataset can be found here: https
tomosynthesis-database/

A Version 1 of the dataset contains only a subset of all data described in the paper ab

Please visit this discussion forum for any questions related to the data: https://www.r

Required Preprocessing of DBT Images

For some of the images, the laterality stored in the DICOM header and/or image orien
are defined with respect to the **corrected** image orientation in these instances. There



github.com

GitHub - Emory-HITI/EMBED_Open_Data: Data descriptor and sample notebooks for the Emory Breast Imaging Dataset (EMBED) hosted on the AWS Open Data Program

README.md

EMBED_Open_Data

Data descriptor and sample notebooks for the Emory Breast Imaging Dataset (EMBED) hosted on the AWS Open Data Program **EMBED Overview:**

I. Summary

EMBED contains 364,000 screening and diagnostic mammographic exams for 110,000 patients from four hospitals over an 8-year period. The EMBED AWS Open Data release represents 20% of the dataset divided into two equal cohorts at the patient level. This release of the dataset includes 2D and C-view images. Digital breast tomosynthesis, ultrasound, and MRI exams will be added at a later date.

II. Primer on Mammography

A full discussion on screening mammography is beyond the scope of this documentation, however we will provide a brief primer. In the United States, women are recommended annual screening mammography beginning at age 40 and biennially after age 55. Screening mammograms are recommended for **asymptomatic** patients to detect occult breast cancer. Approximately 90% of screeninng mammograms are normal, and assigned a BIRADS 1 (negative) or BIRADS 2 (benign) after which the woman will return to screening in 1-2 years. However, approximately 10% of screening exams demonstrate an abnormality that requires further imaging and are assigned BIRADS 0 (additional evaluation). BIRADS 3 is a special case where there is low suspicion and the woman will return for screening again in 6 months.

If a woman is assigned BIRADS 0, she will proceed to diagnostic mammogram with special views including compression and magnification paddles. She may also undergo concurrent ultrasound. In 2/3 of diagnostic mammograms, the finding disappears or is resolved as benign and the woman will be assigned BIRADS 1 or 2. In 1/3 of diagnostic mammograms, the finding persists and the woman is assigned a BIRADS score of 4 (suspicious) or 5 (highly suspicious) and will proceed to biopsy. Approximately 2/3 of biopsy results are benign and the woman will return to annual screening. Approximately 1/3 of biopsy results are malignant in which case she will go for further imaging (MRI to evaluate extent of disease) and/or surgical resection (lumpectomy or mastectomy). The patient will receive diagnostic mammograms for several years after this before returning to annual screening.

Lastly, **symptomatic** patients do not receive screening mammography. Any patient with pain, discharge, or other symptoms would go straight to diagnostic mammography where they will be evaluated, assigned a

Releases

No releases published

Packages

No packages published

Languages

Jupyter Notebook 100.0%

https://registry.opendata.aws/emory-breast-imaging-dataset-embed/

The BCS-DBT Dataset

- Normal:
  5129 (91.4%) studies
  4609 patients
- Actionable:
  280 (5.0%) studies
  278 patients
- Biopsy-proven Benign:
  112 (2.0%) studies
  112 patients
- Biopsy-proven Malignant:
  89 (1.6%) studies
  89 patients

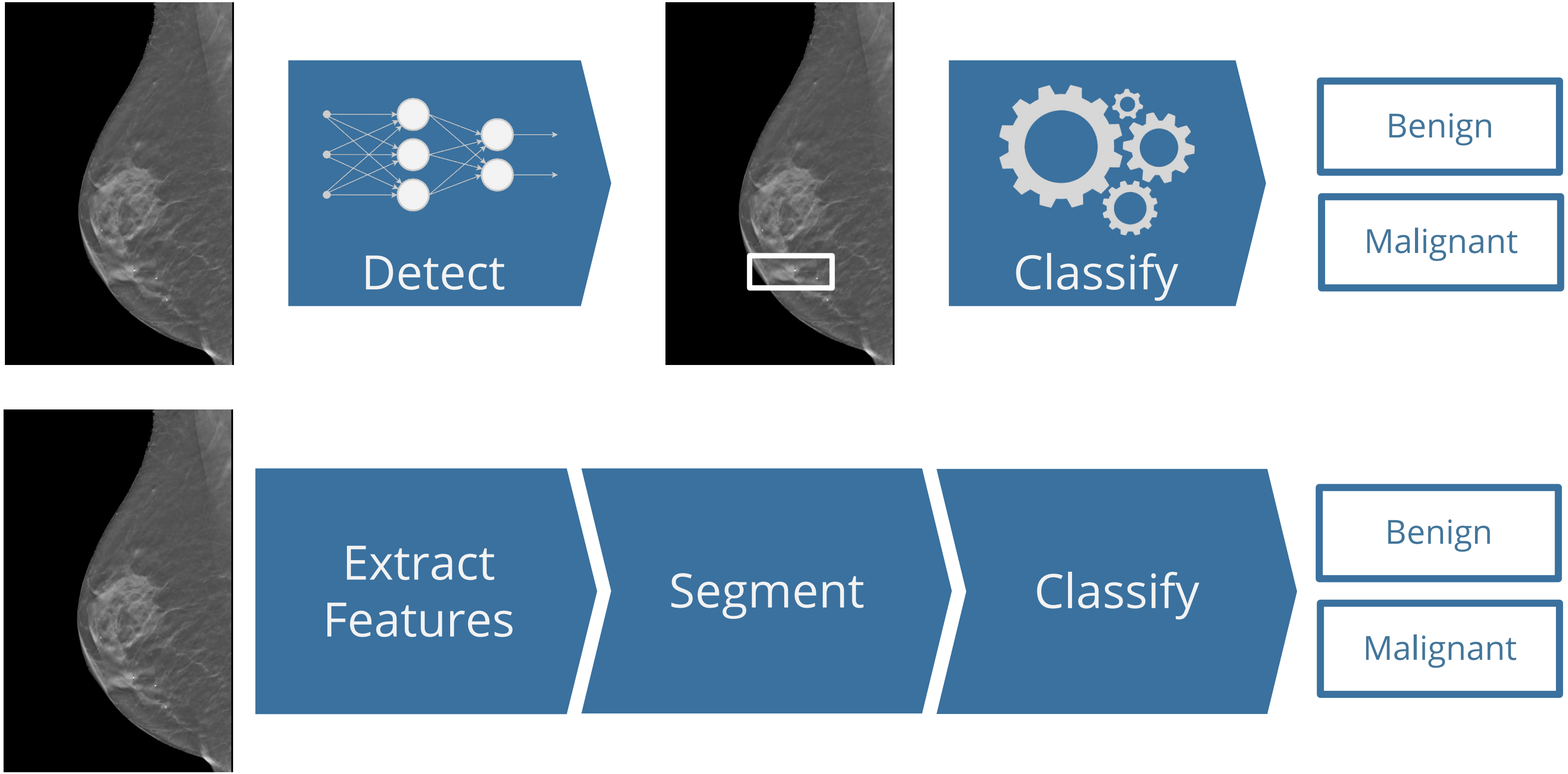
Normal vs non-Normal

```
### RMLO ###
Number of rmlo = 4781
Number of Normal = 4558
Number of non-Normal = 223
Percentage of Normal = 0.953

### LMLO ###
Number of lmlo = 4788
Number of Normal = 4558
Number of non Normal = 230
Percentage of Normal = 0.952

### RCC ###
Number of rcc = 4779
Number of Normal = 4558
Number of non Normal = 221
Percentage of Normal = 0.954

### LCC ###
Number of lcc = 4791
Number of Normal = 4558
Number of non Normal = 233
Percentage of Normal = 0.951
```

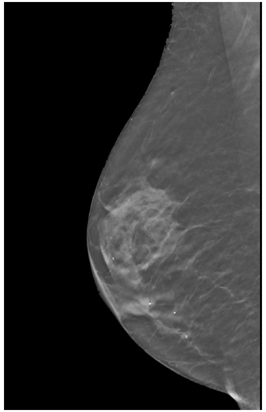
Benign and Malignant

```
### RMLO ###
Number of rmlo = 4781
Number of Benign = 24
Number of Malignant = 20
Number of Benign + Malignant = 44
Percentage of Benign = 0.005
Percentage of Malignant = 0.004

### LMLO ###
Number of lmlo = 4788
Number of Benign = 36
Number of Malignant = 15
Number of Benign + Malignant = 51
Percentage of Benign = 0.008
Percentage of Malignant = 0.003

### RCC ###
Number of rcc = 4779
Number of Benign = 23
Number of Malignant = 19
Number of Benign + Malignant = 42
Percentage of Benign = 0.005
Percentage of Malignant = 0.004

### LCC ###
Number of lcc = 4791
Number of Benign = 36
Number of Malignant = 18
Number of Benign + Malignant = 54
Percentage of Benign = 0.008
Percentage of Malignant = 0.004
```

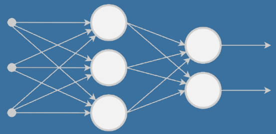
Our approach

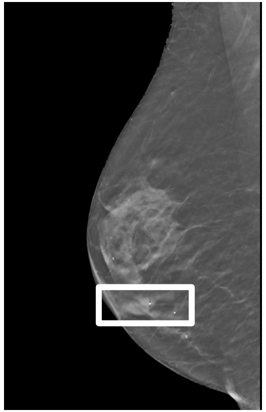


Our approach



Detect

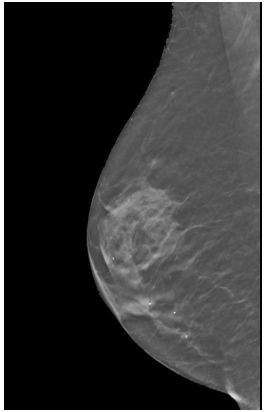




Classify

Benign

Malignant



MatRadiomics

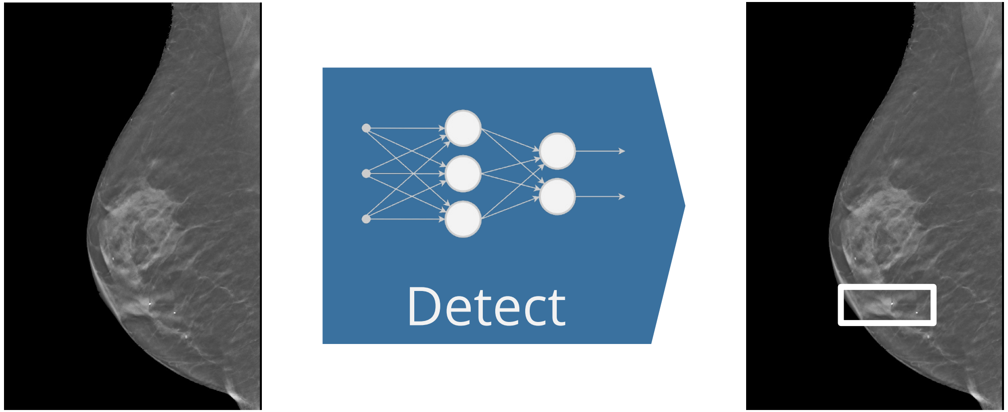
with Alessandro Stefano & Giovanni Pasini

Benign

Malignant

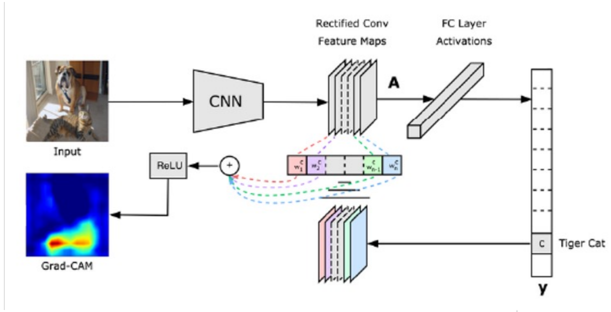
Detection

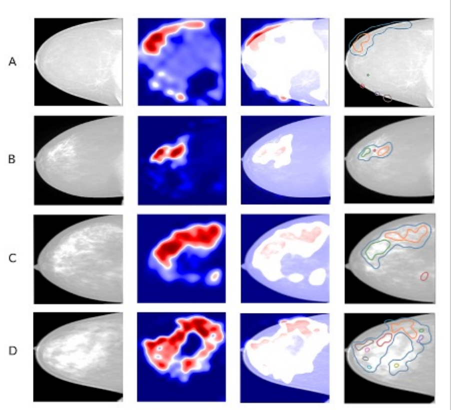
- Large number of ”normal” patients (no mass)
- GAN & UNet
- Anomaly detection approach



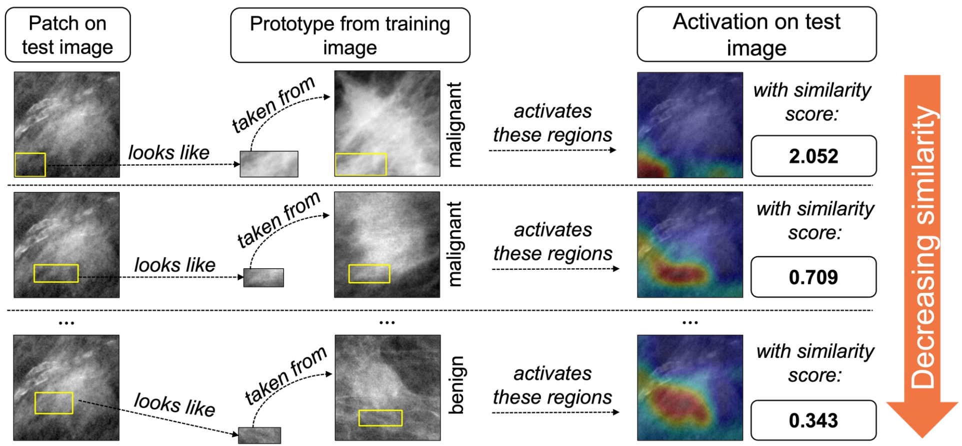
Explainable Classification

CNN + GradCam

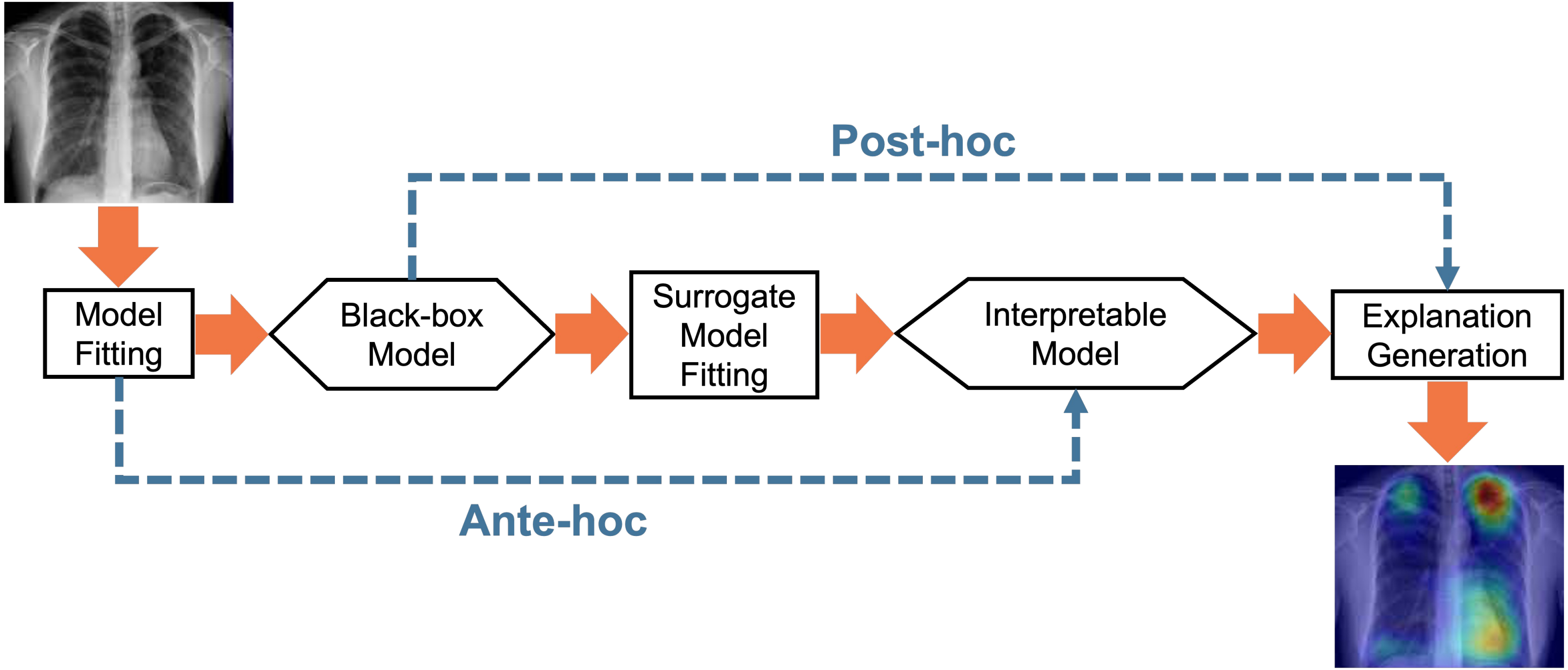




ProtoPNet

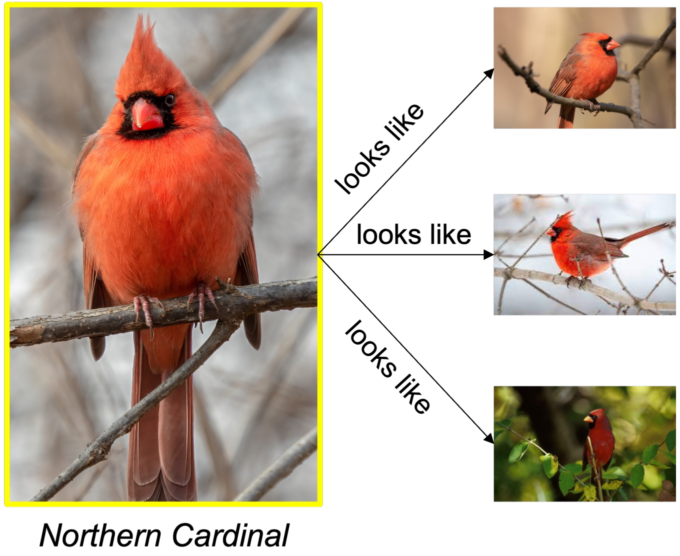


Ante-hoc explainability

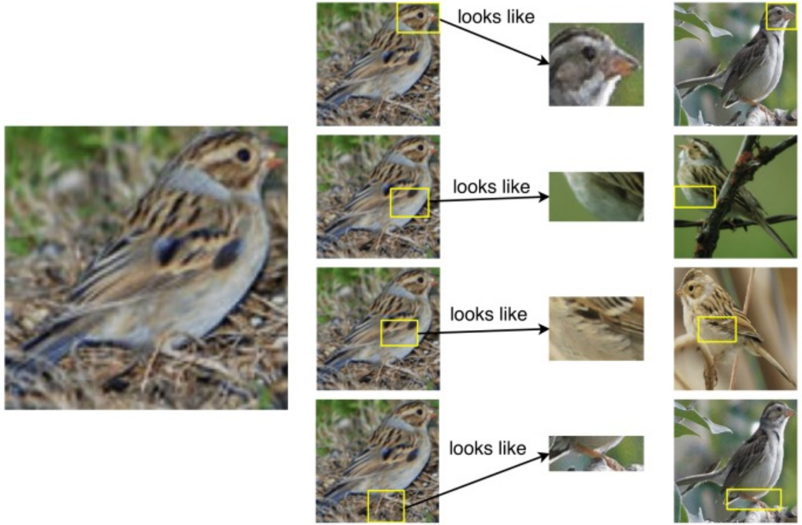


Case-based reasoning

Prototypical Learning



Prototypical Part Learning

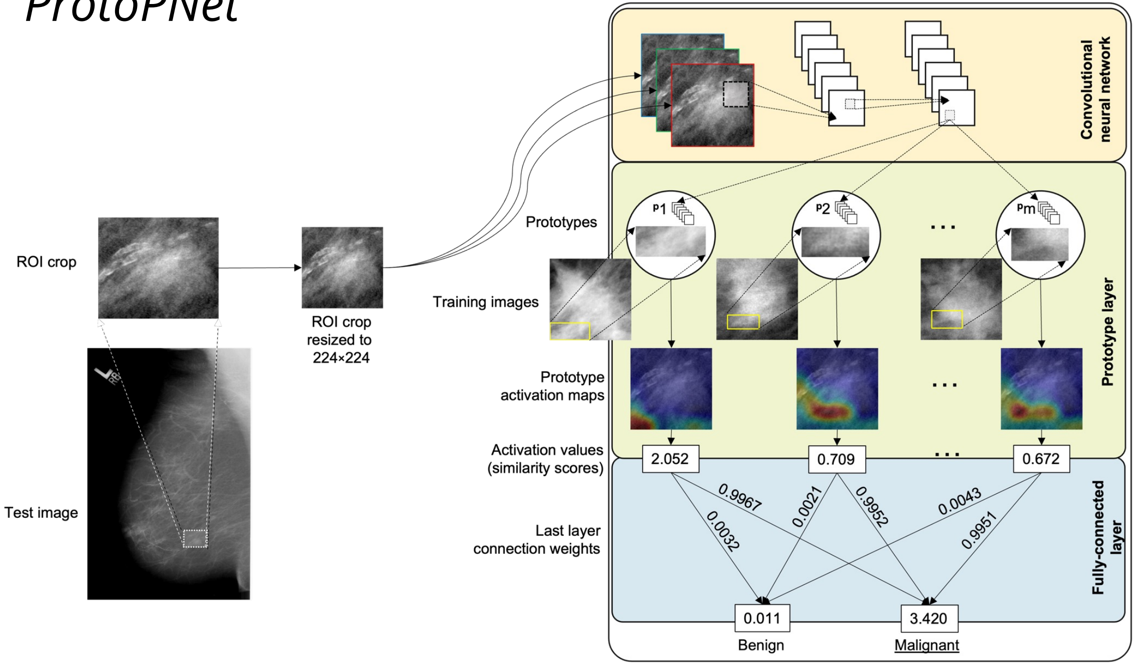


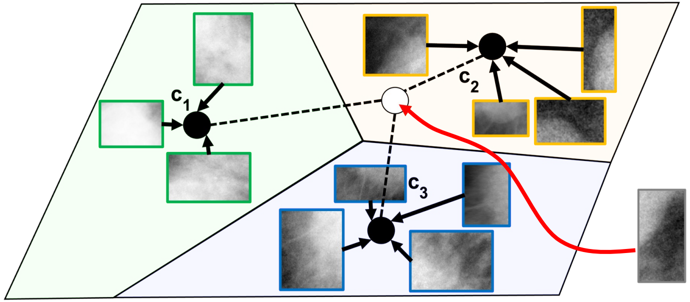
Chen et al. 2019

ProtoPNet on benign vs malignant breast masses

Previously…

On the Applicability of Prototypical Part Learning in Medical Images: Breast Masses Classification Using ProtoPNet



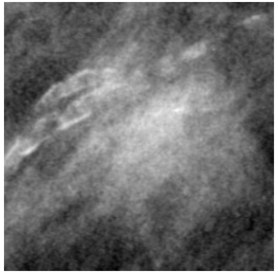


$$Loss := CrossEntr + \alpha\ Clst + \beta\ Sep$$

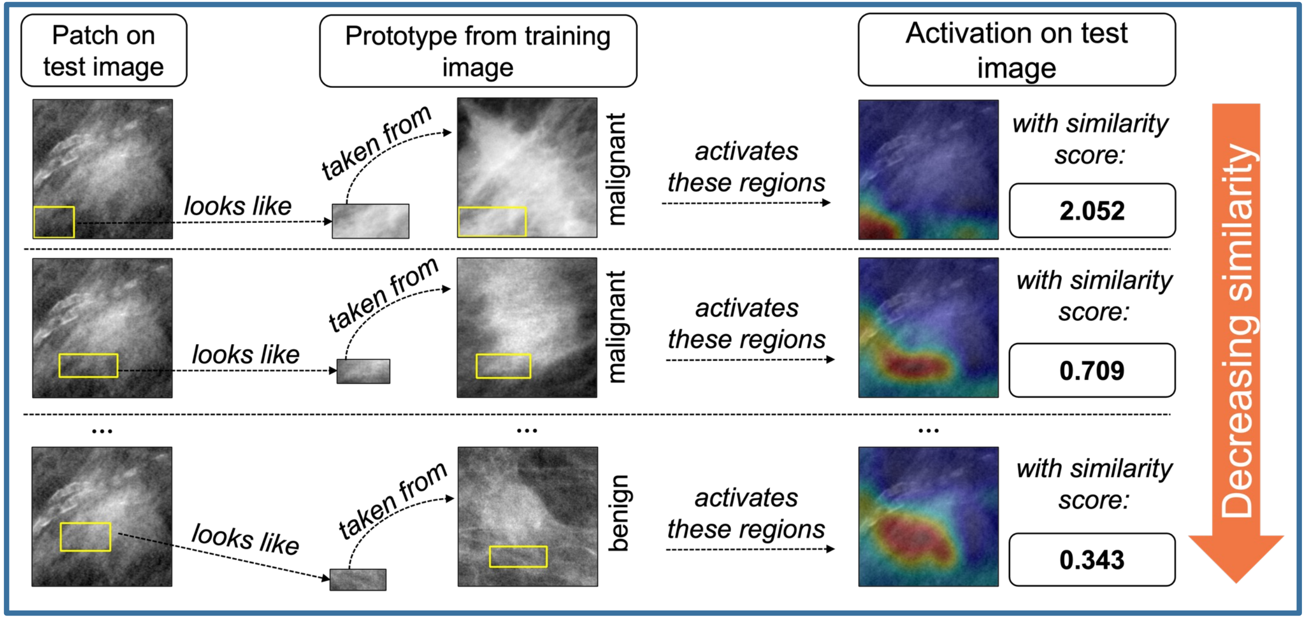
AIHA Workshop
International Conference on Pattern Recognition (Canada)

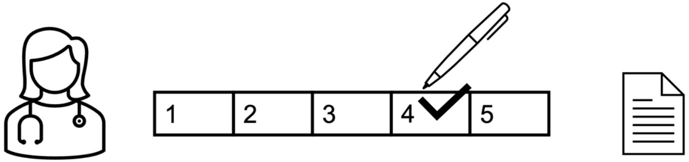
with Gianluca Carloni and Sara Colantonio (ISTI-CNR)

The output of ProtoPNet



Test image: **malignant**
Predicted as: **malignant**





Artificial Intelligence in Medicine

An explainable unified approach for mass detection and classification in DBT

Andrea Berti

INFN

Istituto di Scienza e Tecnologie dell'Informazione "A. Faedo"
Consiglio Nazionale delle Ricerche

I paper su DBT

2D Convolutional Neural Networks for 3D Digital Breast Tomosynthesis Classification *(Zhang et al.)*

- *Classification maligni vs negativi*
- *2D-CNN preallenata su ImageNet per estrarre le feature sulle singole slice*
- *Pooling sulle feature-map delle slice per classificazione 3D*
- *Dataset privato:*
  - *3018 negativi (non maligni)*
  - *272 maligni*
- *AUC = 0.854 (meglio delle 3D-CNN)*

I paper su DBT

A deep learning classifier for digital breast tomosynthesis
(Ricciardi, Mettivier et al.)

- *Classification massa vs non-massa*
- *Custom D-CNN vs AlexNet/VGG*
- *Sembra che non sia patient-stratified*
- *Grad-CAM*
- *Dataset privato 109 pazienti:*
  - *3166 (H1) + 152 (H2) massa*
  - *1526 (H1) + 90 (H2) non-massa*
- $Acc = 0.94$

I paper su BCS-DBT

Intelligent Computer-Aided Model for Efficient Diagnosis of Digital Breast Tomosynthesis 3D Imaging Using Deep Learning *(Adel El-Shazli et al.)*

- *Classification normali vs maligni vs benigni*
- *Mod_AlexNet vs modelli standard*
- *Transfer learning*
- *Dataset:*
  - *499 Normal, 62 Benign, 39 Malignant*
- *Acc = 91.61% sul test*

I paper su BCS-DBT

Applying Graph Convolution Neural Network in Digital Breast Tomosynthesis for Cancer Classification *(Bai et al.)*

- *Classification normal vs cancer*
- *Graph CNN con self-attention pooling layer (risolvere 2D e 3D)*
- *Dataset BCS + privato:*
  - *158 (BCS) + 75 (priv) normal*
  - *75 (BCS) + 94 (priv) cancer (da dove i 75?)*
- *Acc = 0.84, AUC = 0.87*

I paper su BCS-DBT

Trainable Summarization to Improve Breast Tomosynthesis Classification *(Tardy et al.)*

- *Classification normal vs cancer*
- *Multiple instance learning, slabbing + classification (ResNet)*
- *Dataset privato mammo -> pretraining, BCS –> fine-tuning:*
    - *Priv: 1250 benign, 1250 malignant*
    - *BCS: 100 normal, 75 cancer*
- *AUC = 0.73*

Altri paper su BCS-DBT: object detection

- **Detection of masses...** *(Buda et al.): Focal loss per sbilanciamento*
- **Developing breast leasion detection...** *(Hossain et al.): 3 slice consecutive per RGB, (falsi positivi)*
- **Lightweight transformer...** *(Zhang et al.): VIT per risolvere scarsità casi positivi. ResNet - LightweightVIT - ResNet*

Classificazione

- Possibili classificazioni:
  - Normal vs (Actionable + Benign + Malignant)
  - Normal vs (Benign + Malignant)
  - Benign vs Malignant -> ma sono pochi casi
- Come utilizzare le immagini:
  - 2D vs 3D