

Project Techniques of AI

2021 - 2022

2 parts

- Introductory assignment on Canvas with general questions and the selection of the option.
- Project (with several options).

3 options in brief

- Implement a machine learning algorithm and prove to us that it works correctly.
- Select two machine learning algorithms and analyze their performance on an interesting problem. Alternatively, select one algorithm and analyze its behaviour on two interesting problems.
- Analyze a given problem/dataset.

Option 1: Implement yourself

- You are free to select the algorithm, as long as it is an AI algorithm.
- Should not be too obvious and not too hard ...
- In any language you want (more or less).
- How are you going to convince us it works?
- You should inform us about the algorithm, the programming language and the approach for validation before you start.

Option 2: comparison / analysis

- Two algorithms on a dataset.
- One algorithm on two datasets.
- Focus is the interpretation of the differences / similarities.
- Be careful: if you pick two random algorithms and/or two random datasets, you will always find differences. Are they that interesting to investigate?
- You should us inform us on the algorithms, the datasets and what you are going to investigate / what you expect to find.

Option 3: Work on a dataset we provide

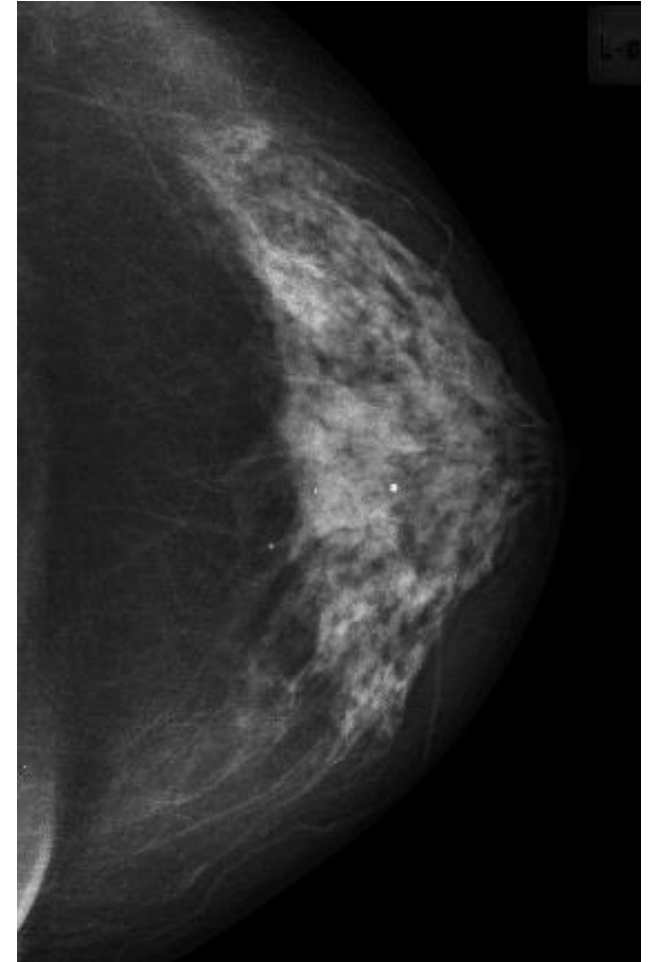
- Detecting breast cancer from images.
- Binary classification problem on 3500 instances, 150 attributes.
- But there are some clear challenges ...

Some background on breast cancer CAD

1mm

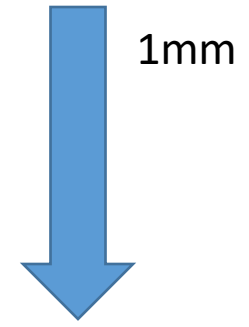


- State of the art is mammography
 - 2 X-RAY images
 - Due to anatomy of the breast, tumours are hard to see
 - Unless you are lucky or they are quite big
- Micro-calcifications:
 - Are tiny white calcium deposits in the breast
 - Are very easy to see on mammography
 - Micro-calcifications appear as a natural process of ageing



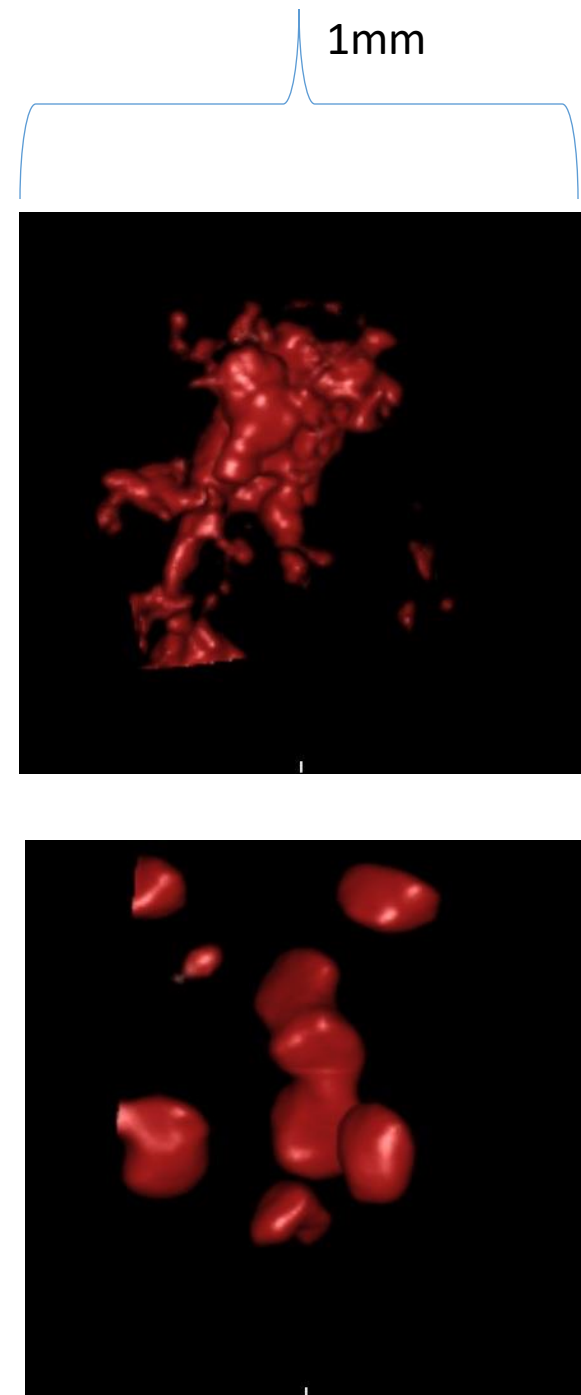
Some more background

- The presence of certain groups of micros is indicative for breast cancer.
- So, there can be a tumour without (visible) micros and the other way around, but there is a correlation.
- Radiologists look for clusters of micros on the mammography



Research hypothesis

- There is even a link between individual micros and cancer: “Shape and texture properties of individual micros allow to predict malignancy”.
- “malignant micro”: micro-calcification present in the neighbourhood of a tumour.
- “benign micro”: micro-calcification not in the neighbourhood of a tumour.
- Problem is having 3D high resolution images of micros



Your challenge

- Based on ~50 properties computed on the ~3500 micros, classify them as benign or malignant.
- **Problem: Every subject in the dataset presents multiple micros and we only know for sure whether the subject has breast cancer or not.**
- **So, 50 benign cases and 50 malignant cases result in 3700 micros in total.**
- Task 1: how well can you classify individual micros assuming all micros per subject have the same label?
- Task 2: how well can you classify whether a subject has cancer based on your classification of the multiple micros per subject?
- We want you to work on both task 1 AND task 2
- You should inform us on the approach you plan to take for both tasks before you start.

Remarks

- Think carefully about the option to select.
- The three options look very different, but require the same amount of effort.
- Each of three options require critical thinking, analysis, understanding and insight.
- Alone or in teams of two. Think about how to make $1+1=3$.

Important

- Project is 50% of total score !!!
- Selection of topic + introductory assignment: March 25
- Submission of project report on Canvas: **May 31**
- Oral discussion: during exam session, June 17 or 18
- Contact us in case of doubts through dedicated Discussion session in Canvas.