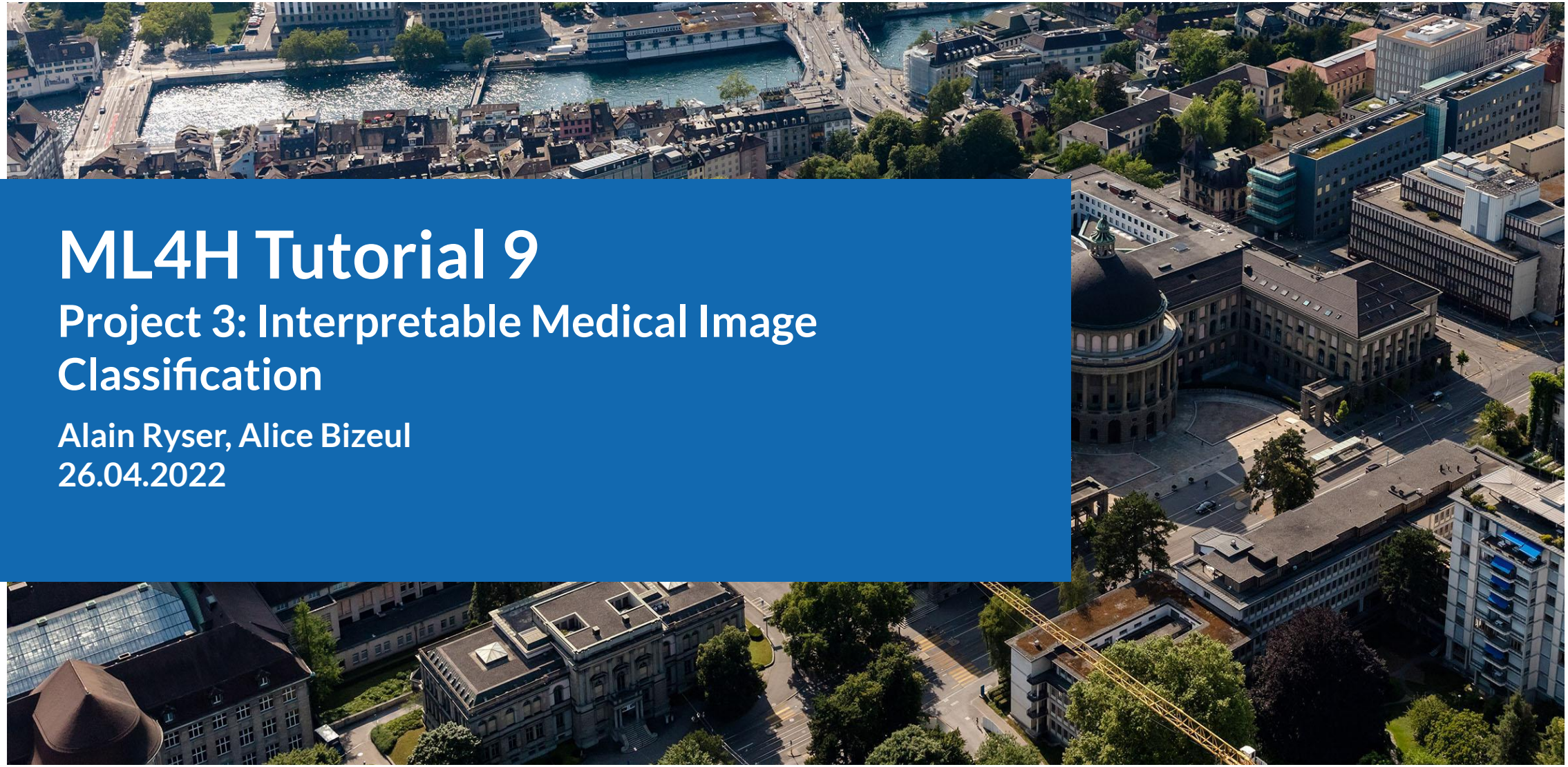
ETH zürich



ML4H Tutorial 9

Project 3: Interpretable Medical Image Classification

Alain Ryser, Alice Bizeul

26.04.2022

Medical Imaging Techniques

Medical Imaging Techniques are at the heart of our medical systems:


Magnetic Resonance Imaging (MRI), Computed Tomography (CT), Echographies, PET scans , X-rays, …


1.18M CT exams & 1.06M MRI exams per year in Switzerland in 2019(1)


Prevalence is increasing in Switzerland (+25% of scanners in CH over the past 5 years)(1)


… increased workload for practitioners


Interest from investment funds into Medical Imaging AI Companies (estimated 600M$ in 2020)(2)

(1) https://www.bfs.admin.ch/bfs/en/home/statistics/catalogues-databases/press-releases.assetdetail.16584130.html
(2) https://www.signifyresearch.net/medical-imaging/vc-funding-for-medical-imaging-ai-companies-tops-2-6-billion

… it remains a challenge to incorporate AI in clinical practice

95% of interrogated practitioners do not trust AI-based clinical tools (1)(2)

Increased efforts should be made to:

- improve **generalisation** of models
- improve **fairness** of predictions
- improve **interpretability** of models (2)

(1) https://www.fda.gov/media/143310/download

(2) Varghese, J. (2020). Artificial intelligence in medicine: Chances and challenges for wide clinical adoption. *Visceral medicine*, *36*(6), 443-449.

Motivation

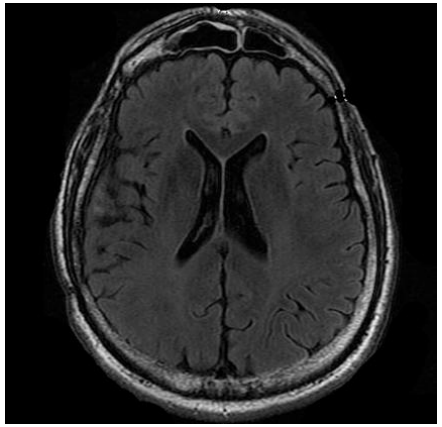
- Brain cancer is one of the deadliest cancer type with ~20% 5-year survival rate in CH (1)
- **Magnetic Resonance Imaging** is one of the first clinical exam when there is brain tumor suspicion
- Can AI provide a **reliable & interpretable** automated tool to identify and classify tumors ?

As a first step: let's focus on reliable & interpretable **tumor identification** using **MRI** scans

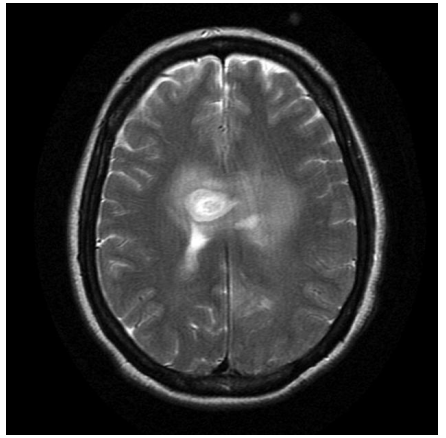
(1) https://www.liguecancer.ch/a-propos-du-cancer/les-chiffres-du-cancer/-dl-/fileadmin/downloads/sheets/chiffres-le-cancer-en-suisse.pdf

Dataset: Kaggle Brain Tumor Detection dataset

- 278 brain slices **with or without** any brain tumor (111 w/o tumor; 167 with tumor)
- tumors are **visible** to non-expert eyes
- Train-Validation-Test (80-10-10%) provided



No tumor



Tumor

Magnetic Resonance Imaging

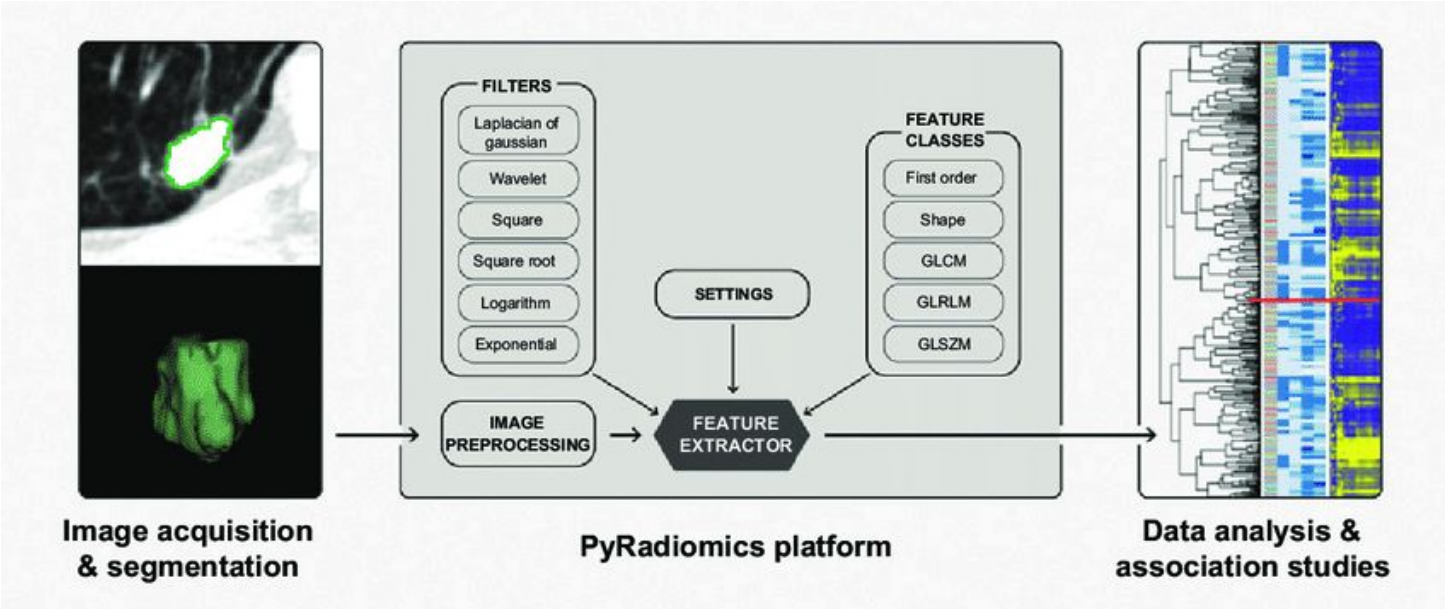
- **non-invasive** imaging technique for **soft tissues** imaging (brain, abdomen, …)
- **safe** imaging method (no X-rays, ionisation radiation)
- based on **Nuclear Magnetic Resonance** (NMR) of water molecules in our body
- the discovery of NMR and MRI was awarded a Nobel Prize in Medicine and Physiology in 2003

Project Tasks

1. Task 1: Implement Random Forest baseline using provided dataset of PyRadiomics features

Pyradiomics Features https://pyradiomics.readthedocs.io/en/latest/

- open-source Python library for **automated radiomics features extraction**
- several families of features to choose from
  - **first-order** statistics
  - **texture** features
  - **shape** features
- automated pre-processing, extraction & storage on **original** and **filtered** images



Pyradiomics Features

Shape Features https://pyradiomics.readthedocs.io/en/latest/features.html

- extracts shape feature on masked regions (e.g brain, tumor, ROI)
- volume, surface, perimeter, sphericity, elongation, …
- uses original image to extract features (2D or 3D)

Texture Features https://pyradiomics.readthedocs.io/en/latest/features.html

- extracts gray-level based/homogeneity features on original and filtered images
- gray-level **co-occurrence** matrix reports the probability of two gray levels to co-occur in a neighborhood
- gray-level **size zone** matrix reports the area where gray-levels co-occur in a neighborhood
- gray-level **run length** matrix reports the amount of patches with consecutive gray-level
- **gray-tone difference** matrix reports the difference between gray-level and neighboring average level
- gray-level **dependence** matrix reports areas where pixels are connected based on gray level

Image Filters https://pyradiomics.readthedocs.io/en/latest/index.html?highlight=filter#filter-classes

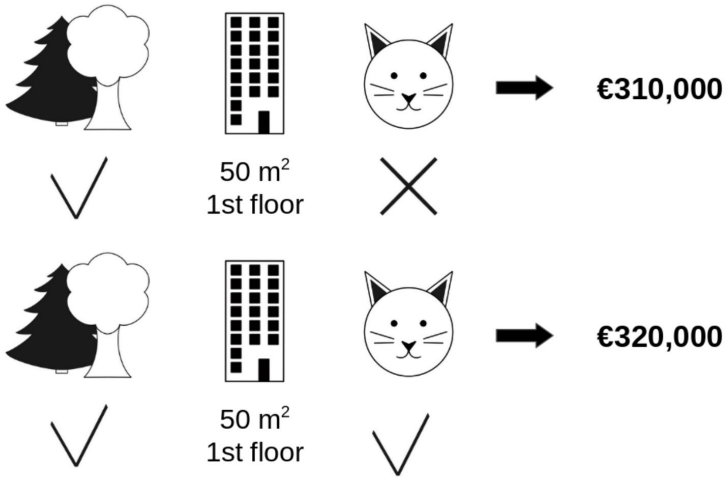
- provides a range of image filters: Laplacian, Wavelets, Square, Square Root, Logarithm, Exponential, …
- filtering is automatically embedded in the pre-processing

Project Tasks

1. **Task 1**: Implement Random Forest baseline using provided dataset of PyRadiomics features
2. **Task 2**: Implement CNN baseline and provide SHAP values interpretation of results using the provided MRI Dataset

Shapley Values https://papers.nips.cc/paper/2017/file/8a20a8621978632d76c43dfd28b67767-Paper.pdf

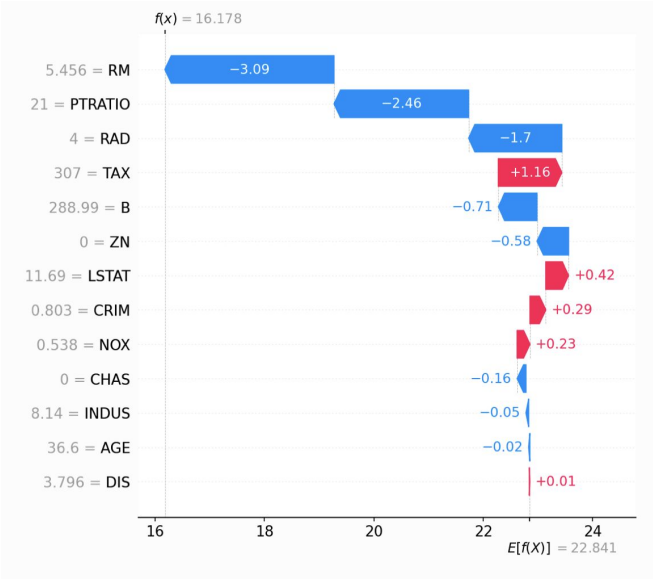
- Feature contribution metric originating from **cooperative game theory**
- Adapted and used for **post-hoc** ML interpretability purposes
- Shapley Values are the **average marginal contribution** of each feature to the difference between the given prediction and the average prediction
- Shapley Values provide information on **relative importance** of each input feature for a given prediction v



$$\Phi_i(v) = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!\,(N - |S| - 1)!}{N!} \left(v(S \cup \{i\}) - v(S) \right)$$

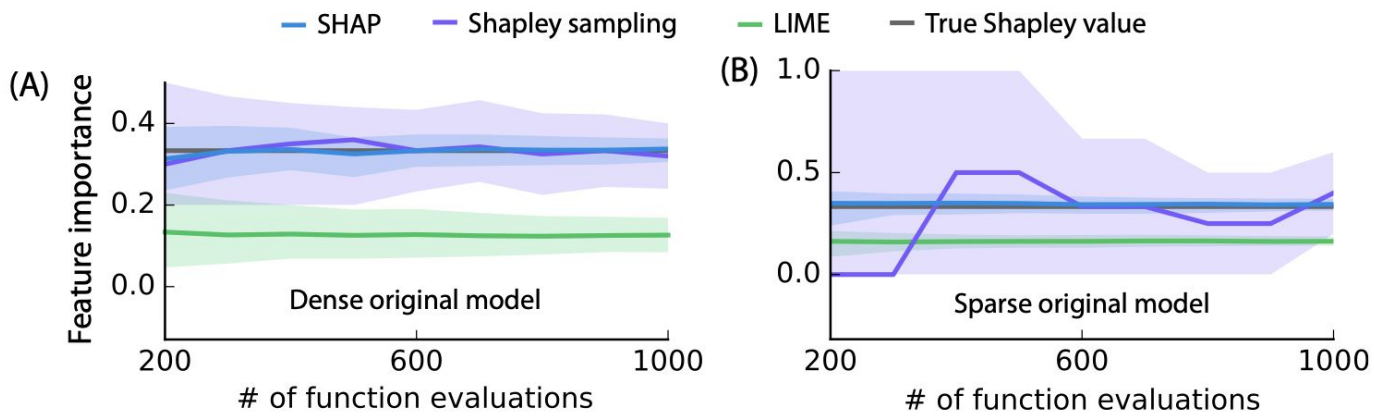
Shapley Values https://papers.nips.cc/paper/2017/file/8a20a8621978632d76c43dfd28b67767-Paper.pdf

- Axiomatic metric:
  - *Efficiency*: sum of values across input features add up to the difference between given and average prediction
- Average prediction is computed on a reference group/background samples, depending on the question we are trying to answer
- Pros:
  - Model agnostic approach, only requires input-output pairs
  - Information about local predictions as well as the global model
- Cons:
  - Amount of input perturbation to consider scales expo. with input size
  - Requires estimates computation in most real-world datasets.



SHapley Additive exPlanations (SHAP)

- Shapley sampling values using **sampling methods** to approximate true Shapley values
- **SHapley Additive exPlanations** (SHAP) combines sampling and other interpretable approaches (i.e. Local Surrogate Models - LIME, DeepLift):
  - Increased **computational efficiency**
  - **Improved approximation** of true Shapley Values
- Various approximation methods :
  - Kernel SHAP: model-agnostic
  - Linear SHAP: independance assumption
  - Deep SHAP: deep networks



SHAP Python library https://shap.readthedocs.io/en/

- provides **automatic estimates** of Shapley values for wide range of ML/DL models types (1)
  - Model Agnostic
  - Linear models
  - Tree based models
  - Neural networks
    - single & multi output
    - including transformers
    - including image classification, image captioning, text generation
- supports Tensorflow, Pytorch, Sklearn models

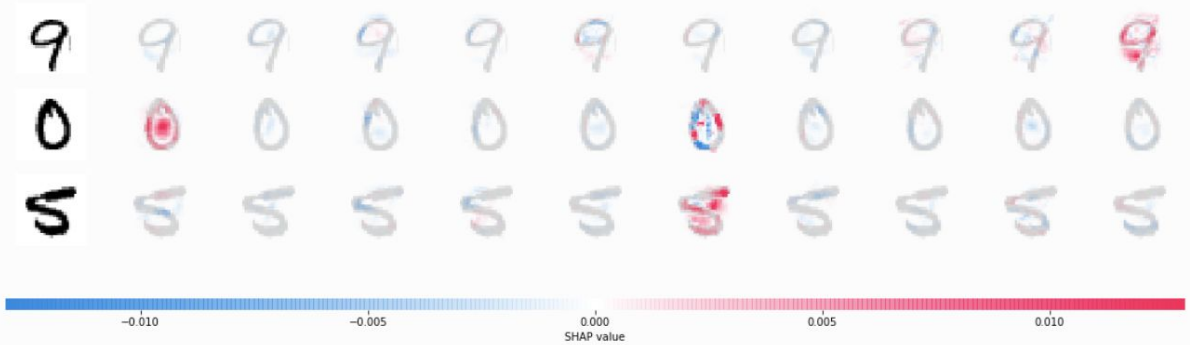
```
# since shuffle=True, this is a random sample of test data
batch = next(iter(test_loader))
images, _ = batch

background = images[:100]
test_images = images[100:103]

e = shap.DeepExplainer(model, background)
shap_values = e.shap_values(test_images)
```

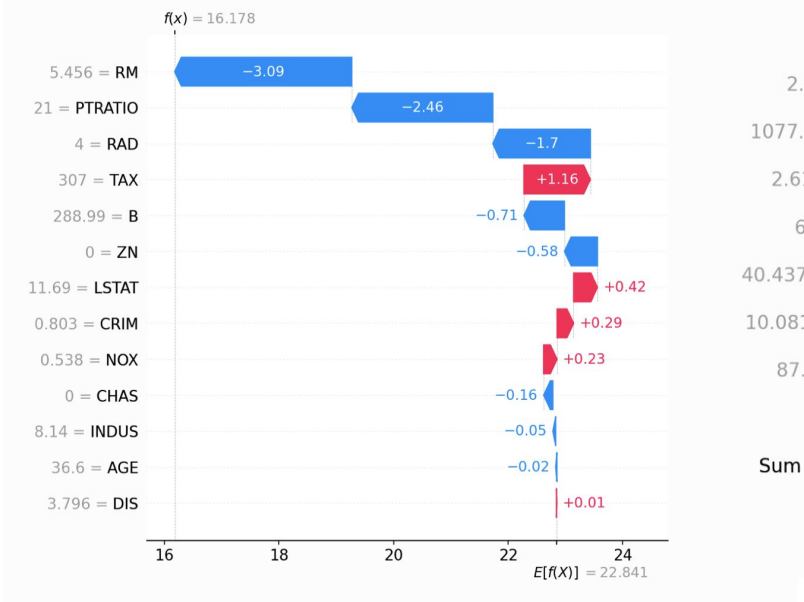
```
shap_numpy = [np.swapaxes(np.swapaxes(s, 1, -1), 1, 2) for s in shap_values]
test_numpy = np.swapaxes(np.swapaxes(test_images.numpy(), 1, -1), 1, 2)
```

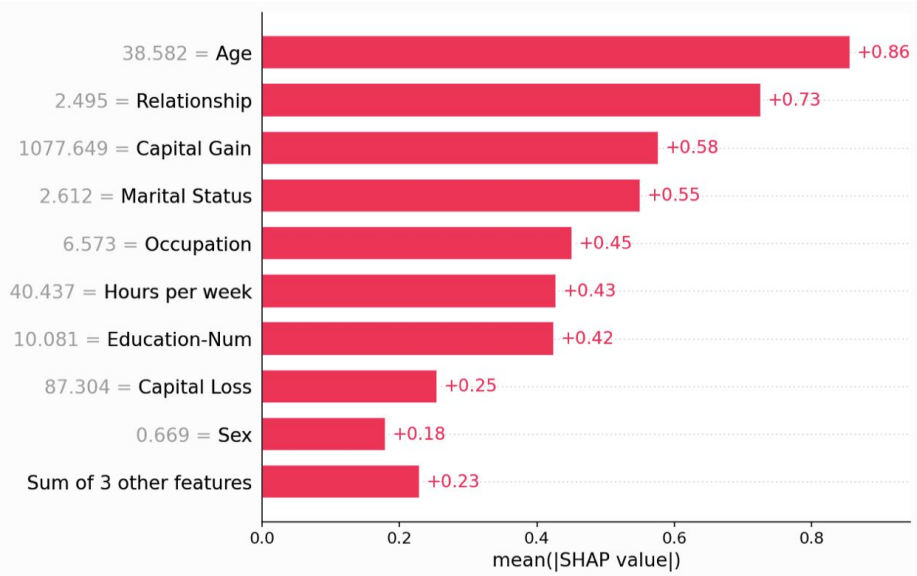
```
# plot the feature attributions
shap.image_plot(shap_numpy, -test_numpy)
```

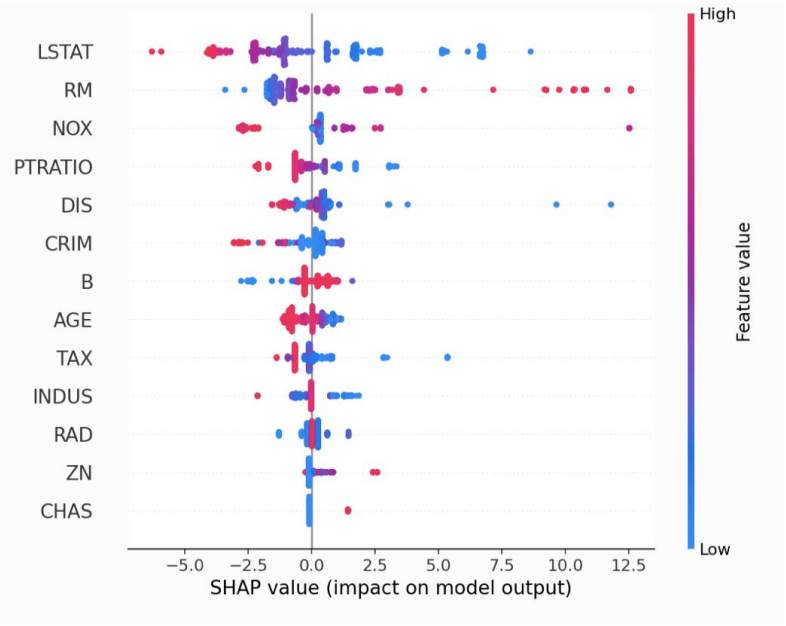


SHAP Python library https://shap.readthedocs.io/en/

- provides **automatic estimates** of Shapley values for wide range of ML/DL models types (1)
- provides **visualisations** of Shapley values for interpretability purposes (2)





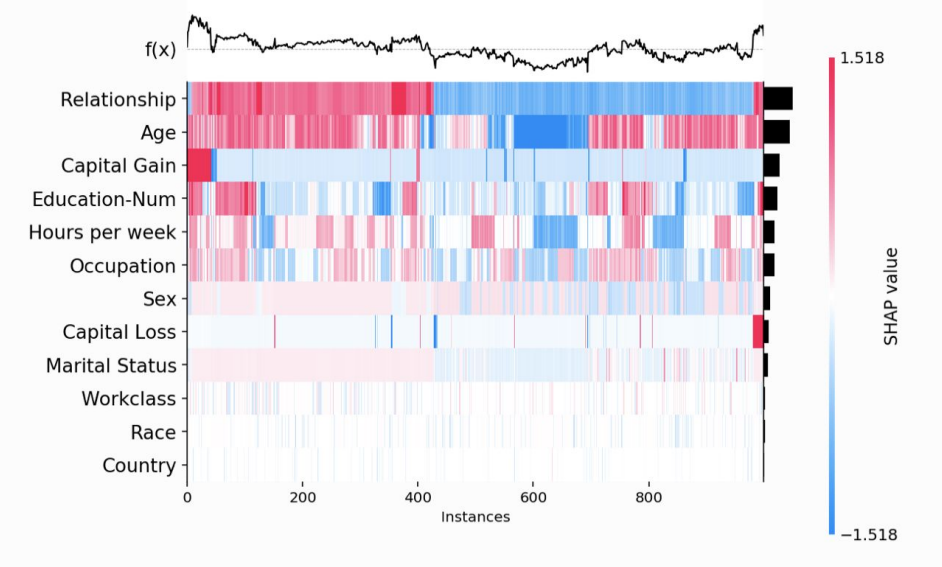


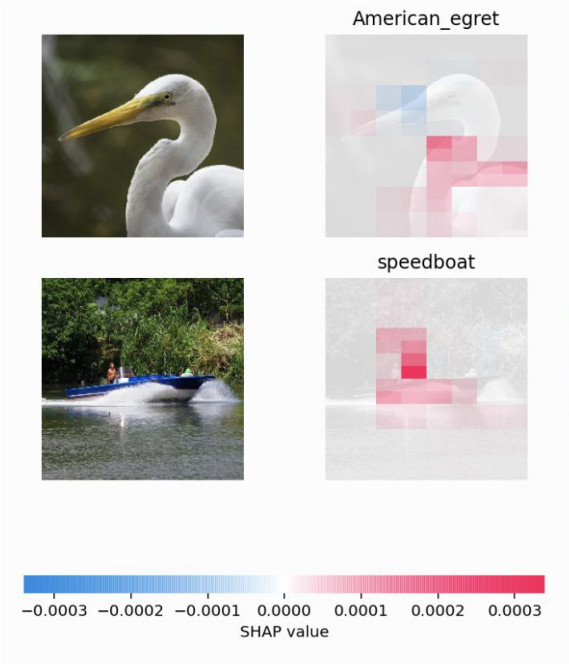
(1) https://shap.readthedocs.io/en/latest/api.html#explainers

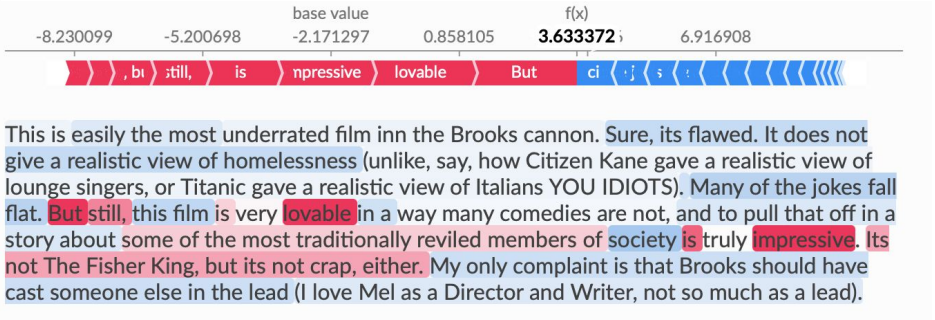
(2) https://shap.readthedocs.io/en/latest/api.html#plots

SHAP Python library https://shap.readthedocs.io/en/

- provides **automatic estimates** of Shapley values for wide range of ML/DL models types (1)
- provides **visualisations** of Shapley values for interpretability purposes (2)







(1) https://shap.readthedocs.io/en/latest/api.html#explainers

(2) https://shap.readthedocs.io/en/latest/api.html#plots

Project Tasks

1. Task 1: Implement Random Forest baseline using provided dataset of PyRadiomics features
2. Task 2: Implement CNN baseline and provide SHAP values interpretation of results using the provided MRI dataset
3. Task 3: Implement at least two additional interpretable classification methods on either dataset and one additional post-hoc interpretation method
4. [OPTIONAL] Improve model accuracy using approaches of your choice (e.g transfer learning, data augmentations, …)
5. Compare results and motivate the choice of your final classifier based on performance-interpretability trade-off

- Code Template and Dataset can be accessed **here**.

Important: Elaborate on the interpretability aspect of the implemented classifier/post-hoc method in the report for each subtask!

Project Deliverables

- Solve all tasks.
- **Report** of max. 4 pages, 11pt (+ 1 page for references + 1 page of appendix if needed).
- **Well-commented code/jupyter notebooks** with conda environment and README.
- Do not hardcode any results! We will run your code.
- Ensure sequential execution and **reproducibility**.
- **Do not copy solutions from previous projects!** We are aware of all existing solutions on github and kaggle. We run code similarity checks and check for plagiarism in the reports from previous years solutions. Any plagiarism will result in a 0 grade for all projects.
- **Deadline**: 17.05.2022

Project Grade

- To grade the project we will focus (on equal parts) on:
  - the content, organisation, clarity, quality and writing of the **final report**.
  - the quality of the **implementation** (reproducibility and clarity).
  - the **performance*** of the methods used to solve the tasks, and the **justification** of the choices.
- The **prerequisites** to get the maximum grade are:
  - write a clear and **good report.**
  - submit a **clean code** with **easy instructions** on how to reproduce each result of the report.
  - solve every task with **well-justified** methods and choices.
  - bonus: implement **creative** models for one/the optional task

*We will consider resource constraints. Aside from correct baseline implementation, the aim is not to get the best performance but to explore and discuss relevant methods.

Reading List

Shapley Values

- https://papers.nips.cc/paper/2017/file/8a20a8621978632d76c43dfd28b67767-Paper.pdf
- https://apps.dtic.mil/sti/pdfs/AD0604084.pdf

GradCAM

- https://arxiv.org/pdf/1610.02391.pdf

Integrated Gradients

- http://proceedings.mlr.press/v70/sundararajan17a/sundararajan17a.pdf

Local Surrogate Models (LIME)

- https://arxiv.org/pdf/1602.04938.pdf?ref=morioh.com

DeepLIFT

- https://arxiv.org/pdf/1704.02685.pdf

ETH zürich

Questions?

Also on Moodle (preferred: your classmates probably have similar questions!)
or by email at alain.ryser@inf.ethz.ch , alice.bizeul@inf.ethz.ch .