xLungs

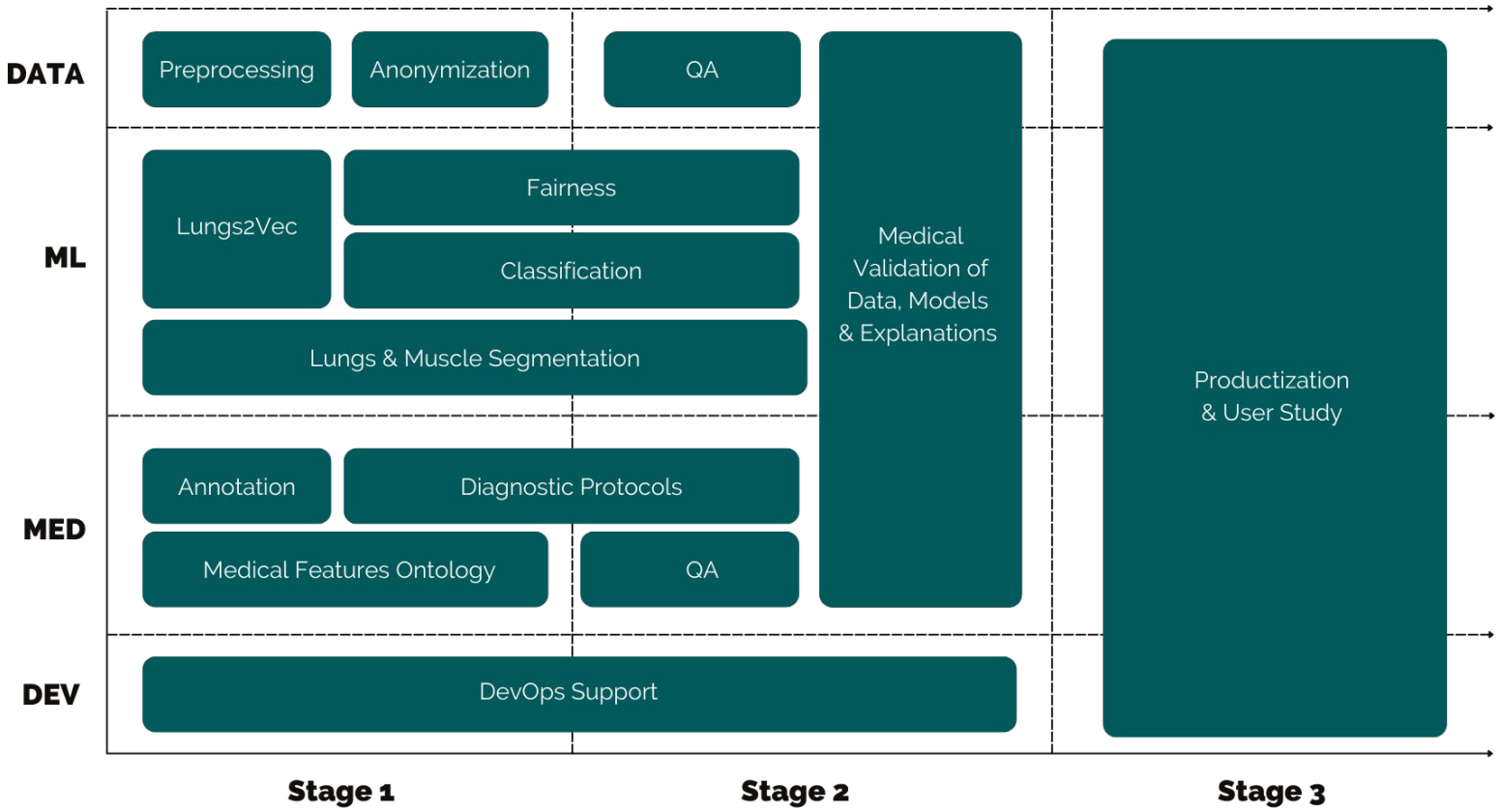
Efficient and Explainable Diagnostics' Support for Radiologists

MI²DataLab Winter Seminar 2022

MI

Let's **Support Radiologists** with **Diagnostics of Lungs Diseases** by Using Skills We Are Experts In: **EDA**, **ML** & **XAI**

MI


DATA
ML
MED
DEV
Preprocessing
Anonymization
QA
Lungs2Vec
Fairness
Classification
Lungs & Muscle Segmentation
Annotation
Diagnostic Protocols
Medical Features Ontology
QA
DevOps Support
Medical Validation of Data, Models & Explanations
Productization & User Study
Stage 1
Stage 2
Stage 3

Task I

Goal: Acquisition of data from different sources to create one anonymized dataset.

1. Data sources
2. Anonymization

Data acquisition

During our project we acquired (and are still acquiring) data from

- PGRP (Polish Group of Lung Cancer) consisting of CT scans and meta data from screening tests ~40 000 patients
- B50 consisting of CT, X-ray and metadata ~50 patients
- B2000 consisting of X-ray and metadata ~2000 patients

Elżbieta Sienkiewicz

Data anonymization

- Medical imaging data in the DICOM format == a lot of precise metadata
- A DICOM dataset is a dictionary consisting of many predefined attributes
- The anonymization procedure consists of several stages
  - On-site (in the hospital) blacklisting
  - Off-site (in the faculty) whitelisting
  - Independent quality verification
- Not only the patient's personal information must be anonymized, but also:
  - Timestamps
  - Globally unique identifiers
  - Attributes with questionable content
  - Data concerning the medical personnel
  - Characteristic information describing the equipment used
- Sensitive data may also be present in the images themselves

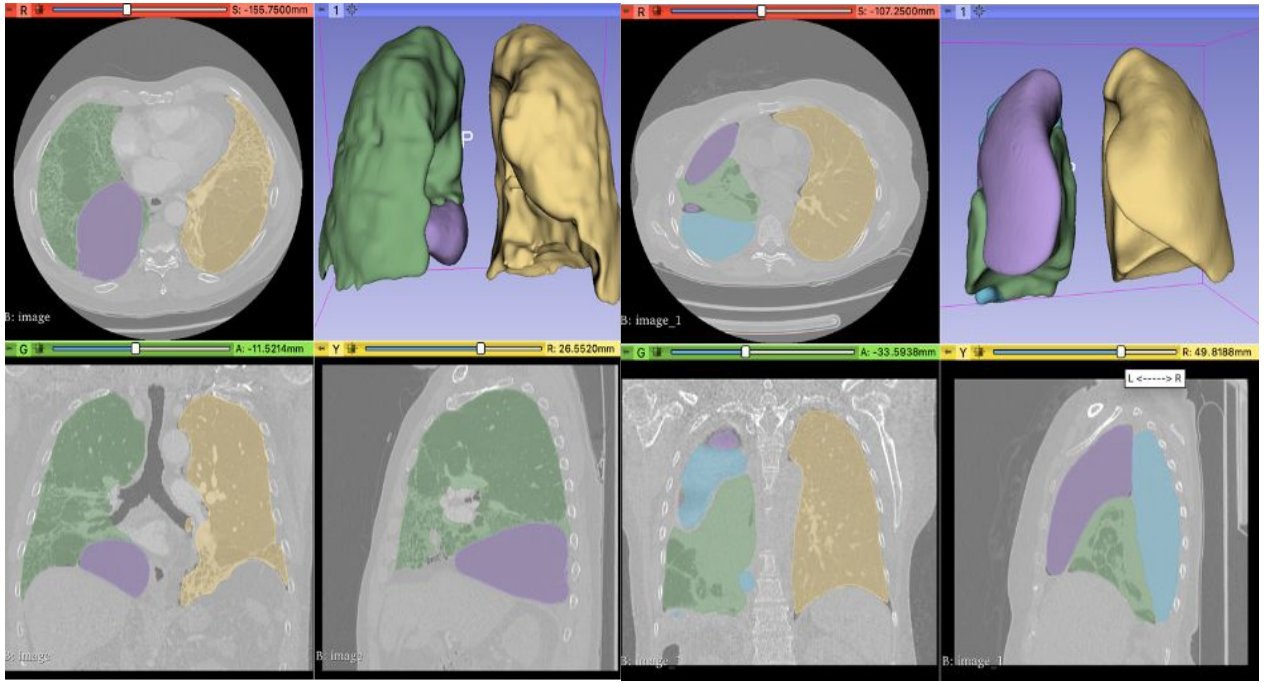
Mateusz Grzyb

Task II

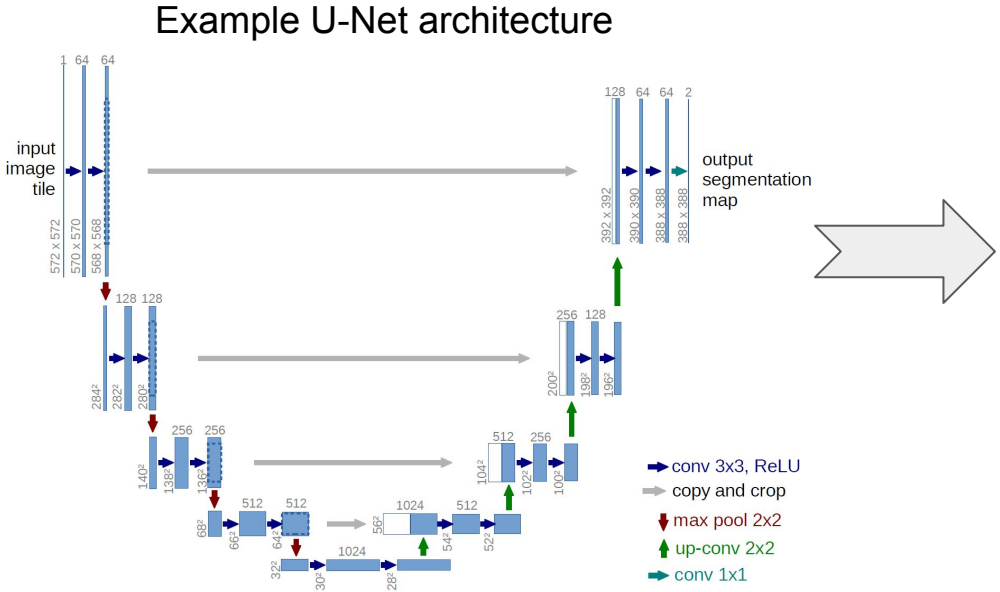
Goal: Modeling using data collected at task I.

- Segmentation
  - lungs
  - muscles
- Classification
  - vector representation
  - classification of lungs tumor
- Fairness

Segmentation



nnU-Net



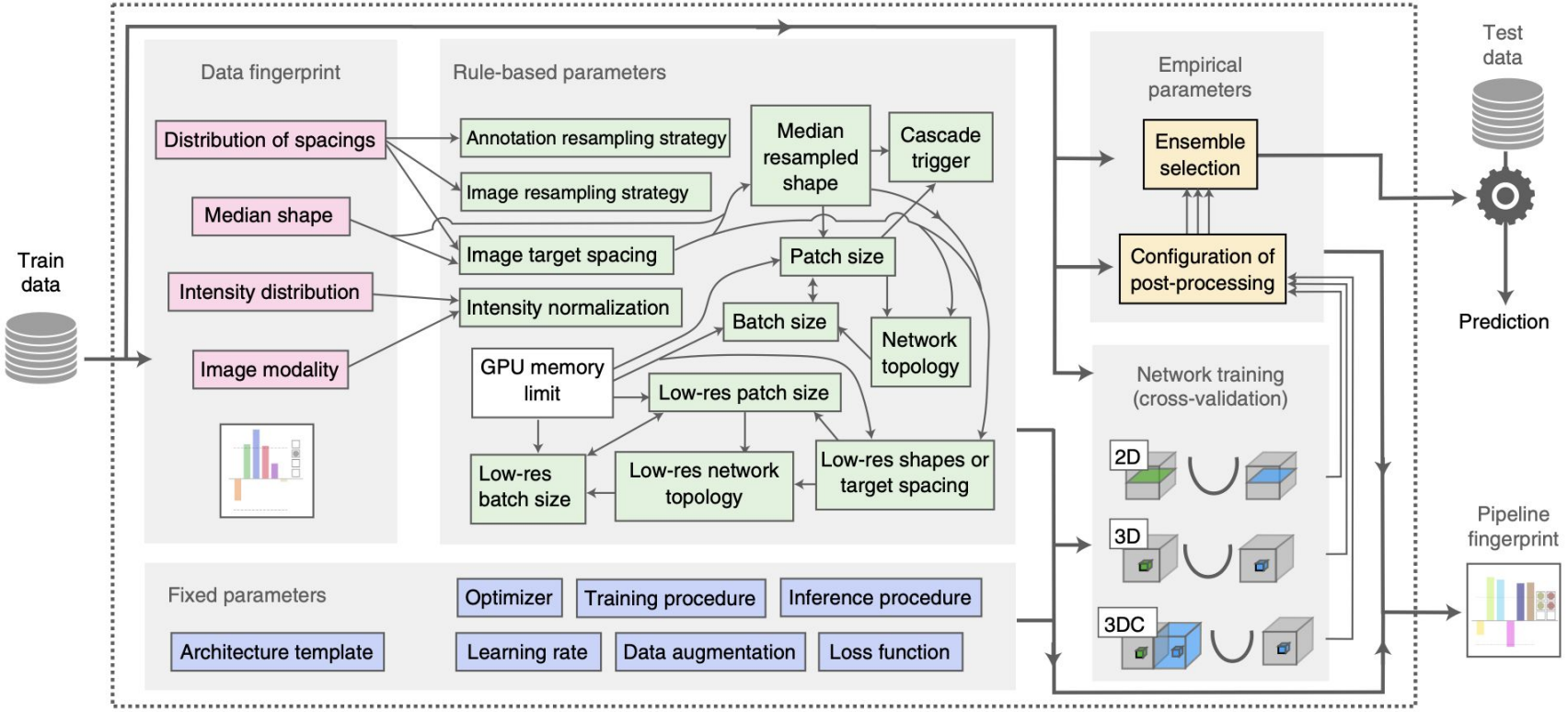
1. Collect design decisions that do not require adaptation between tasks (**'fixed parameters'**).

2. For as many of the remaining decisions as possible, formulate explicit dependencies between specific dataset properties (**'dataset fingerprint'**) and design choices (**'pipeline fingerprint'**) in the form of heuristics for almost-instant adaptation (**'rule-based parameters'**).

3. Learn only the remaining decisions empirically from the data (**'empirical parameters'**).

Isensee, F., Jaeger, P. F., Kohl, S. A., Petersen, J., & Maier-Hein, K. H. (2020). nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. Nature Methods, 1-9.

nnU-Net



Isensee, F., Jaeger, P. F., Kohl, S. A., Petersen, J., & Maier-Hein, K. H. (2020). nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. Nature Methods, 1-9.

Bartłomiej Sobieski

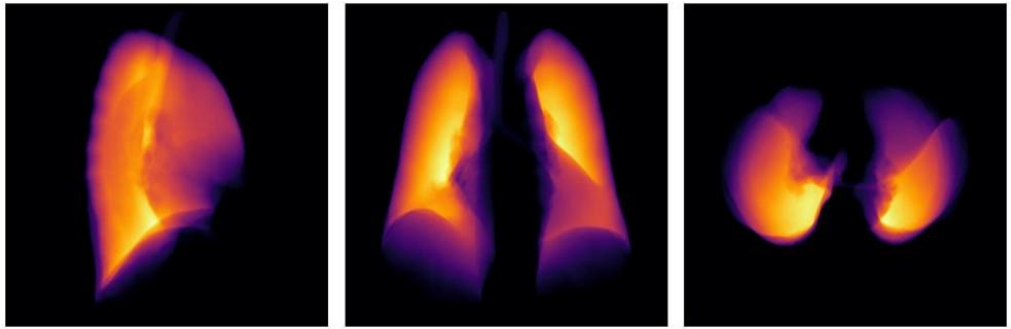
Segmentation of skeletal muscles

- low muscle density and volume is a predictor of many diseases (cancer, sarcopenia, severe COVID-19, etc.)
- segmentation and analysis of various muscle groups may improve work of radiologist
- Segmentation with classical image processing methods

Adam Kozłowski

Classical image processing methods

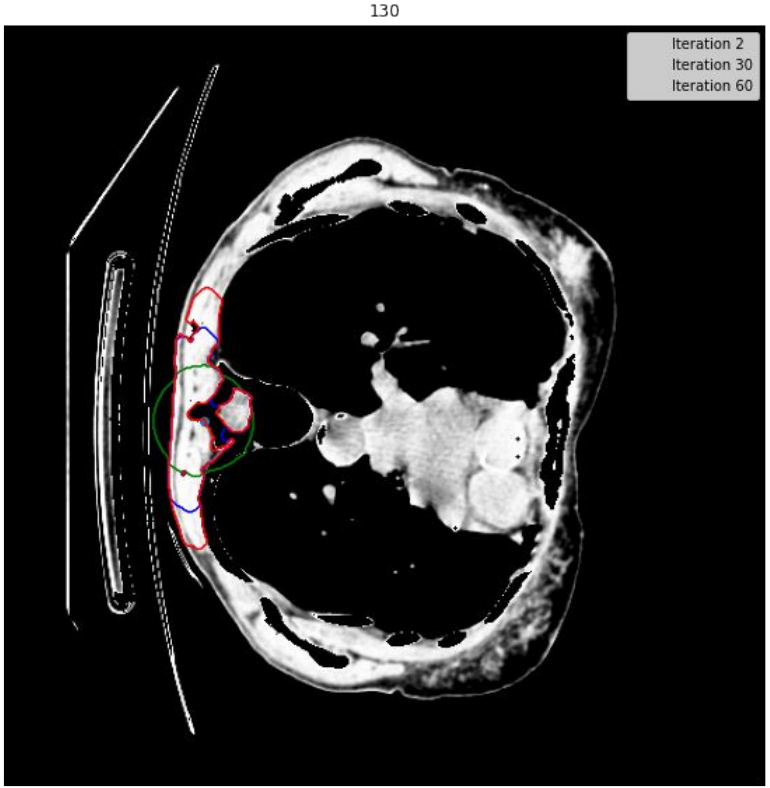
- Pre-DL algorithms in image segmentation
- CT images pixel values are in Hounsfield Units (each value represents tissue radiodensity)
- relatively poor performance on scale
- **does not require labeled data**



Adam Kozłowski

Segmentation of skeletal muscles

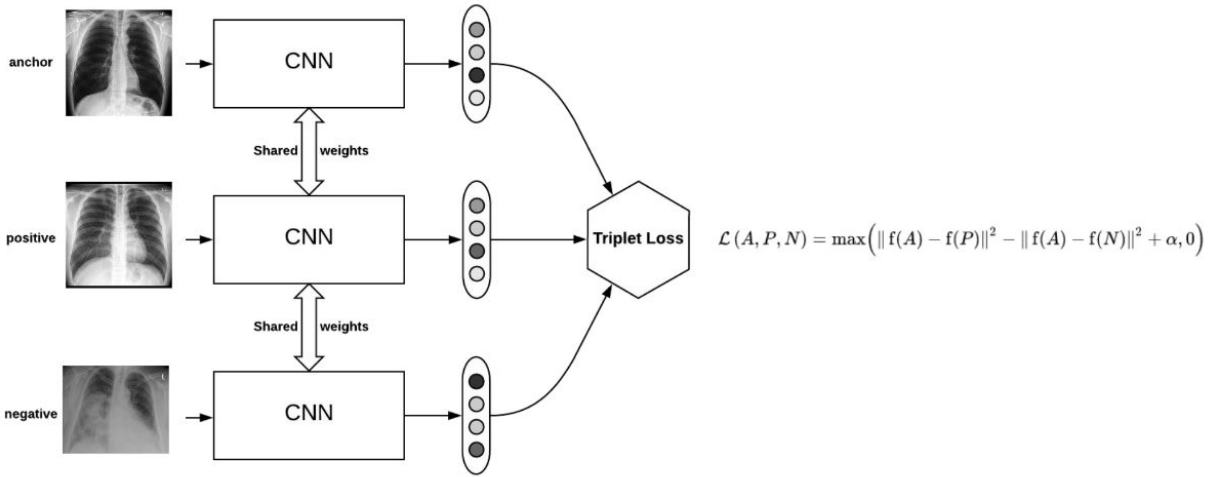
- Used algorithm: active contour (Snake) by Chan and Vese
- Contour wraps around areas based on its energy equation
- Requires complex preprocessing,
- Produces Dice score ~0.7 (to be evaluated in further experiments)



Adam Kozłowski

Classification from X-rays

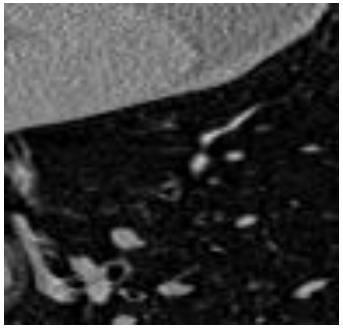
We want to classify lungs diseases using X-rays. To tackle this problem we created siamese network which encodes X-rays. Next we are using those embeddings to train classical ML models for multilabel classification.



Weronika Hryniewska-Guzik

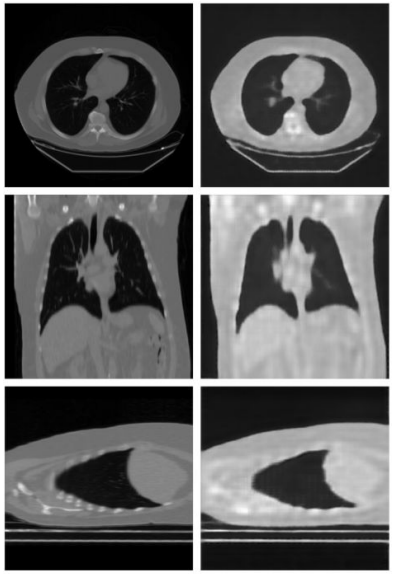
Autoencoders

We wanted to create embeddings of lungs CT scans.
To achieve that we tested classical autoencoders with CNNs and autoencoders using Transformers architecture.



Real CT scan

Reconstructed CT scan



Maciej Chrabąszcz

Fairness

Basing on the PGRP data we are building models to predict different targets basing on the volumetric features of different masks and organs. In the next step we are checking whether the models perform well on different population subgroups such as different ages and sexes.

Ilona Grabowicz

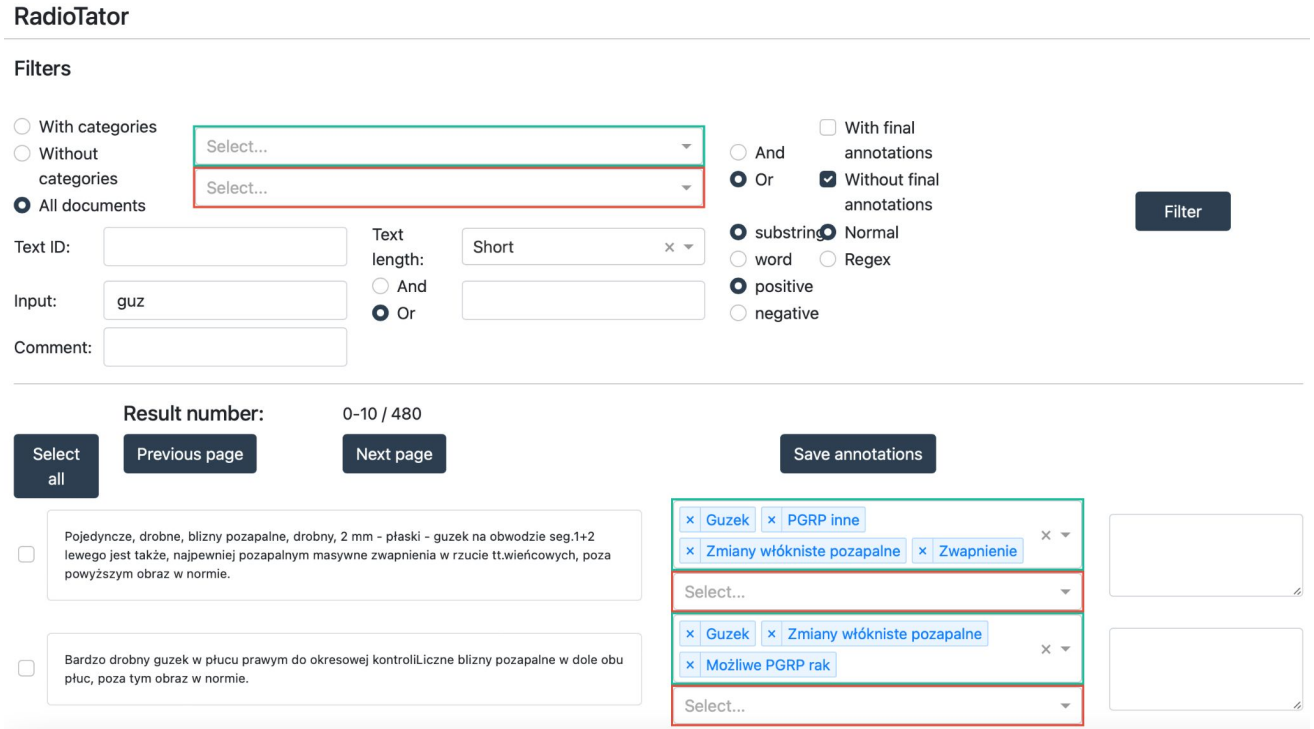
Ontology

We prepared medical ontology of anatomic features and pathological symptoms for both X-rays and CT scans. Our work was supported by RadLex - lexicon of radiological information that was prepared by the Radiological Society of North America. It is an ontological system with a useful vocabulary for radiologists. Our ontology is both in Polish and English.

SUBTAPE ID (RADLEX)	PL TYPE	PL SUBTYPE	ENG TYPE	ENG SUBTYPE	DEFINITION	NOTE
RID28531	Zagęszczenia (miąższowe)	Zagęszczenia typu matowej szyby	Opacity (parenchymal)	Ground-glass opacity	Hazy increased opacity of lung, with preservation of bronchial and vascular margins. It is caused by partial filling of airspaces, interstitial thickening (due to fluid, cells, and/or fibrosis), partial collapse of alveoli, increased capillary blood volume, or a combination of these, the common factor being the partial displacement of air.	
RID43255		Konsolidacja		Consolidation	Consolidation refers to an exudate or other product of disease that replaces alveolar air, rendering the lung solid (as in infective pneumonia).	
RID28493		Niedodma globalna		Atelectasis (global)	Atelectasis is reduced inflation of all or part of the lung. Reduced volume is seen, accompanied by increased atenuation (CT). Atelectasis is often associated with abnormal displacement of fissures, bronchi, vessels, diaphragm, heart, or mediastinum. The distribution can be lobar, segmental, or subsegmental. Local atelectasis is often qualified by descriptors such as linear, discoid, or platelike.	Radlex does not stand out division to global and local atelectasis.
		Niedodma lokalna		Local atelectasis		

Patryk Szatkowski

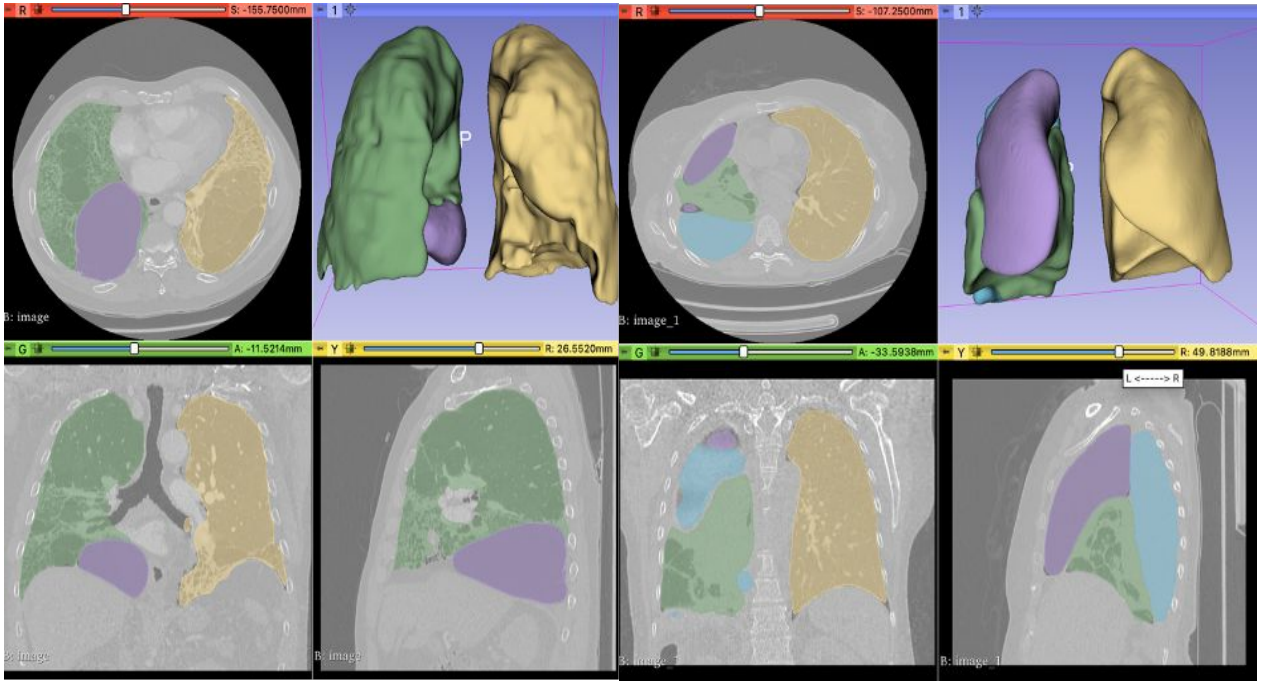
Data annotations - RadioTator



RadioTator is a tool developed particularly for annotation of significant amount of textual medical data - in this case from radiological reports.
In a short period of time we can categorize each result to one or few categories (classes) of pathologies due to medical ontology.

Patryk Szatkowski

Visual annotations - masks.



Using open-source software 3dSlicer we prepared masks for organs and some specific pathologies at CT scans.
Our work was concentrated on annotation of lung parenchyma (including atelectasis, consolidations, ground-glass opacites), pleural effusion and lung nodules.

Patryk Szatkowski

Diagnostic protocols

Diagnostic protocols will help doctors to manage the patients appropriately. We prepare protocols for two diseases - pneumonia and pneumothorax. Depending on findings from X-rays and CT scans (which are in fact characterized by classes from ontologies), protocol will suggest differential diagnosis (i.e. type of pneumonia), potential complications (i.e. pleural effusion, tension pneumothorax) and further management (conservative vs surgical treatment).

Przemysław Bombiński

Task IV

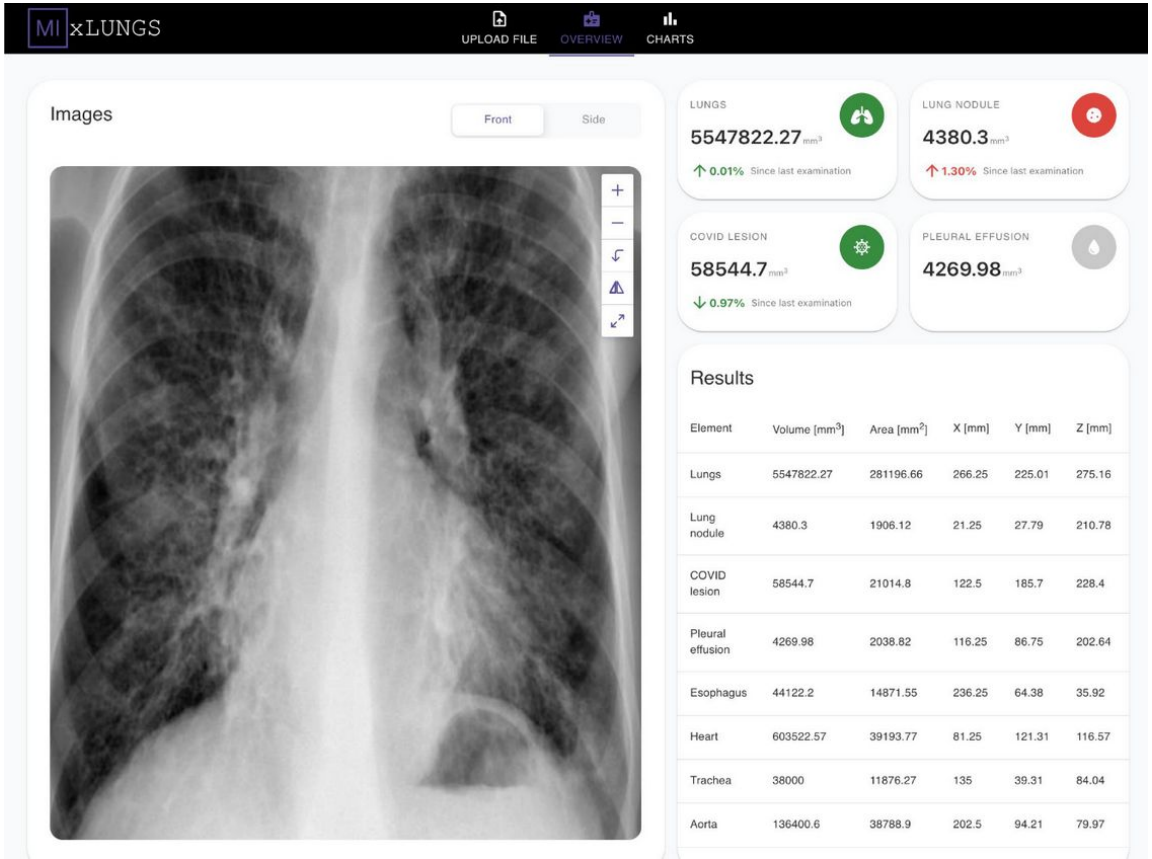
Goal: Turning results of previous tasks into the Application

- Argus - personal health report based on CT scan
- preparing infrastructure for models deployment
- maintaining structure of the model’s code

Argus

This web application reports abnormalities in body structures and lesions.

It checks multiple properties, including size, volume, surface, shape, and texture.

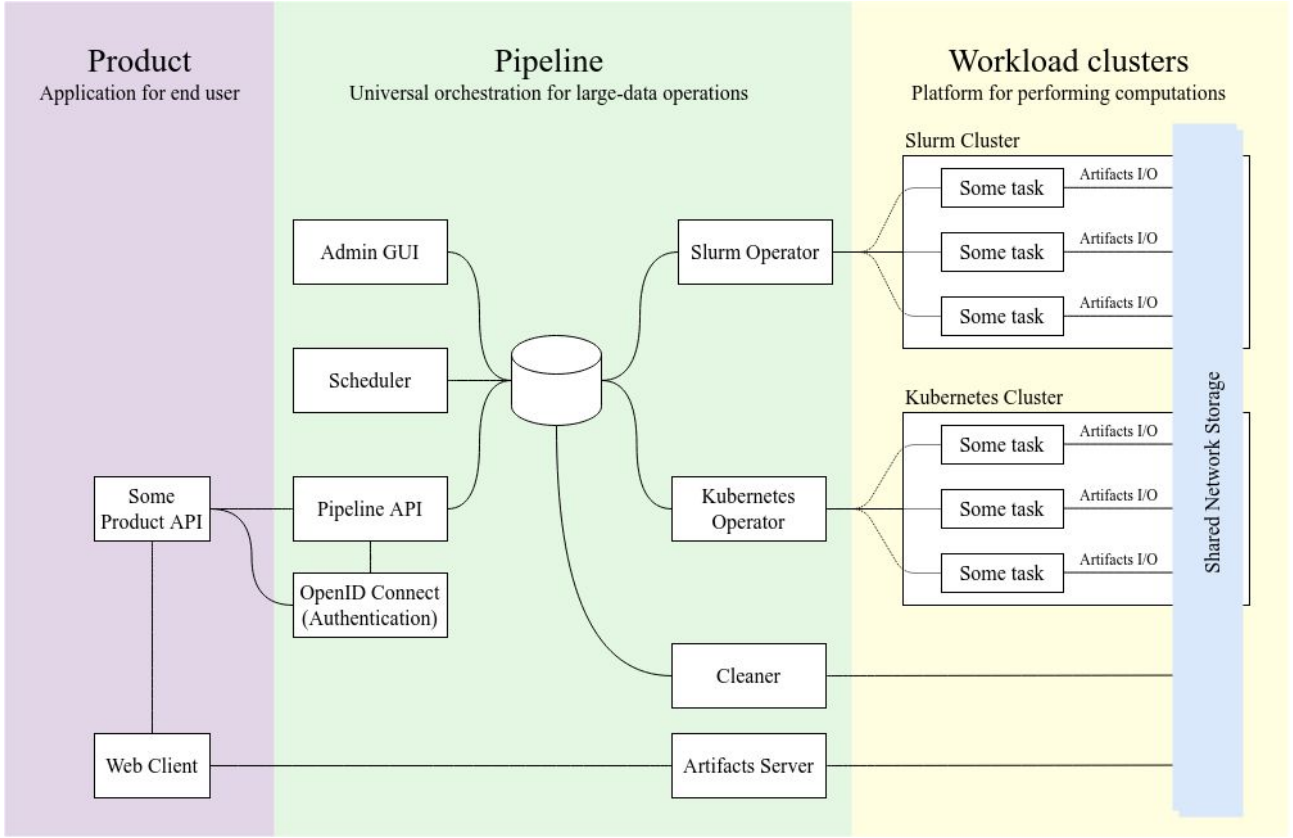


Piotr Piątyszek

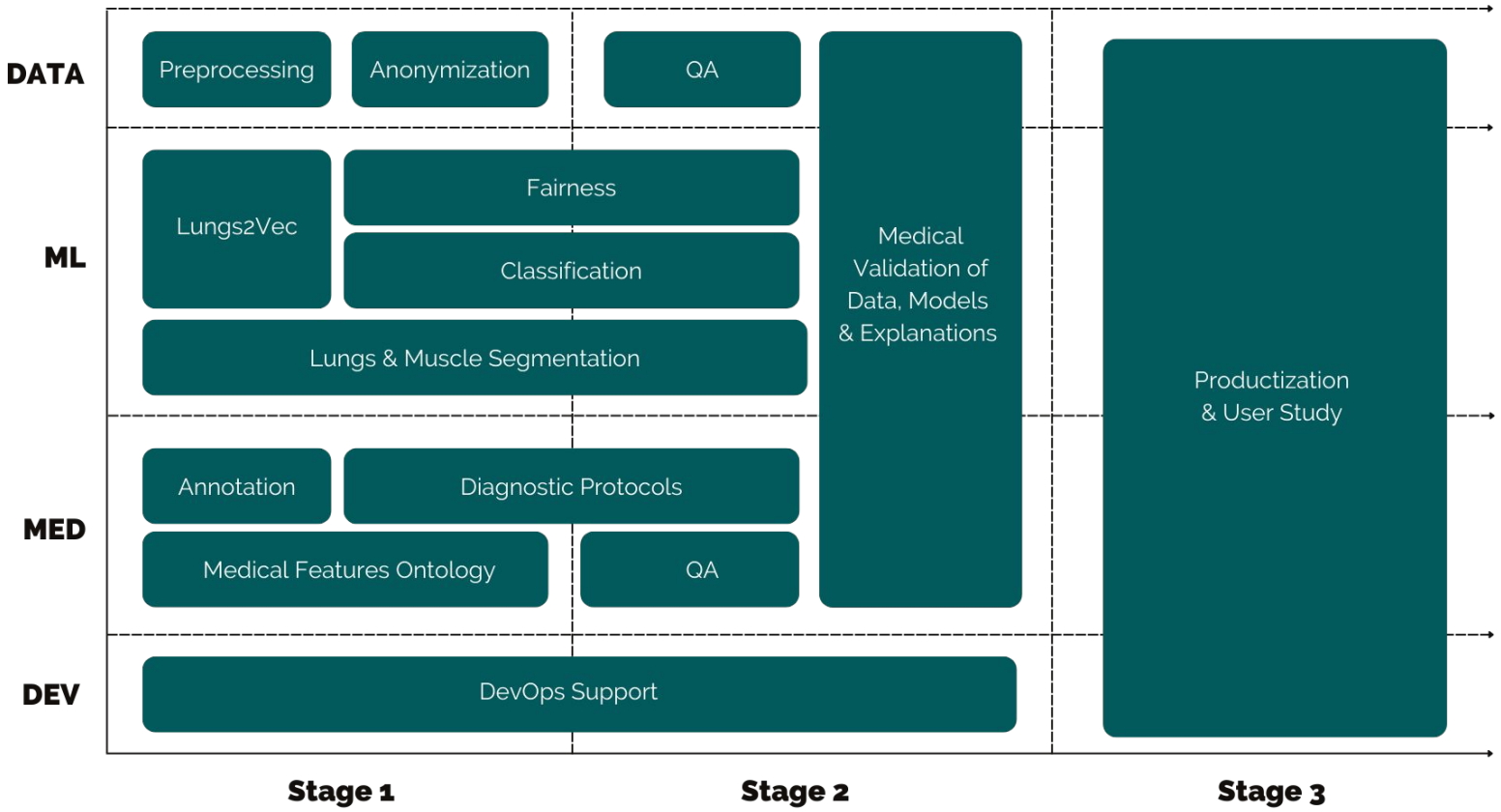
Pipeline

"Pipeline" creates space for medical models to be deployed, connected, and organized.

This project joins several existing and new technologies into the complete infrastructure.



Piotr Piątyszek


DATA
ML
MED
DEV
Preprocessing
Anonymization
QA
Lungs2Vec
Fairness
Classification
Lungs & Muscle Segmentation
Annotation
Diagnostic Protocols
Medical Features Ontology
QA
DevOps Support
Medical Validation of Data, Models & Explanations
Productization & User Study
Stage 1
Stage 2
Stage 3