Kunstmatige Intelligentie in Medische Beeldvorming

Associate Prof. Peter van Ooijen, MSc, PhD, CPHIMS
Dept. of Radiation Oncology, Associate Prof. AI in Medical Imaging
Data Science Center in Health, Coordinator Machine Learning Lab

Data Science Center in Health

Leerdoelen

- Inzicht krijgen in het proces van Kunstmatige Intelligentie
- Weten wat mogelijke valkuilen van KI zijn
- Besef van relevante wet- en regelgeving binnen de EU
- Kennis van bestaande methoden en checklists te gebruiken bij ontwikkeling en uitrol van KI

Machine Learning Lab

DASH
Data Science Center in Health

umcg

Stappen in het KI Process

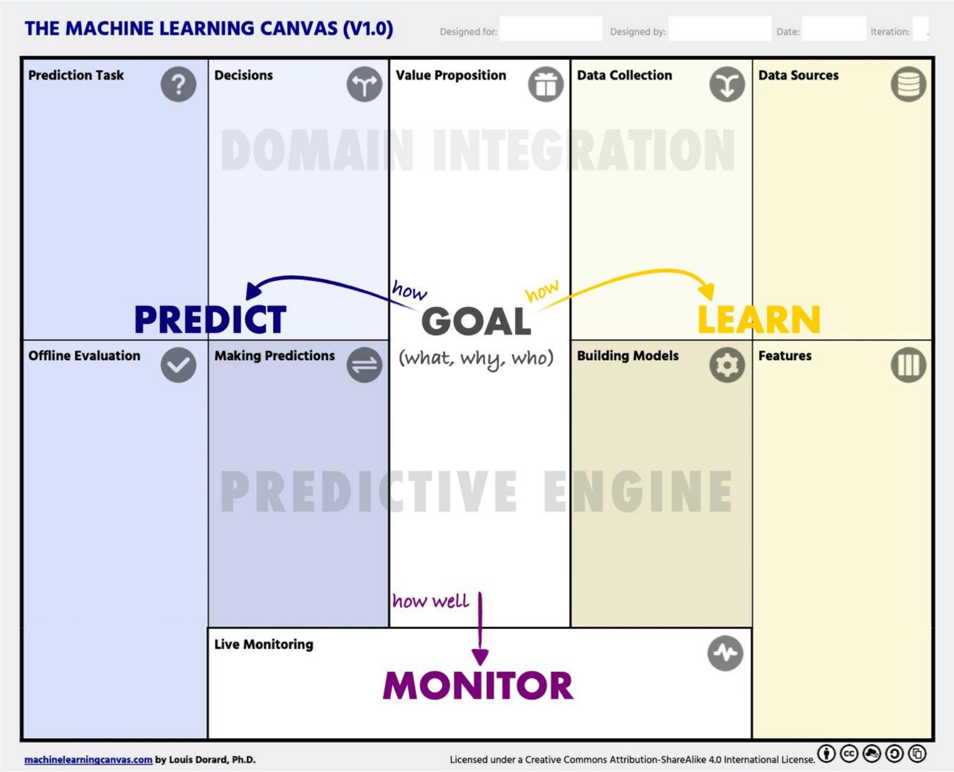
- Clinical Problem
- Data Collection
- Data Annotation
- Training
- Testing
- External Validation

Machine Learning Lab

DASH Data Science Center in Health | umcg

Start van een project
Machine Learning Canvas

- Louis Dorard
- https://www.ownml.co/machine-learning-canvas



Machine Learning Lab

DASH Data Science Center in Health | umcg

Waarom ML Canvas?

- Describe complex ML systems in a comprehensible and structured way.
- ***Early engagement of stakeholders*** in the ML development process
- Get the key elements of a project
- Assess feasibility
- Detect bottlenecks and technical constraints early on
- ***Collaboration in the team***
- Plan work and choose the right tech

THE MACHINE LEARNING CANVAS (V1.0)

Designed for: Designed by: Date: Iteration:

Prediction Task

Type of task? Entity on which predictions are made? Possible outcomes? Wait time before observation?

Decisions

How are predictions turned into proposed value for the end-user? Mention parameters of the process / application that does that.

Value Proposition

Data Collection

Strategy for initial train set & continuous update. Mention collection rate, holdout on production entities, cost/constraints to observe outcomes

Data Sources

Where can we get (raw) information on entities and observed outcomes? Mention database tables, API methods, websites to scrape, etc

DOMAIN INTEGRATION

how PREDICT

how LEARN

GOAL

(what, why, who)

Who is the end-user? What are their objectives? How will they benefit from the ML system? Mention workflow/interfaces.

how well

Offline Evaluation

Methods and metrics to evaluate the system before deployment?

Making Predictions

When do we make real-time / batch pred.? Time available for this + featurization + post-processing? Compute target?

PREDICTIVE ENGINE

Building Models

How many prod models are needed? When would we update? Time available for this (including featurization and analysis)?

Features

Input representations available at prediction time, extracted from raw data sources.

Live Monitoring

Metrics to quantify value creation and measure the ML system's impact in production (on end-users and business)?

Pneumothorax: definition and diagnosis

Pneumothorax (PTX)

- Air in the pleural space, causing partial collapse of the lung
- Occurrence of 15% in blunt thorax trauma cases
- Has to be treated prior to mechanical ventilation to avoid tension PTX (can lead to shock or death)

Diagnosis

- CT is gold standard
- Chest X-ray recommended for initial assessment (ATLS guidelines): fast, simple, cheap, always possible

DASH
Data Science Center in Health

Decisions

How are predictions used to make decisions that provide the proposed value to the end-user?

Provide the score of the Xray to the emergency physician. Based on the score the physician will make a decision on how to further treat the patient or if additional diagnosis (e.g. performing Computed Tomography) is needed.

ML task

Input, output to predict, type of problem.

We want to be able to answer the question "*Is there a pneumothorax in the presented thoracic X Ray?*" right after performing the X Ray.

Input: X Ray image

Output:"Pneumothorax" or "no Pneumothorax" => Binary Classification

Identification of the pneumothorax region => image segmentation

Value Propositions

What are we trying to do for the end-user(s) of the predictive system? What objectives are we serving?

Provide automated diagnosis of pneumothorax in thoracic X-ray images in the emergency department as input for the emergency physician.

Aim to facilitate the process and avoid delays in treatment of patients by faster diagnosis of the -ray images.

In case of doubt, expert review by a radiologist is still required.

Data Sources

Which raw data sources can we use (internal and external)?

X-rays acquired at emergency department

Diagnosis description of the images. Weak or strong labels?

Collecting Data

How do we get new data to learn from (inputs and outputs)?

PACS Export images + Structured Reports

Manual tagging of the images by a human expert.

Alternative could be to automatically retrieve the tagging using NLP from the structured report by the radiologist.

Making Predictions

When do we make predictions on new inputs? How long do we have to featurize a new input and make a prediction?

Predictions have to be made right after the thoracic X Ray is required. The prediction should be provided in seconds because of the emergency setting in which waiting is not possible.

Offline Evaluation

Methods and metrics to evaluate the system before deployment.

False negative predictions should be avoided. System should be fine-tuned for this.

Features

Input representations extracted from raw data sources.

Image features of the thoracic Xrays

Diagnosis information (yes or no pneumothorax) from free text reports using NLP

Building Models

When do we create/update models with new training data? How long do we have to featurize training inputs and create a model?

One model is built for general use based on the retrospectively collected data.

feedback from the expert users will be collected during operation.

New model will be trained after collecting feedback for 6 months. Only 'frozen' models can be used in the clinic, no continuous learning.

Live Evaluation and Monitoring

Methods and metrics to evaluate the system after deployment, and to quantify value creation.

Changes of acquisition setup should be tracked (could influence performance of the trained network).

Comparison of processing time of thoracic X-ray before and after AI implementation.

Recording the agreement and disagreement of the emergency physician with the decision as made by AI

DASH
Data Science Center in Health

Data Collection

Machine Learning Lab

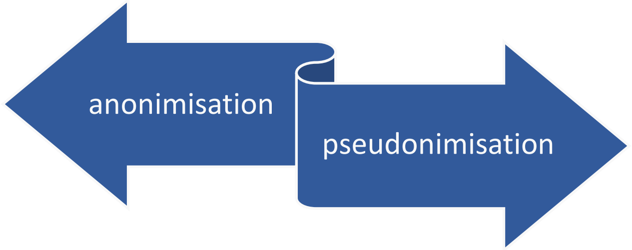
DASH
Data Science Center in Health

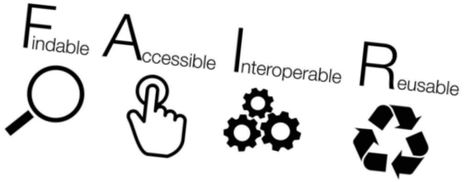
umcg

Data Collection

- Probably the most important part of the process
- Very often the limiting factor

- Data Quality
- Standardization
- Small data set
- Bias





https://www.emedicinehealth.com/informed_consent/article_em.htm

Machine Learning Lab

DASH Data Science Center in Health | umcg

Datasheets for Datasets

arXiv.org > cs > arXiv:1803.09010

Search...

Help | Advanced

Computer Science > Databases

[Submitted on 23 Mar 2018 (v1), last revised 19 Mar 2020 (this version, v7)]

Datasheets for Datasets

Timnit Gebru, Jamie Morgenstern, Briana Vecchione, Jennifer Wortman Vaughan, Hanna Wallach, Hal Daumé III, Kate Crawford

The machine learning community currently has no standardized process for documenting datasets, which can lead to severe consequences in high-stakes domains. To address this gap, we propose datasheets for datasets. In the electronics industry, every component, no matter how simple or complex, is accompanied with a datasheet that describes its operating characteristics, test results, recommended uses, and other information. By analogy, we propose that every dataset be accompanied with a datasheet that documents its motivation, composition, collection process, recommended uses, and so on. Datasheets for datasets will facilitate better communication between dataset creators and dataset consumers, and encourage the machine learning community to prioritize transparency and accountability.

Comments: Working Paper, comments are encouraged

Subjects: **Databases (cs.DB)**; Artificial Intelligence (cs.AI); Machine Learning (cs.LG)

Cite as: arXiv:1803.09010 **[cs.DB]**

(or arXiv:1803.09010v7 **[cs.DB]** for this version)

Data sheets for datasets

- Motivation
- Composition
- Collection process
- Preprocessing/cleaning/labeling
- Uses
- Distribution
- Maintenance

Machine Learning Lab

DASH Data Science Center in Health | umcg

https://arxiv.org/abs/1803.09010
https://www.microsoft.com/en-us/research/project/datasheets-for-datasets/

Data Collection

European Radiology
https://doi.org/10.1007/s00330-020-07453-w

LETTER TO THE EDITOR

COVID-19, AI enthusiasts, and toy datasets: radiology without radiologists

H. R. Tizhoosh[1,2] · Jennifer Fratesi[3]

Received: 4 September 2020 / Revised: 23 September 2020 / Accepted: 2 November 2020

Machine Learning Lab

DASH Data Science Center in Health

umcg

Pneumothorax detection

- Dataset already available in public domain
  - Con: No influence on data collection/selection
  - Pro: easy to start with

- ChestXray14[1]
- 112.120 Xrays

DASH
Data Science Center in Health

[1]X. Wang, et al. ChestX-ray8: Hospital-scale Chest X-ray Database and Benchmarks on Weakly- Supervised Classification and Localization of Common Thorax Diseases, IEEE CVPR, pp. 3462-3471,2017

Data Annotation

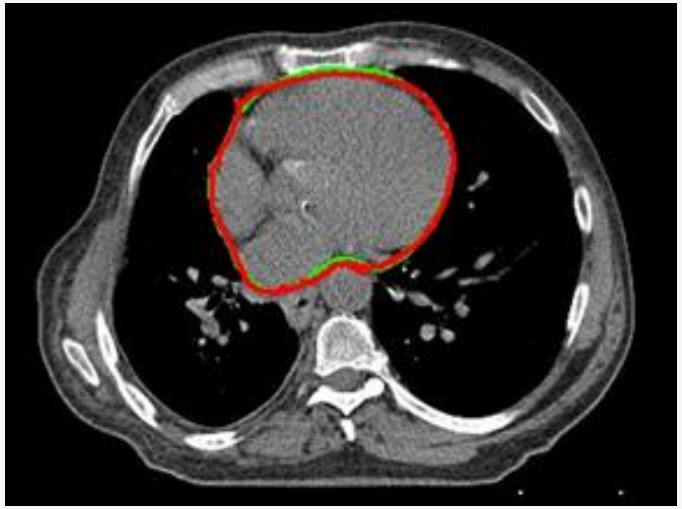
Data Annotation/Labeling

- Annotation/Label quality!
  - Appropriate for the task at hand? (global/local labels)
  - Class balance? (the more imbalance, the more difficult the task, the more data likely needed to solve it)
  - Consistency? (everyone used the same labeling guidelines/procedure?)
  - Label format? (computer readable? Or hidden in e.g. radiology report?)

Machine Learning Lab

DASH Data Science Center in Health | umcg

Data Annotation/Labeling



Heart segmentation:
Local, pixel level labels

Fracture detection:
Global or local labels?

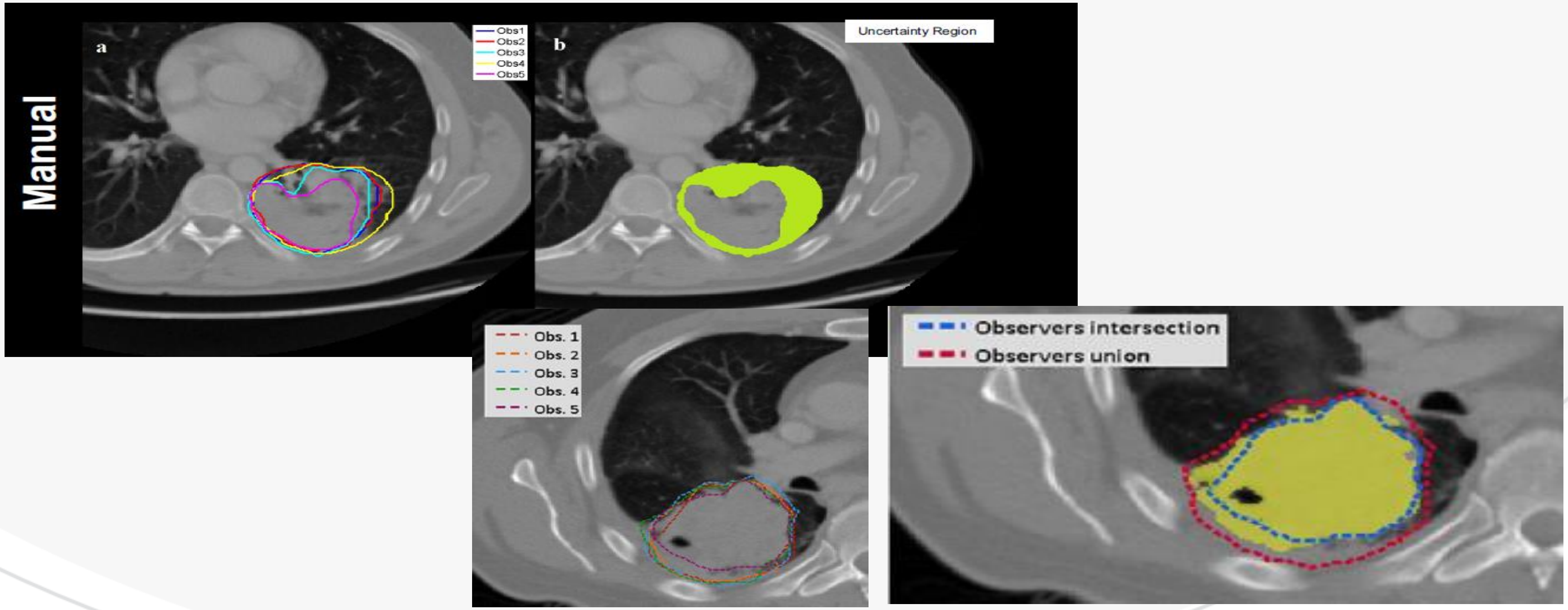
Machine Learning Lab

DASH Data Science Center in Health | umcg

Data Annotation/Labeling

- Manual delineation of regions of interest (ROI)
  - Time consuming
  - inter- en intra-observer variability
- Semi-automatic: some user interaction
  - Dependent on starting ROI
  - Variability of seedpoint selection
  - Less user dependent than manual

Data Annotation/Labeling

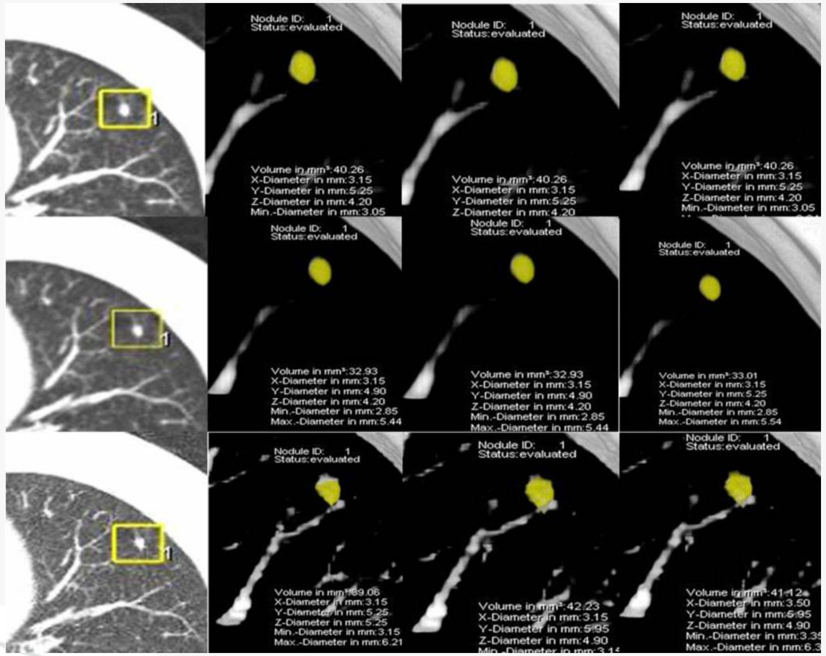


Velazquez et al. Radiotherapy and Oncology 2012 & Scientific Reports 2013

Machine Learning Lab

DASH Data Science Center in Health | umcg

Data Annotation/Labeling



- Variation in CT reconstruction algorithm
- 1 mm soft kernel, 2 mm soft kernel, and 2 mm sharp kernel

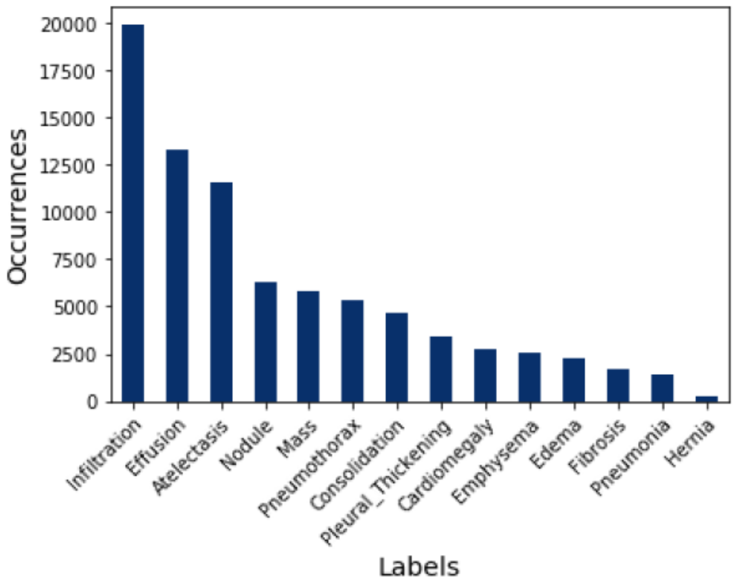
Machine Learning Lab

DASH Data Science Center in Health | umcg

Wang et al. Eur Radiol 2010

Pneumothorax detection

- Publicly available X-ray dataset: ChestXray14[1]
- 112.120 Xrays
- 14 labeled diseases



- ~60.000 images with ‘No Findings’

DASH Data Science Center in Health

[1]X. Wang, et al. ChestX-ray8: Hospital-scale Chest X-ray Database and Benchmarks on Weakly- Supervised Classification and Localization of Common Thorax Diseases, IEEE CVPR, pp. 3462-3471,2017

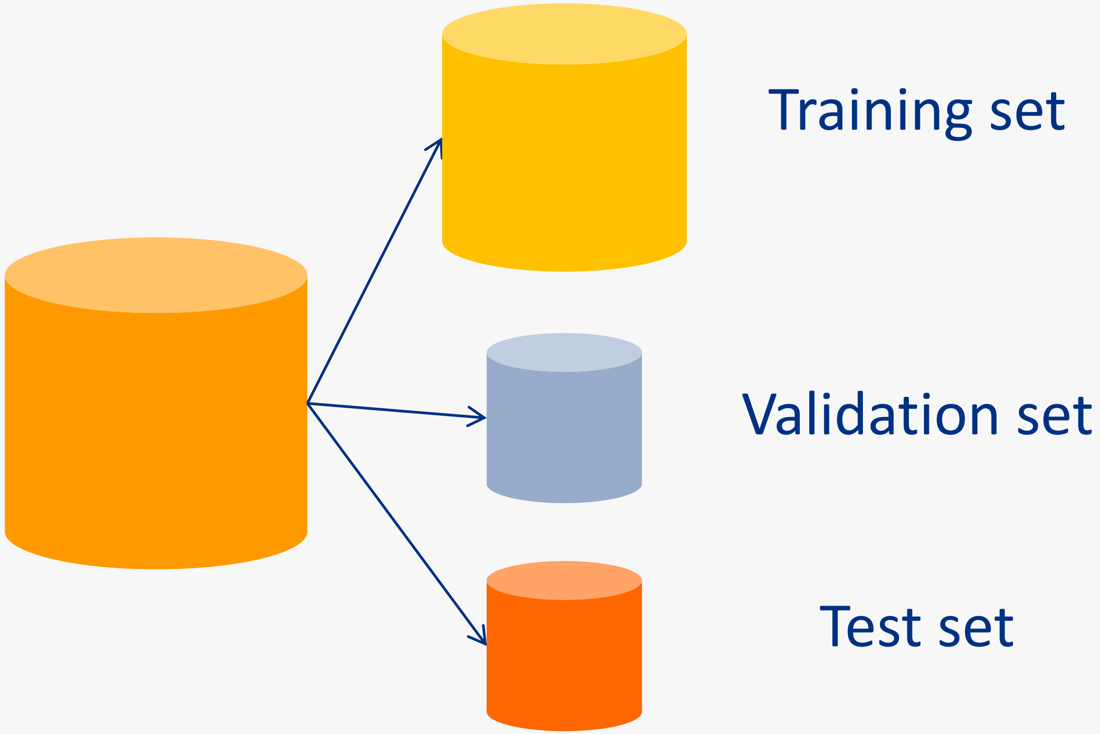
Model training

Machine Learning Lab

DASH Data Science Center in Health

umcg

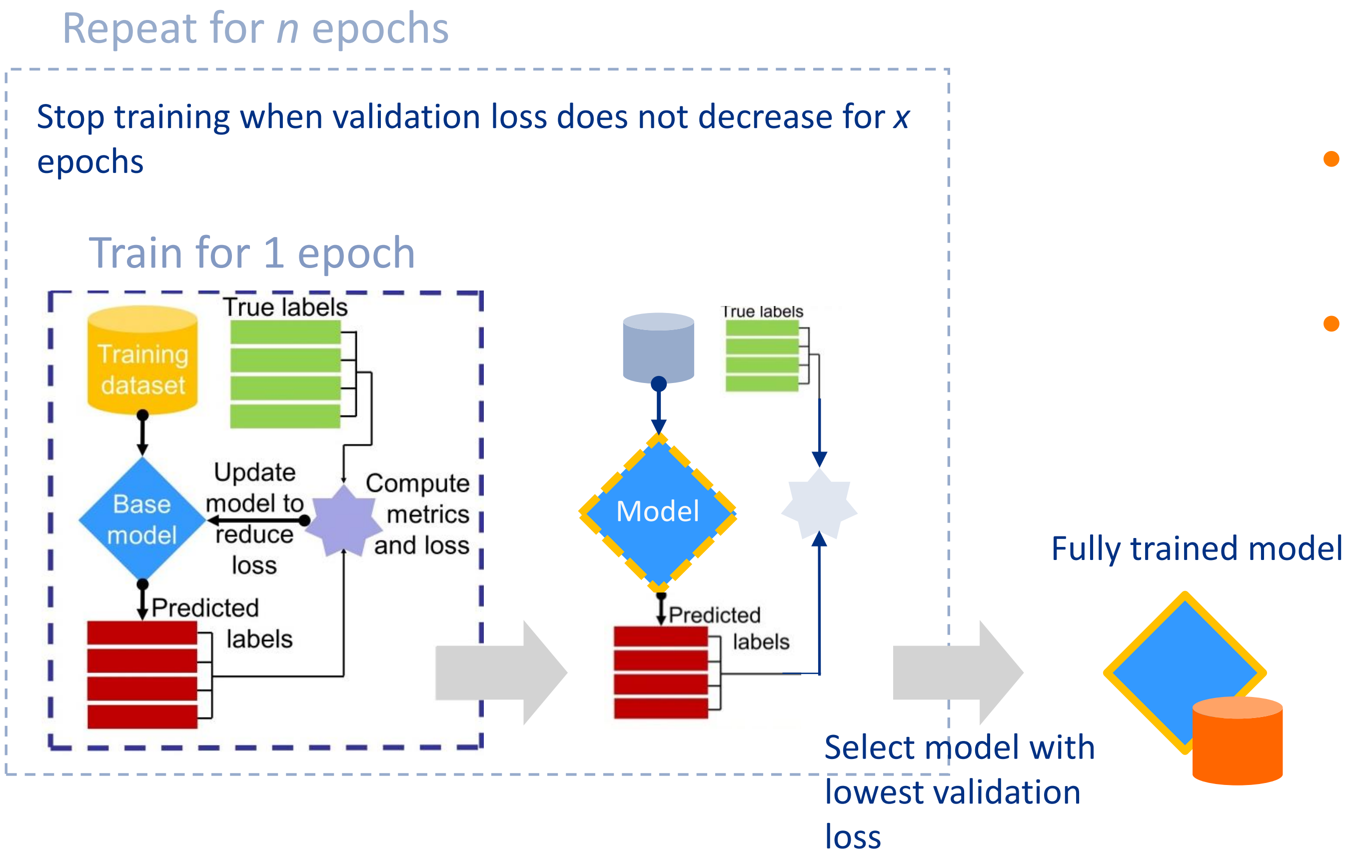
Data split



Machine Learning Lab

DASH Data Science Center in Health | umcg

So how does a computer learn?



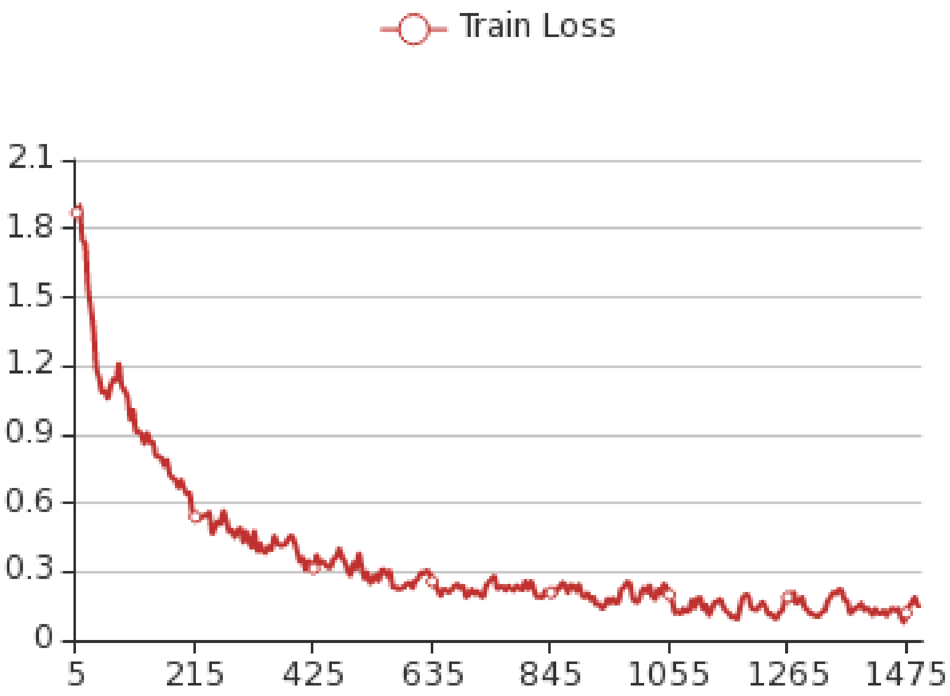
Epoch: one iteration over the full dataset

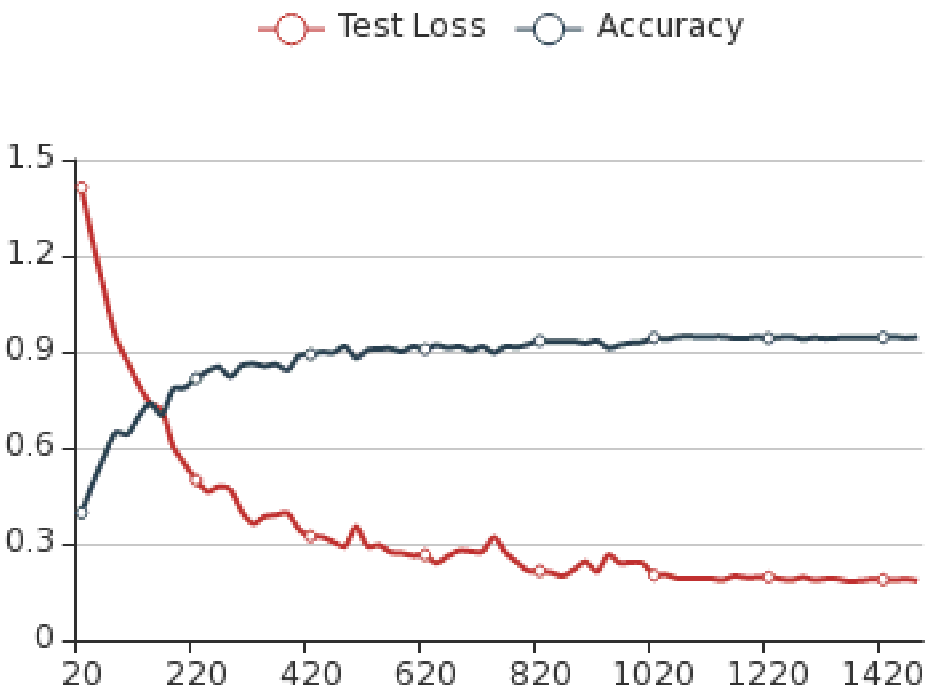
- Learns by calculating how wrong it is: The loss
- Training phase:
  - Provide input and the correct output
  - Computer predicts the output
  - Calculates the error from the correct output
  - Reiterates

Machine Learning Lab

DASH Data Science Center in Health | umcg

Loss graph


Train Loss
2.1
1.8
1.5
1.2
0.9
0.6
0.3
0
5
215
425
635
845
1055
1265
1475


Test Loss
Accuracy
1.5
1.2
0.9
0.6
0.3
0
20
220
420
620
820
1020
1220
1420

Model training

InceptionV3 architecture[1]

Input: 299x299

Data augmentation:

- Rotations $\alpha \in [-25°, +25°]$
- Translations $\Delta x, \Delta y \in [-0.1, +0.1]$
- Brightness $\Delta B \in [-0.3, +0.3]$

Dataset split:

- Train/validation/test: 80/10/10

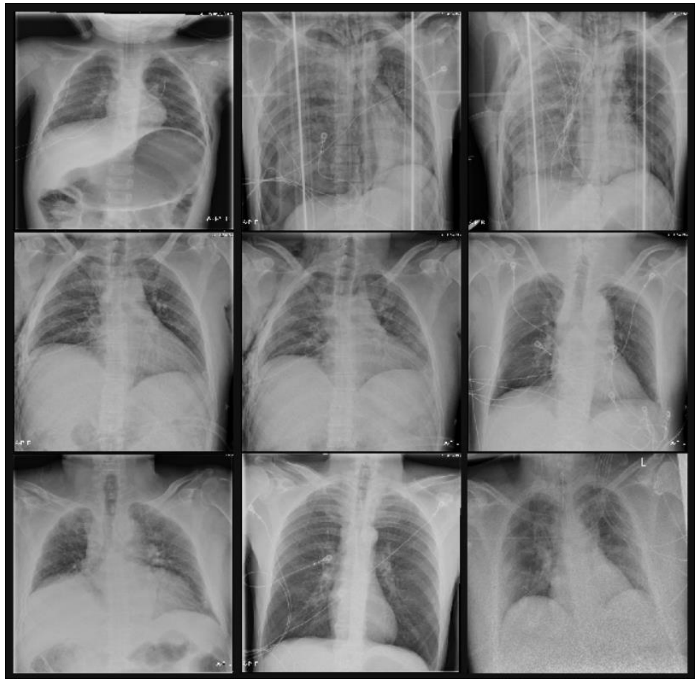
Trained for 100 epochs

Data Science Center in Health

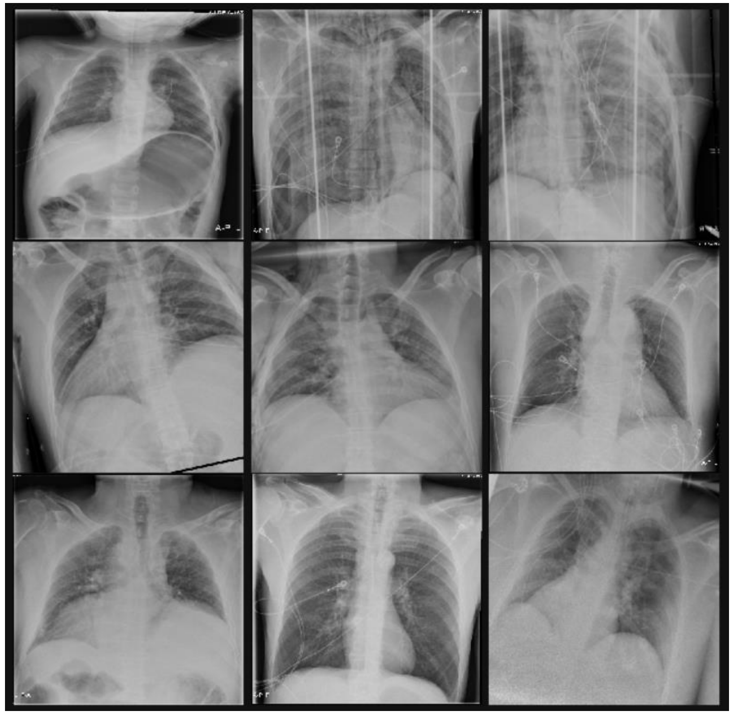
1. Szegedy C, Vanhoucke V, Ioffe S, Shlens J (2016) Rethinking the Inception Architecture for Computer Vision. 2016 IEEE Conf Comput Vis Pattern Recognit.

Data augmentation

No augmentations



Random horizontal flip
Random rotation/scaling/translation
Random brightness/contrast adjustment



DASH
Data Science Center in Health

Model Validation and Testing

Machine Learning Lab

DASH
Data Science Center in Health

umcg

Example Carotid artery ultrasound

AI training Summary	
Task	Segmentation of carotid artery lumen
Number of images	1060 labeled, 2500 total
Annotation procedure	Manual annotation
Performance measure	Pixelwise dice coefficient $$dc = \frac{2TP}{2TP + FP + FN}$$
Current model performance on test set	$dc = 0.93$
Train/validation/test split	848/106/106
Network architecture	Unet

Machine Learning Lab

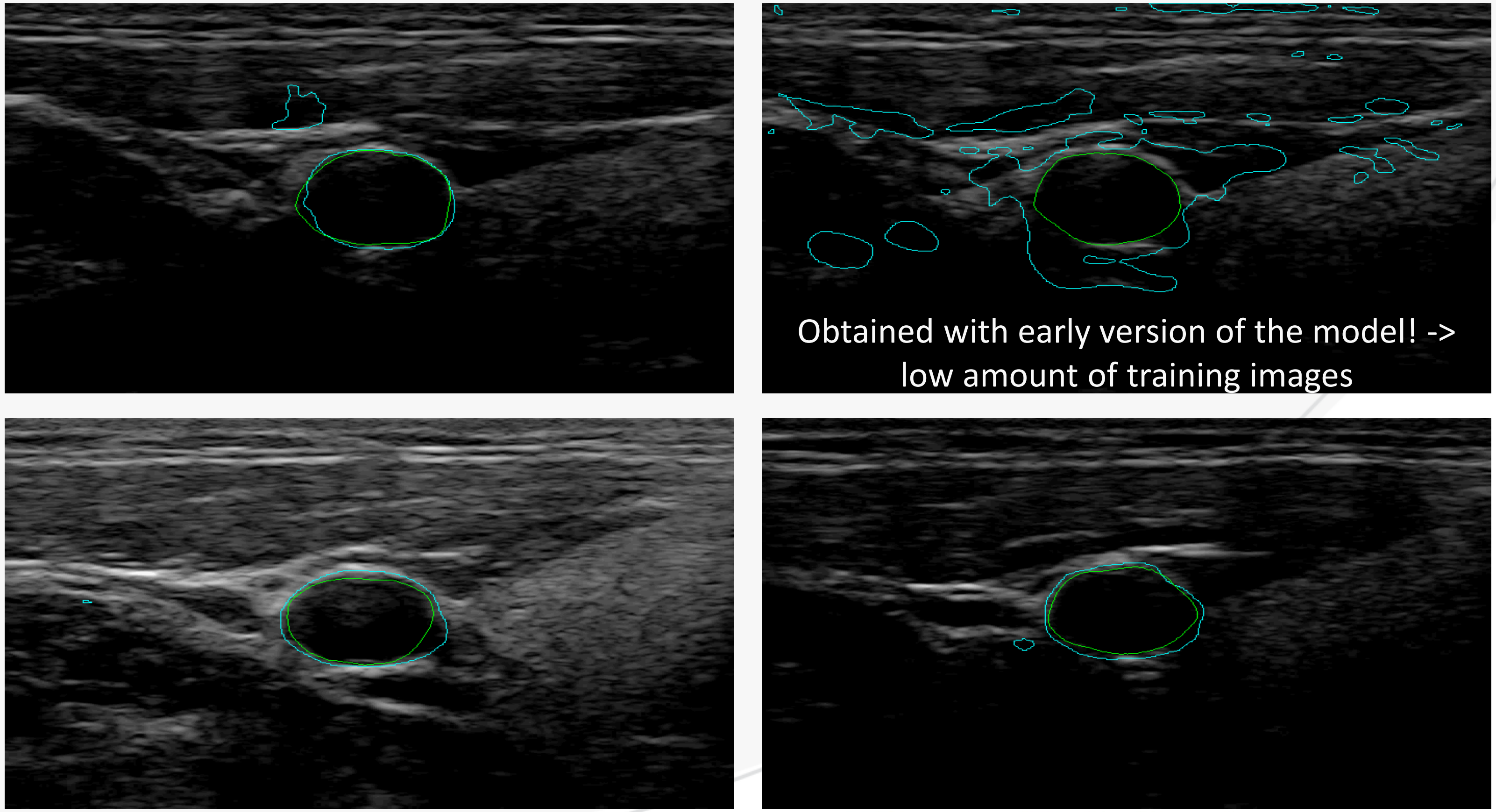
DASH Data Science Center in Health

umcg

Examples of low dice results

Ground Truth

AI output



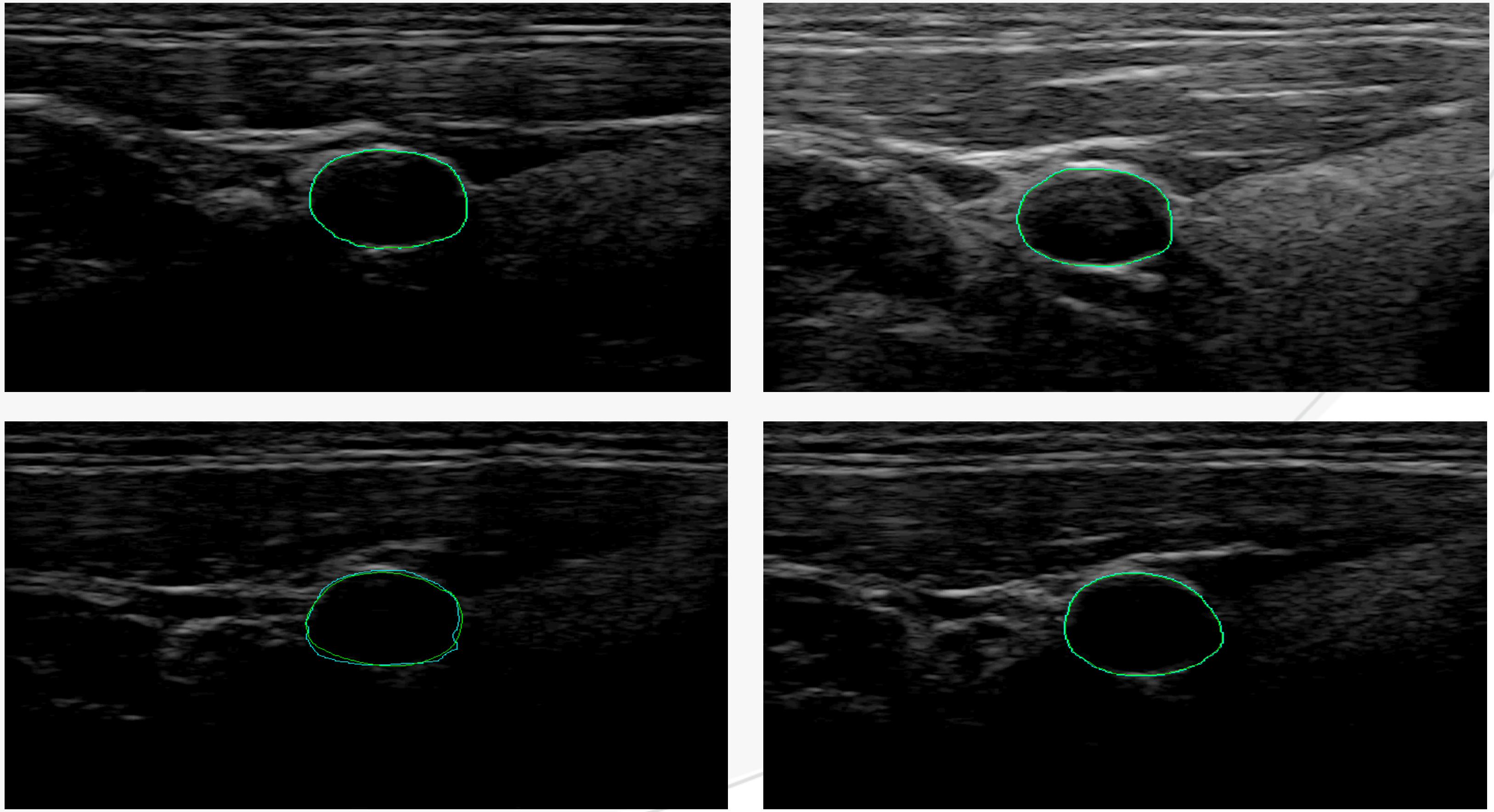
Machine Learning Lab

DASH Data Science Center in Health | umcg

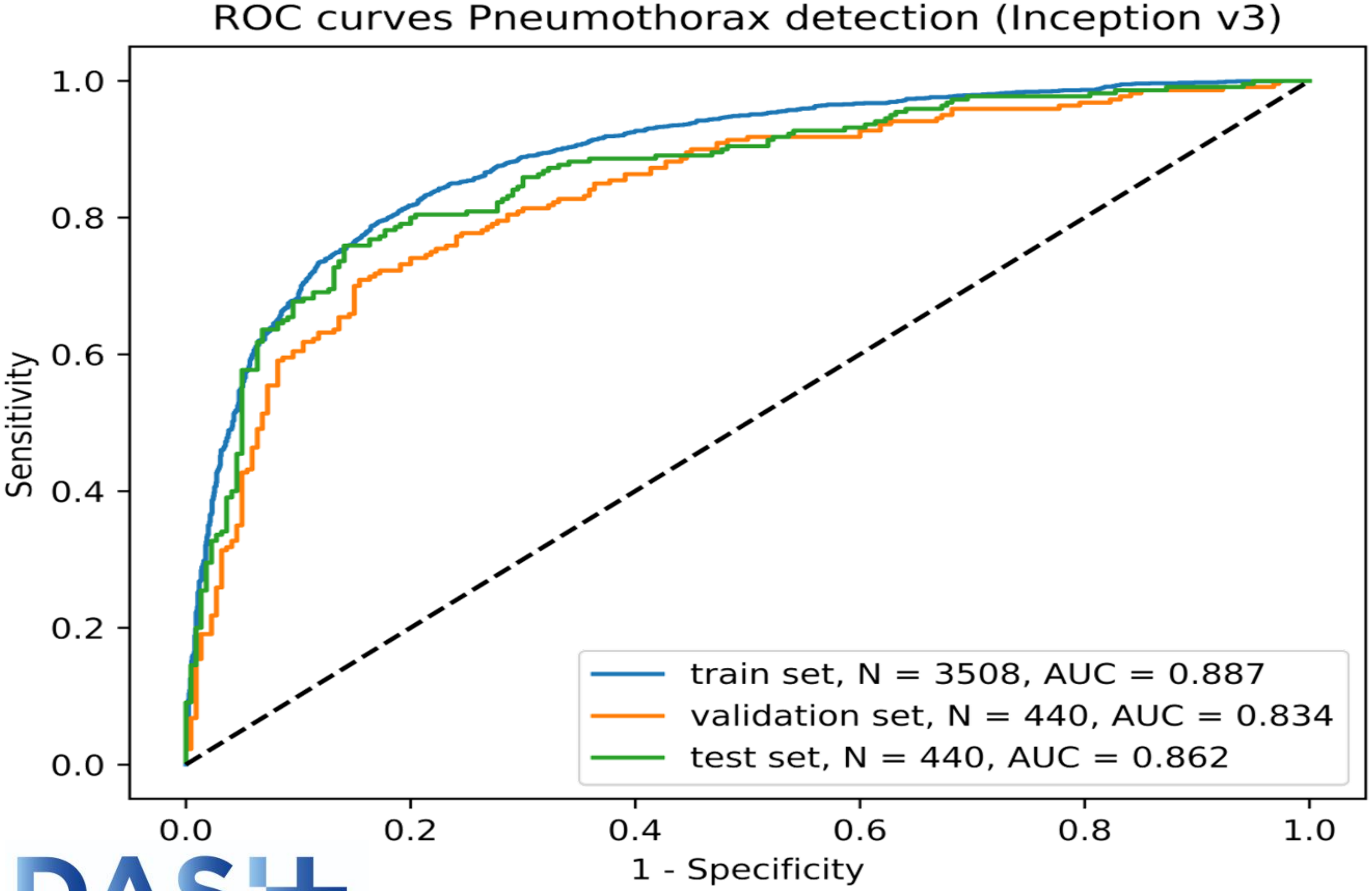
Examples of high dice results

Ground Truth

AI output



Model Validation and Testing



DASH
Data Science Center in Health

Explainability

Understanding “why” the network arrives at its prediction

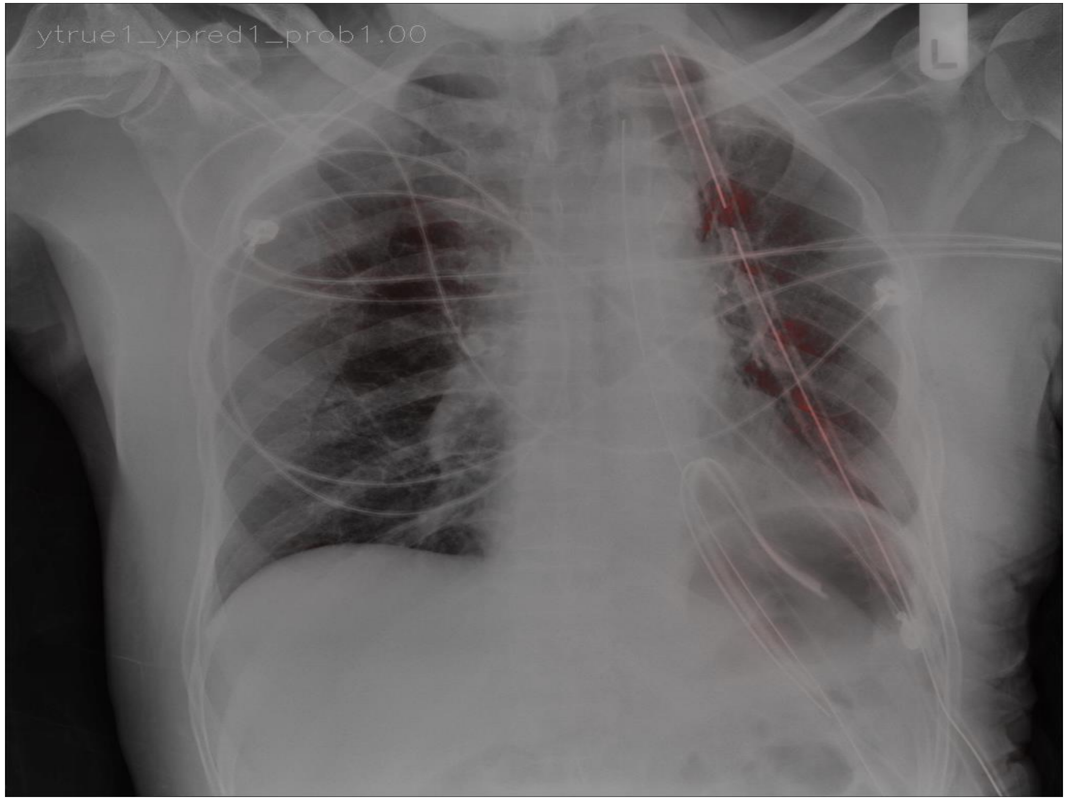
1. Failure mode analysis
    - Inspect misclassified samples, see if there is a pattern
2. Do the classification pixel-wise (i.e. make it a segmentation problem)
    - Requires contours as labels (often not available)
    - Requires different model architecture
3. “Saliency map” or “heatmap” of pixel relevance
    - Indicates how much each pixel in the input contributes to model output
    - Layer-wise relevance propagation

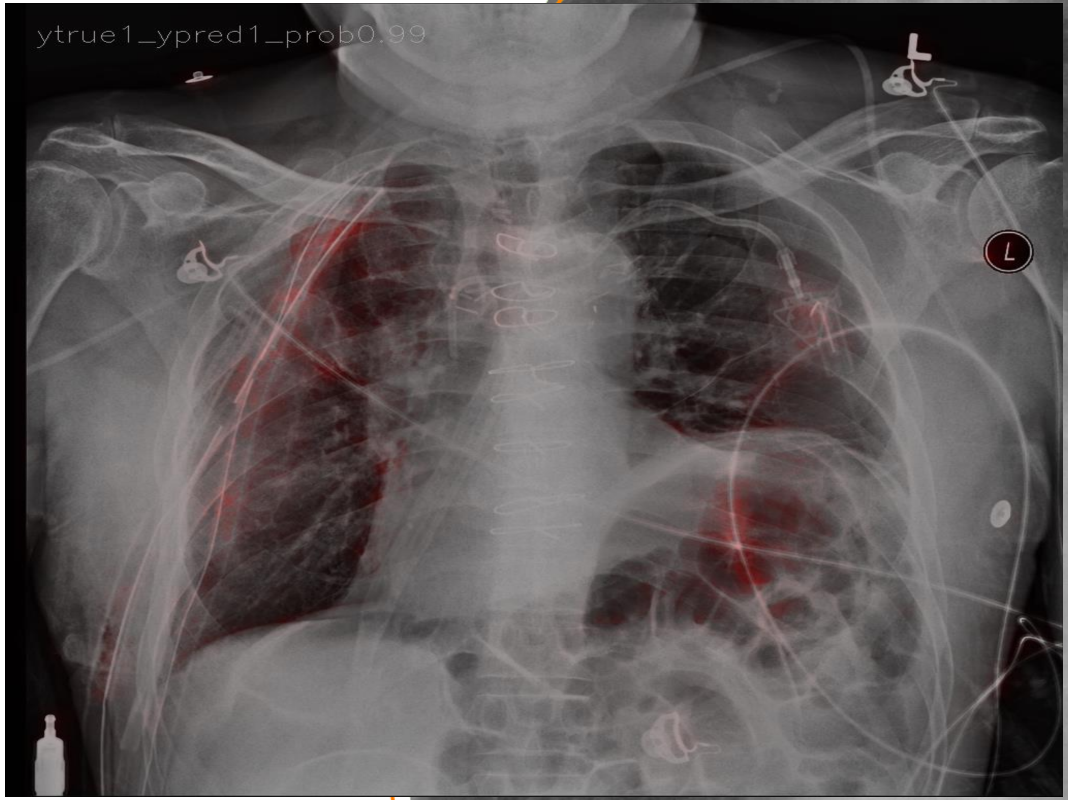
DASH
Data Science Center in Health

Explainability – Grad-CAM

DASH
Data Science Center in Health

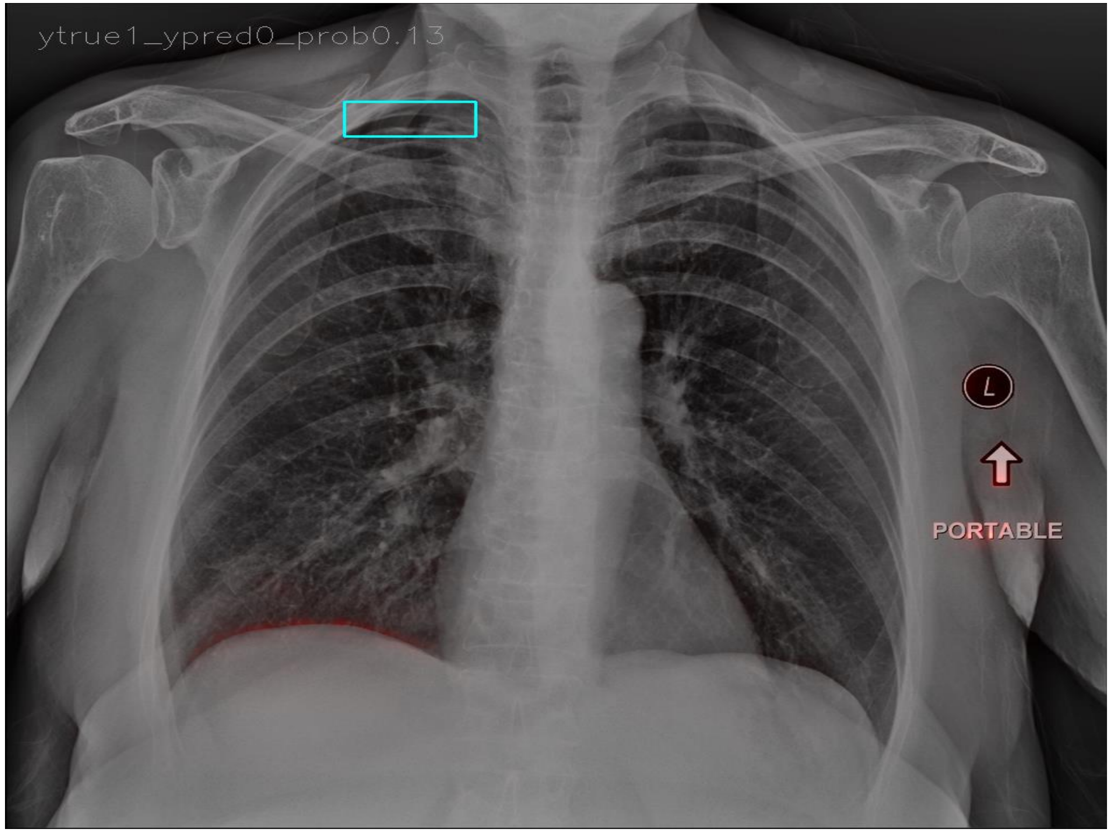
Explainability – Deep Taylor Decomposition

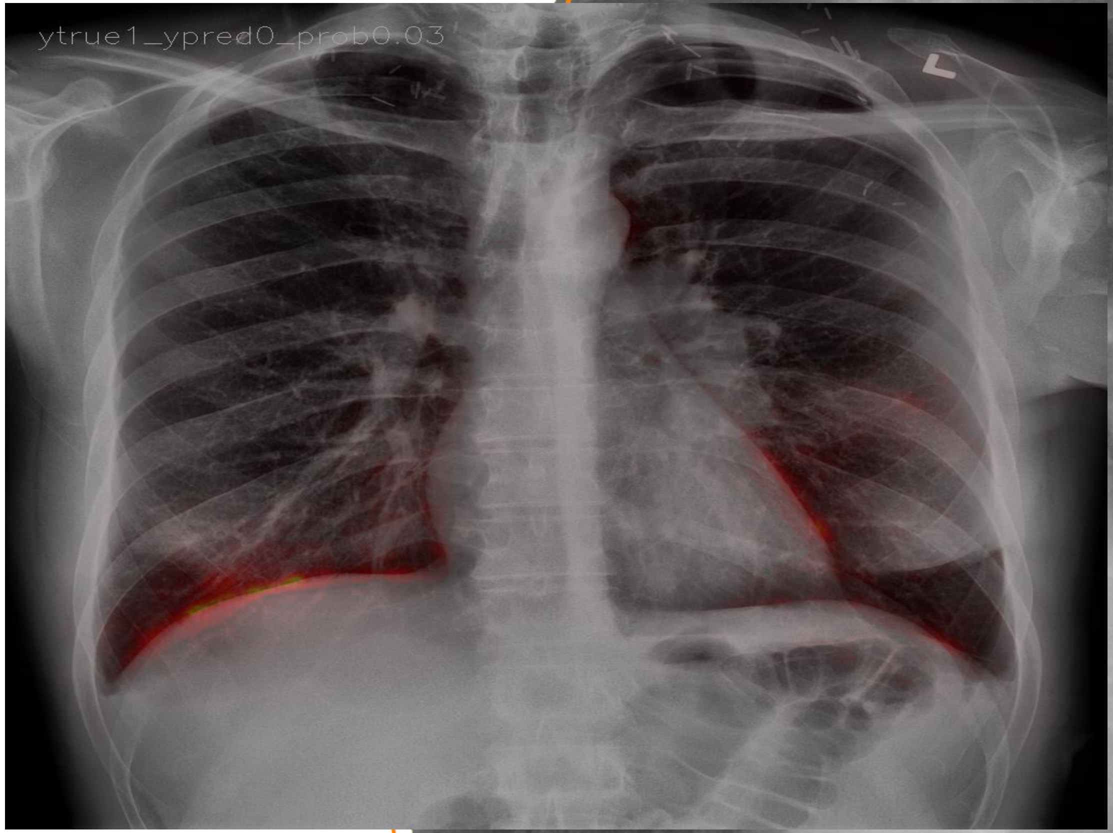




DASH
Data Science Center in Health

Explainability – Deep Taylor Decomposition

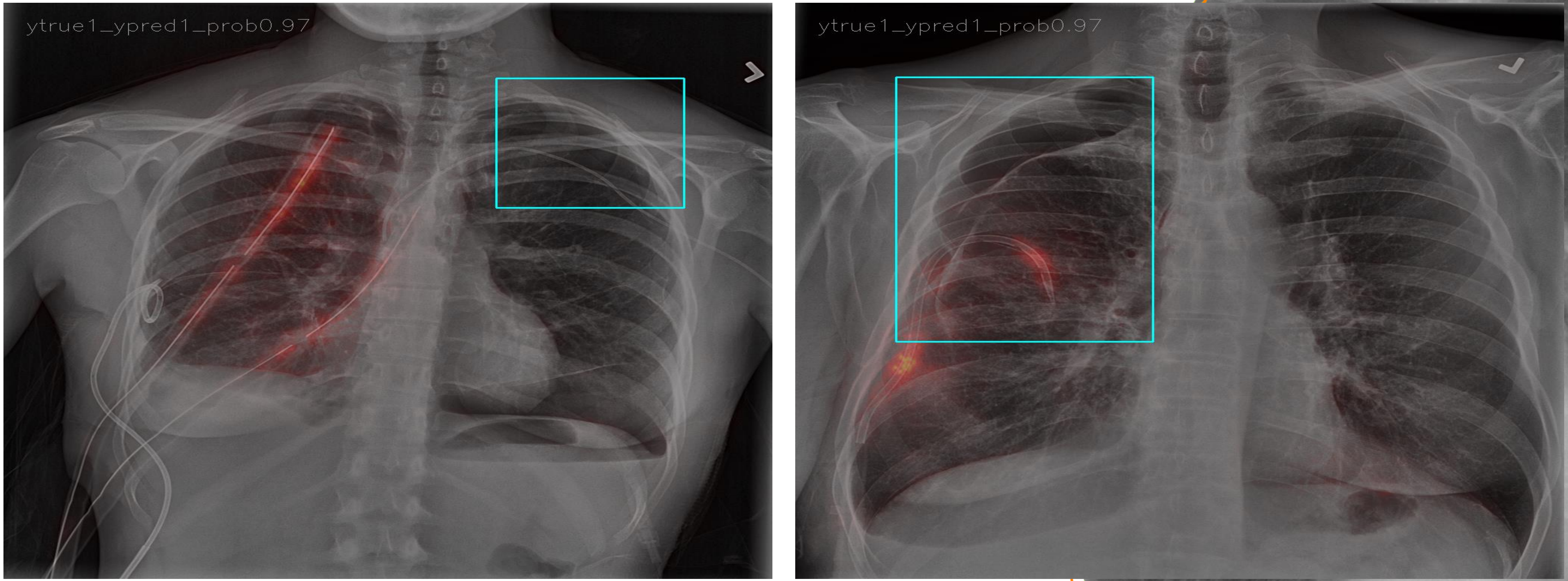




FALSE NEGATIVES

DASH
Data Science Center in Health

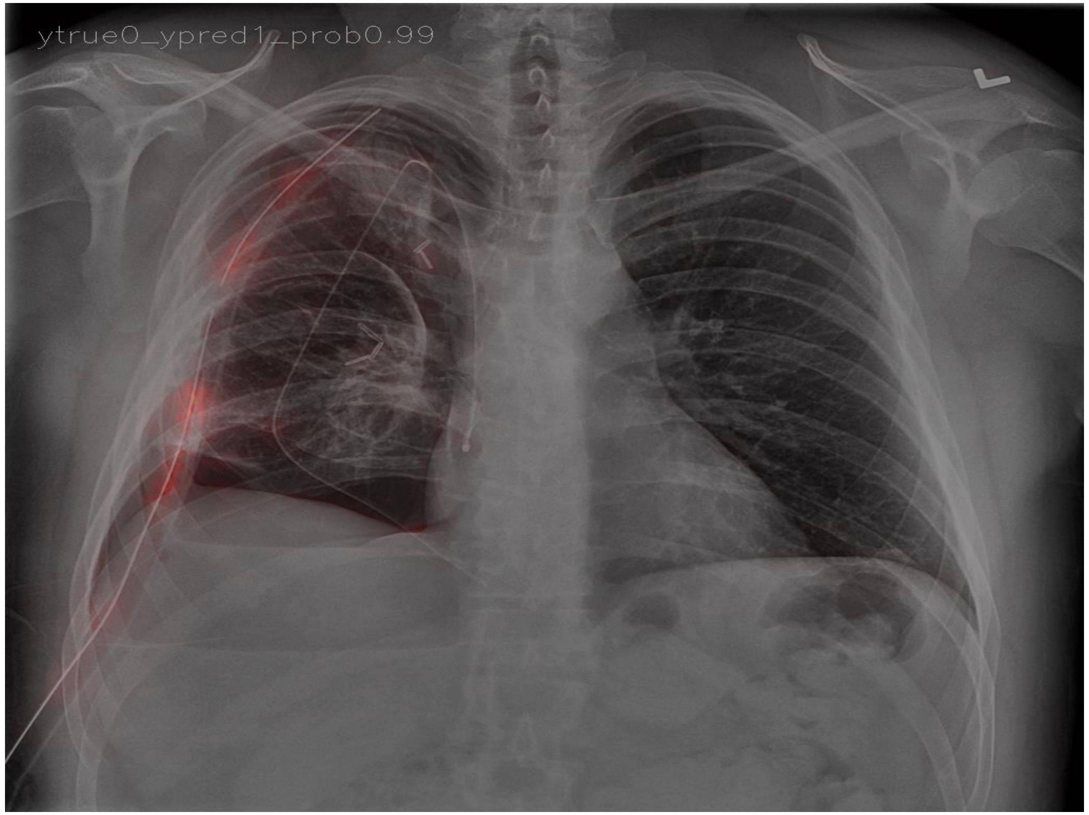
Explainability – Deep Taylor Decomposition

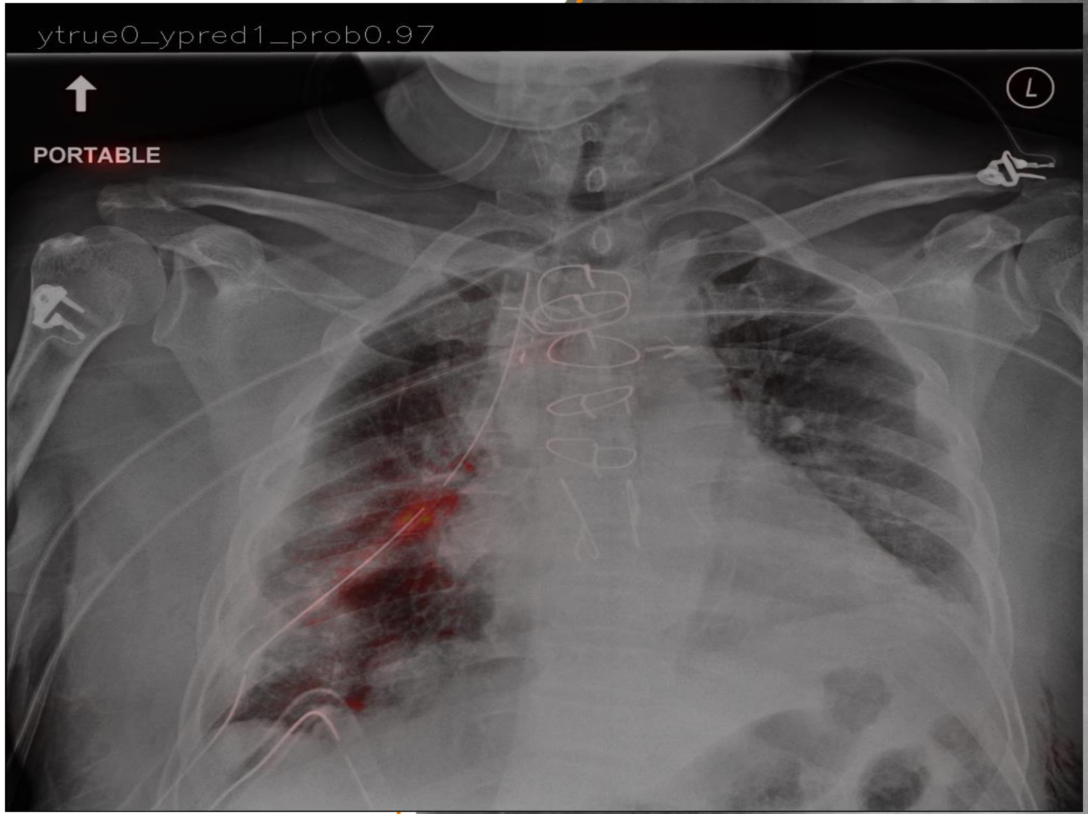


TRUE POSITIVES

DASH
Data Science Center in Health

Explainability – Deep Taylor Decomposition

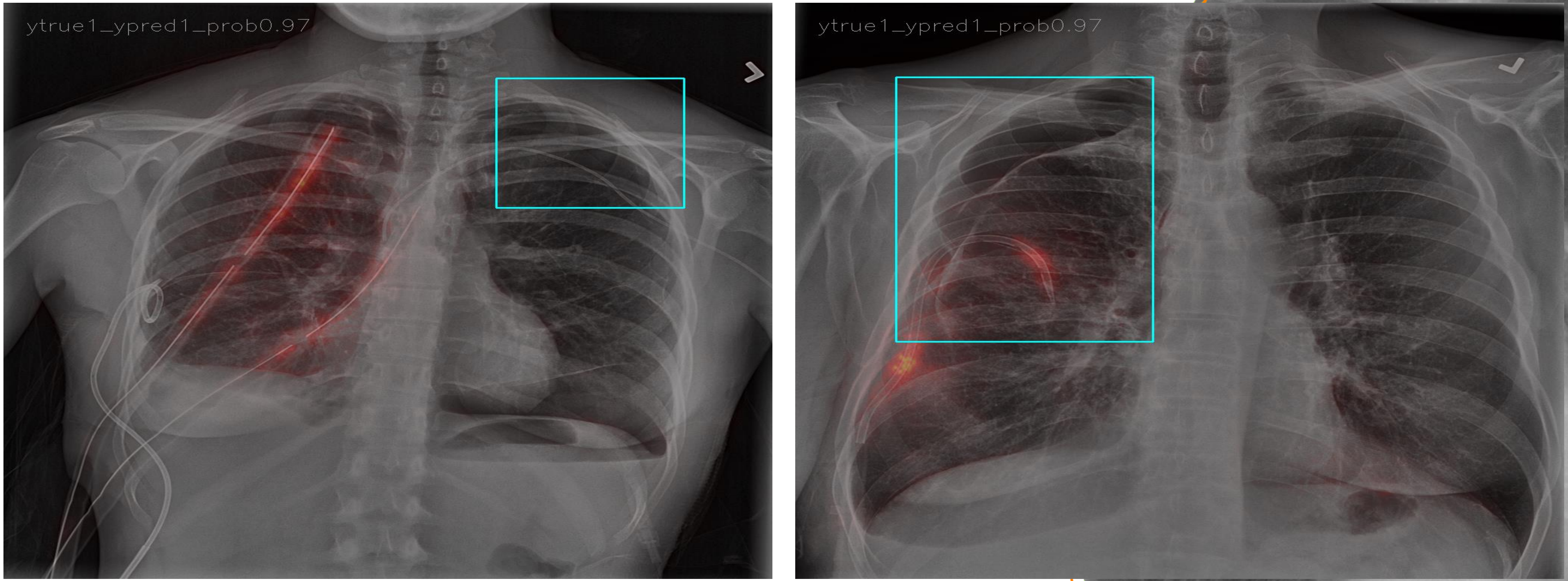




FALSE POSITIVES

DASH
Data Science Center in Health

Explainability – Deep Taylor Decomposition



TRUE POSITIVES

Dataset Bias!

- CHX 14: PTX images are biased
  - Large fraction of post-treatment images in dataset
- Causes a bias towards X-rays with drain
- Correct classification, but not suitable for diagnostics
- Impossible to tell from the metrics alone
  - Heatmap/relevance map useful tool to understand model decision making

Data Science Center in Health

Model Deployment and Quality Assurance

Machine Learning Lab

DASH Data Science Center in Health

umcg

AI Market situation

- https://grand-challenge.org/aiforradiology/

European Radiology (2021) 31:3797–3804
https://doi.org/10.1007/s00330-021-07892-z

AI for Radiology
an implementation guide

IMAGING INFORMATICS AND ARTIFICIAL INTELLIGENCE

Check for updates

Artificial intelligence in radiology: 100 commercially available products and their scientific evidence

Kicky G. van Leeuwen[1] · Steven Schalekamp[1] · Matthieu J. C. M. Rutten[1,2] · Bram van Ginneken[1] · Maarten de Rooij[1]

Machine Learning Lab

DASH Data Science Center in Health | umcg

AI Market situation

- 100 CE Marked AI products from 54 vendors
- 64/100 have no peer review evidence
- 36/100 have evidence from 237 papers
- 116/237 papers were independent and not (co-)funded or (co-)authored by the vendor
- Only 18/100 AI products have demonstrated (potential) clinical impact!!

European Radiology (2021) 31:3786–3796
https://doi.org/10.1007/s00330-020-07684-x

IMAGING INFORMATICS AND ARTIFICIAL INTELLIGENCE

To buy or not to buy—evaluating commercial AI solutions in radiology (the ECLAIR guidelines)

Patrick Omoumi[1] • Alexis Ducarouge[2] • Antoine Tournier[2] • Hugh Harvey[3] • Charles E. Kahn Jr[4] • Fanny Louvet-de Verchère[5] • Daniel Pinto Dos Santos[6] • Tobias Kober[7] • Jonas Richiardi[1]

Machine Learning Lab

DASH Data Science Center in Health | umcg

Checklist AI assessment

1. Relevance
2. Performance and Validation
3. Usability and Integration
4. Regulatory and Legal Aspects
5. Financial and Support services considerations

Machine Learning Lab

DASH Data Science Center in Health | umcg

Table 2 Top 10 questions to consider

1. What problem is the application intended to solve, and who is the application designed for?
2. What are the potential benefits and risks, and for whom?
3. Has the algorithm been rigorously and independently validated?
4. How can the application be integrated into your clinical workflow and is the solution interoperable with your existing software?
5. What are the IT infrastructure requirements?
6. Does the application conform to the medical device and the personal data protection regulations of the target country, and what class of regulation does it conform to?
7. Have return on investment (RoI) analyses been performed?
8. How is the maintenance of the product ensured?
9. How are user training and follow-up handled?
10. How will potential malfunctions or erroneous results be handled?

Machine Learning Lab

DASH Data Science Center in Health

umcg

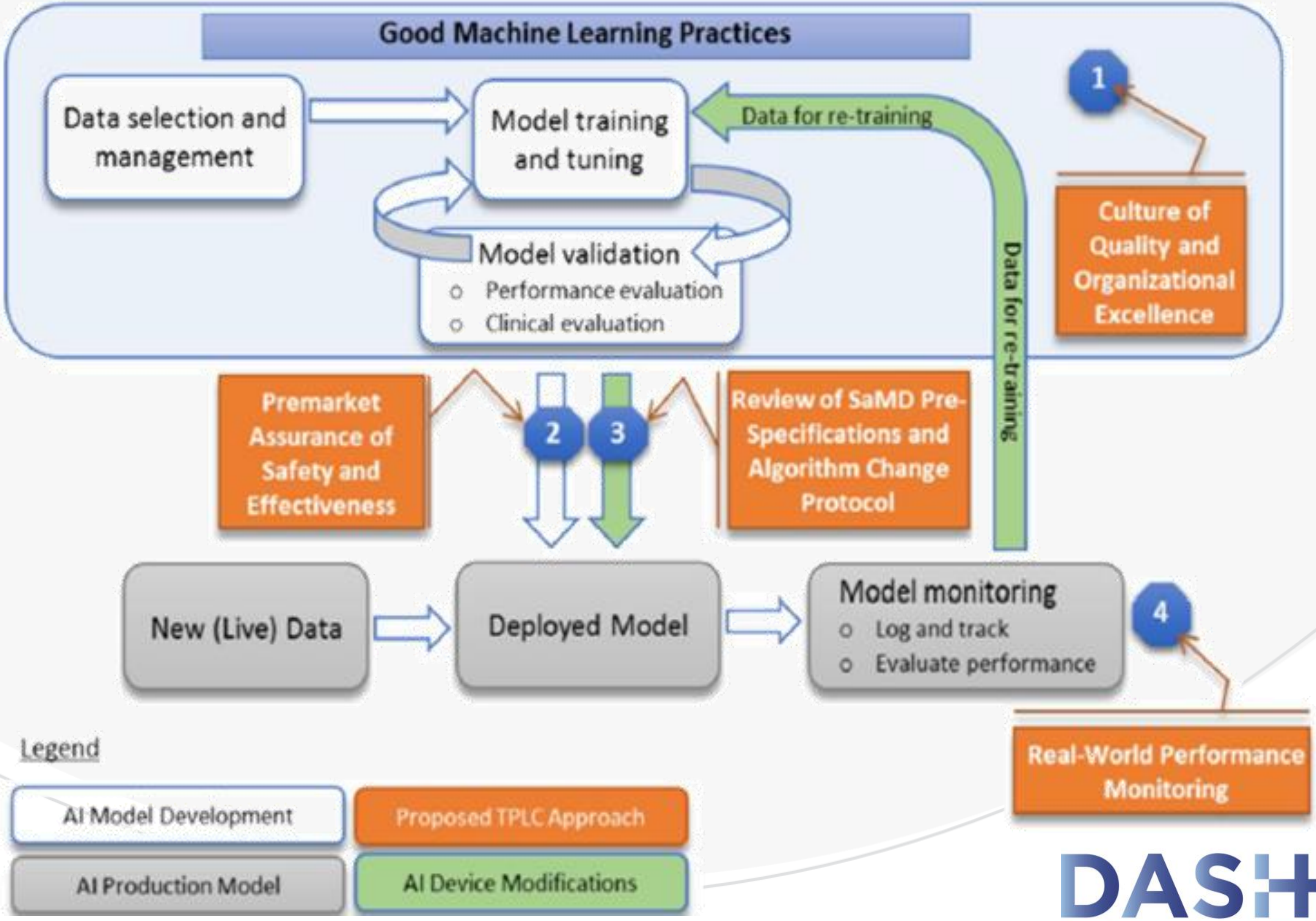
Challenges in QA

Performance of AI is influenced by:

- Data Drift
  - New scanner coming in
- Concept Drift
  - Ideas about diagnosis/conclusion changes
- Population Drift
  - Patient groups change

Machine Learning Lab

DASH Data Science Center in Health | umcg



Machine Learning Lab

DASH Data Science Center in Health | umcg

U.S. FOOD & DRUG ADMINISTRATION

ELSA Issues

Machine Learning Lab

DASH Data Science Center in Health

umcg

ELSA Issues

Ethical

- Informed Consent
- Safety and Transparency
- Algorithmic Fairness and Bias
- Data Privacy

Legal

- Safety and effectiveness
- Liability
- Data protection and privacy
- Cybersecurity
- Intellectual Property law

Machine Learning Lab

DASH Data Science Center in Health | umcg

Regulatory Concepts EU

- General Data Protection Regulation (GDPR)
- Medical Device Regulation (MDR)
  - Software as a Medical Device (SaMD)
- Ethical Guidelines for Trustworthy AI

Regulatory Concepts - GDPR

"Right for an individual to obtain a meaningful explanation when automated (algorithmic) decision-making is involved."

Machine Learning Lab

EuSoMII

DASH Data Science Center in Health | umcg

Regulatory Concepts - MDR

Machine Learning Lab

EuSoMII

DASH Data Science Center in Health | umcg

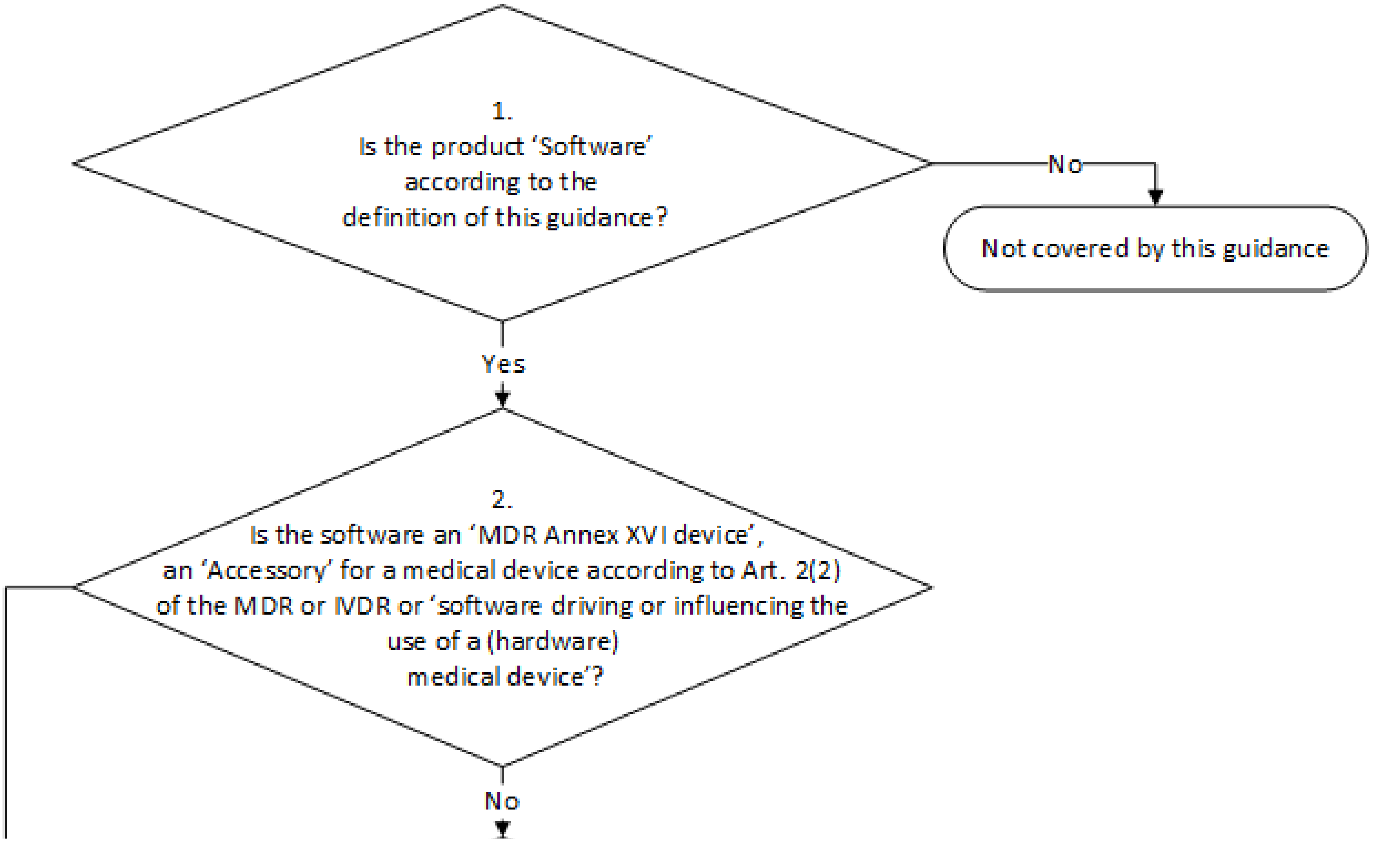
Beckers. Kwade, Zanca. Physica Medica 2021;83:1-8

Regulatory Concepts - MDR

State of Healthcare / Situation of patient condition	Significance of information provided by the MDSW to a healthcare situation related to diagnosis/therapy: **High** Treat or diagnose ~IMDRF 5.1.1	**Medium** Drives Clinical Management ~IMDRF 5.1.2	**Low** Informs Clinical Management (everything else)
Critical situation or patient condition ~IMDRF 5.2.1	Class III Category IV.i	Class IIb Category III.i	Class IIa Category II.i
Serious situation or patient condition ~IMDRF 5.2.2	Class IIb Category III.ii	Class IIa Category II.ii	Class IIa Category I.ii
Non-Serious situation or patient condition (everything else)	Class IIa Category II.iii	Class IIa Category I.iii	Class IIa Category I.i

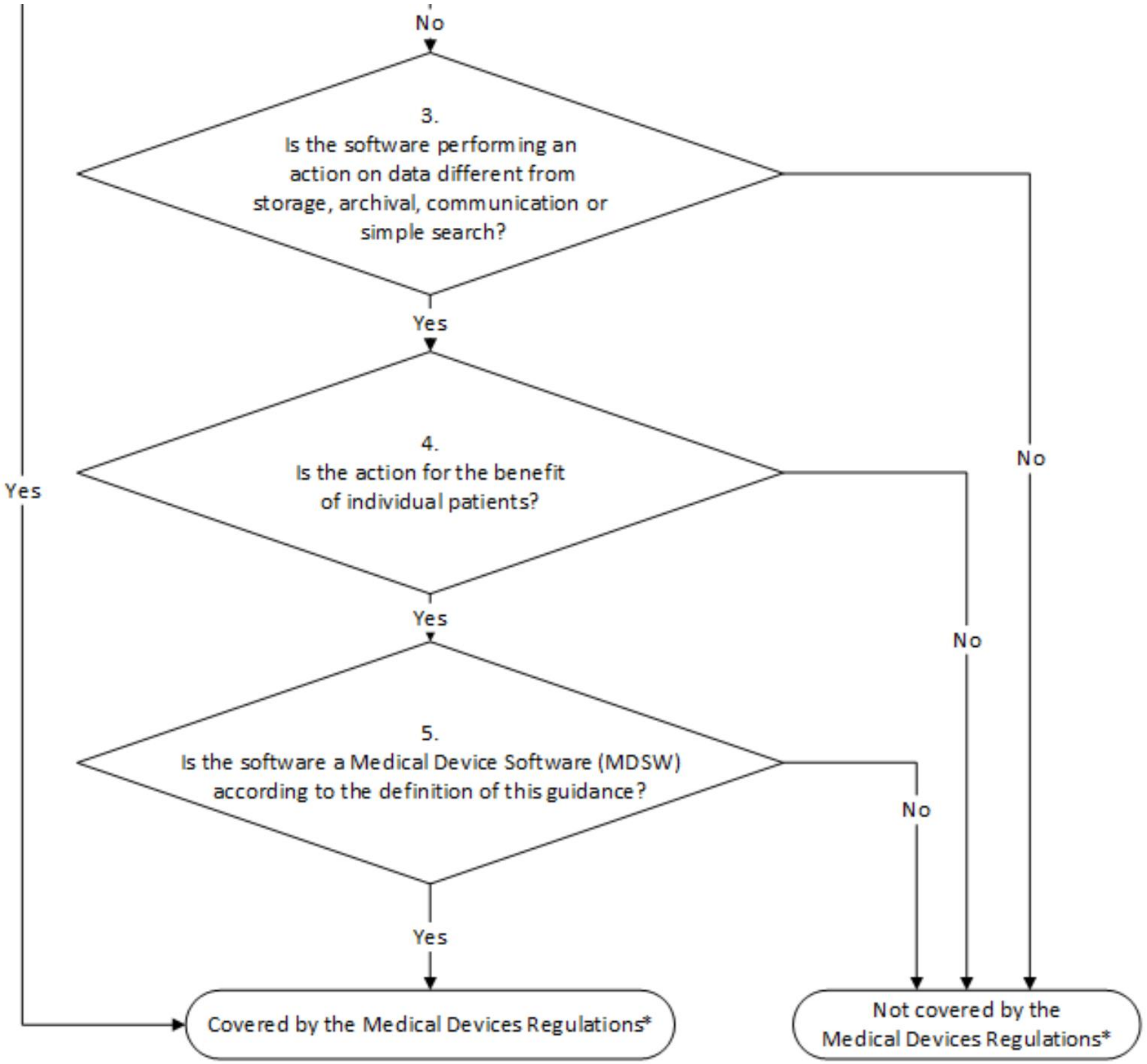
MDCG 2019-11 – Guidance on Qualification and Classification of Software in Regulation (EU) 2017/745 - MDR

Regulatory Concepts - MDR



MDCG 2019-11 – Guidance on Qualification and Classification of Software in Regulation (EU) 2017/745 - MDR

Regulatory Concepts - MDR



MDCG 2019-11 – Guidance on Qualification and Classification of Software in Regulation (EU) 2017/745 - MDR

Regulatory Concepts - MDR

- Also for in-house developed software!
- CE marking is not required when software is developed, used, and maintained only within a health institution, under certain conditions:
  - software cannot be transferred to another legal entity
  - health institution justifies that the target patient group's specific needs cannot be met or cannot be met at the appropriate level of performance by an equivalent device available on the market.
  - The health institution needs to prepare documentation containing e.g. design and performance information of the device AND to review experience gained from clinical use of the software and take necessary corrective actions.

Machine Learning Lab

EuSoMII

DASH Data Science Center in Health

umcg

Beckers. Kwade, Zanca. Physica Medica 2021;83:1-8

Regulatory Concepts - Ethical Guidelines Trustworthy AI

- Ethical Guidelines for trustworthy AI
- Independent High-Level experts group on AI
- Started by the European Commission in June 2018

https://altai.insight-centre.org/

Regulatory Concepts - Ethical Guidelines Trustworthy AI

7 Requirements:

1. Human Agency and Oversight;
2. Technical Robustness and Safety;
3. Privacy and Data Governance;
4. Transparency;
5. Diversity, Non-discrimination and Fairness;
6. Societal and Environmental Well-being;
7. Accountability.

Machine Learning Lab

EuSoMII

DASH Data Science Center in Health

umcg

AI Home How To Complete The ALTAI Fundamental Rights My ALTAIs Hello p.m.a.van.ooijen@umcg.nl! Logout

ALTAI for Test

Notes

Sections of the ALTAI

- Human Agency and Oversight
- Technical Robustness and Safety
- Privacy and Data Governance
- Transparency
- Diversity, Non-Discrimination and Fairness
- Societal and Environmental Well-being
- Accountability

Legend of progression symbols

Human Agency and Oversight

AI systems should support human autonomy and decision-making, as prescribed by the principle of respect for human autonomy. This requires that AI systems should both act as enablers to a democratic, flourishing and equitable society by supporting the user's agency and upholding fundamental rights, which should be underpinned by human oversight. In this section, we are asking you to assess the AI system in terms of the respect for human agency, as well as human oversight.

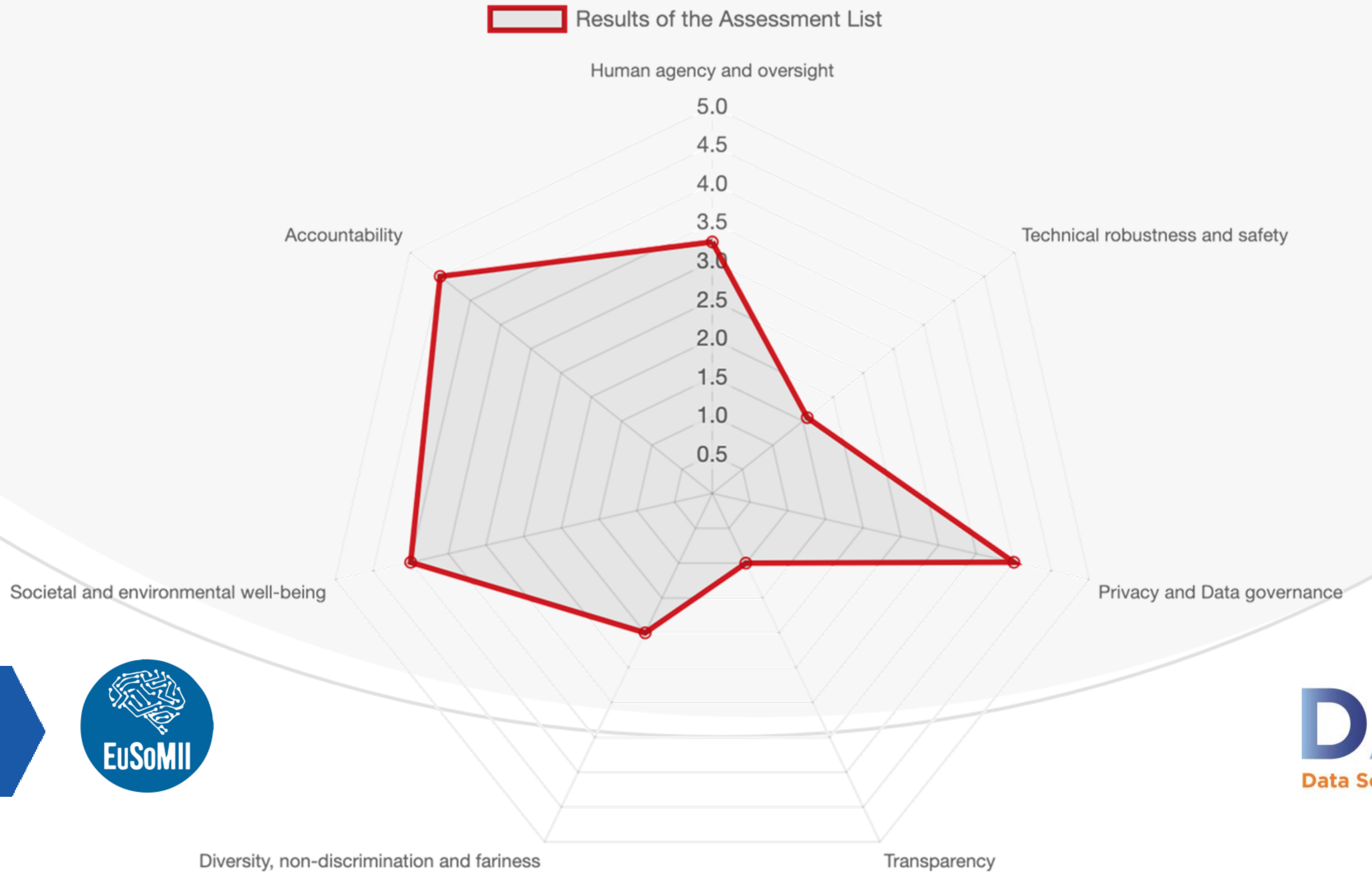
Human Autonomy

https://altai.insight-centre.org/

This subsection deals with the effect AI systems can have on human behaviour in the broadest sense. It deals with the effect of AI systems that are aimed at guiding, influencing or supporting humans in decision making processes, for example, algorithmic decision support systems, risk analysis/prediction systems (recommender systems, predictive policing, financial risk analysis, etc.). It also deals with the effect on human perception and expectation when confronted with AI systems that 'act' like humans. Finally, it deals with the effect of AI systems on human affection, trust and (in)dependence.

Is the AI system designed to interact, guide or take decisions by human end-users that affect humans ('subjects') or society? ? *

Self assessment results and recommendations



Self assessment results and recommendations

Recommendations

Human agency and oversight

Put in place any procedure to avoid that the system inadvertently affects human autonomy.

Deploy a “stop button” or procedure to safely abort an operation when needed.

Technical robustness and safety

No recommendation for this requirement.

Privacy and Data Governance

Consider the privacy and data protection implications of data collected, generated or processed over the course of the AI system's lifecycle.

Whenever possible and relevant, align the AI-system with relevant standards (e.g. ISO, IEEE) or widely adopted protocols for (daily) data management and governance.

Machine Learning Lab

EuSoMII

DASH Data Science Center in Health | umcg

Hulpmiddel Handelingsruimte Waardevolle AI voor gezondheid en zorg

- Om onderzoekers en ontwikkelaars in het traject van ontwikkeling tot opschaling van waardevolle artificiële intelligentie (AI) te helpen, biedt dit hulpmiddel aanwijzingen in de handelingsruimte binnen de wet- en regelgeving. Zo kan er vroegtijdig gestart worden met het voorbereiden op gevraagde minimale eisen of standaarden. En reflecteren op acties om tot mensgerichte en betrouwbare AI-toepassingen te komen.

General Lesson

The future in medical imaging is in
Data Science Developments

Machine
Learning Lab

DASH
Data Science Center in Health

umcg

General Lesson

The future in medical imaging is in Data Science Developments

When applied carefully with the aim to support and facilitate healthcare practice and continuously evaluated and adapted to current possibilities and needs. All with a strong emphasis on technical, clinical, ethical and legal validation.

Machine Learning Lab

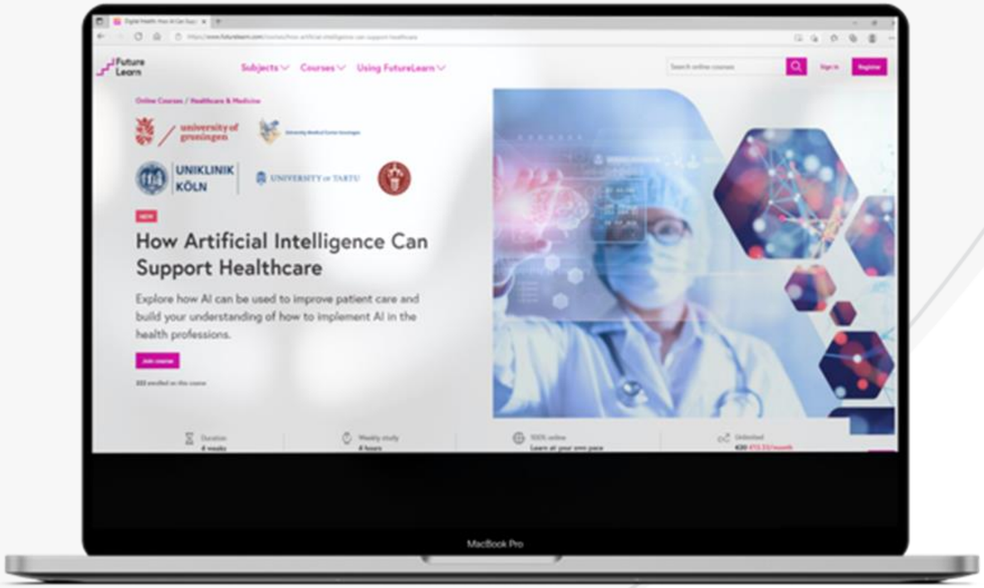
DASH Data Science Center in Health | umcg

Want to learn more about the use of artificial intelligence (AI) in healthcare?

Follow our free online course

'How Artificial Intelligence can support Healthcare'

Discover how AI can be used to improve patient care and gain a deeper understanding of AI implementation in health professions.



https://www.futurelearn.com/courses/how-artificial-intelligence-can-support-healthcare

DASH Data Science Center in Health

umcg

Contact us

linkedin.com/company/dash-umcg

umcgresearch.org/w/dash

dash@umcg.nl

twitter.com/DASH_umcg

DASH Data Science Center in Health

umcg