Physikalisch-Technische Bundesanstalt
Braunschweig and Berlin
National Metrology Institute

Quantitative Measurements in Medicine
-
Accurate data for better Health

Tobias Schaeffter

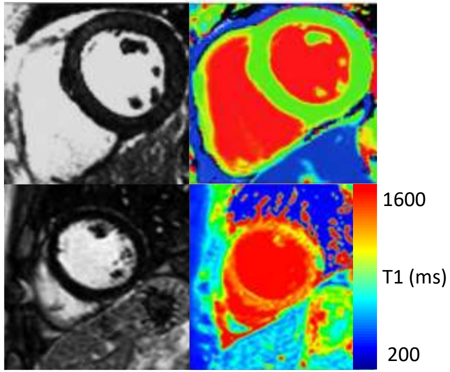
6 June 2024

EINSTEIN CENTER
Digital Future

Modern Healthcare

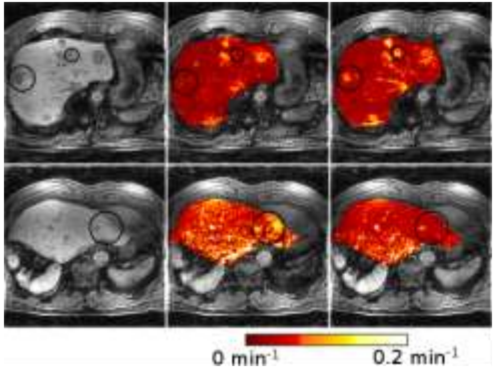
Driving Force: More Chronic Disease - High Costs

Heart Disease



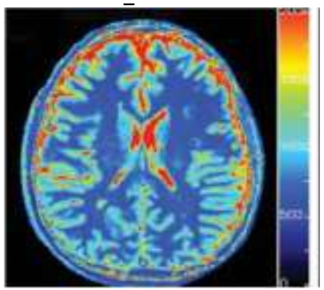
Every 33 seconds a person dies due to CVD[1]

Cancer



Every 3rd person will have cancer[2]

Neurodegenerative



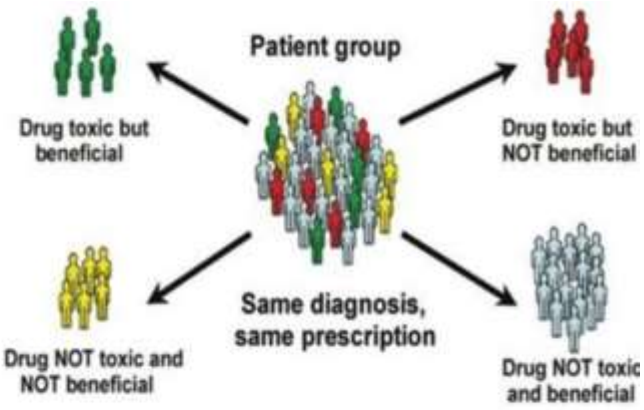
Every 8th person over 65 has Alzheimer[3]

[1]AHA, [2]CRUK, [3] Alzheimer Association

Modern Healthcare

Solution: Stratified Medicine

„Tailoring the right therapeutic strategy for the right person at the right time“



Aim:

- Increase response rate
- Reduce side-effects/toxity

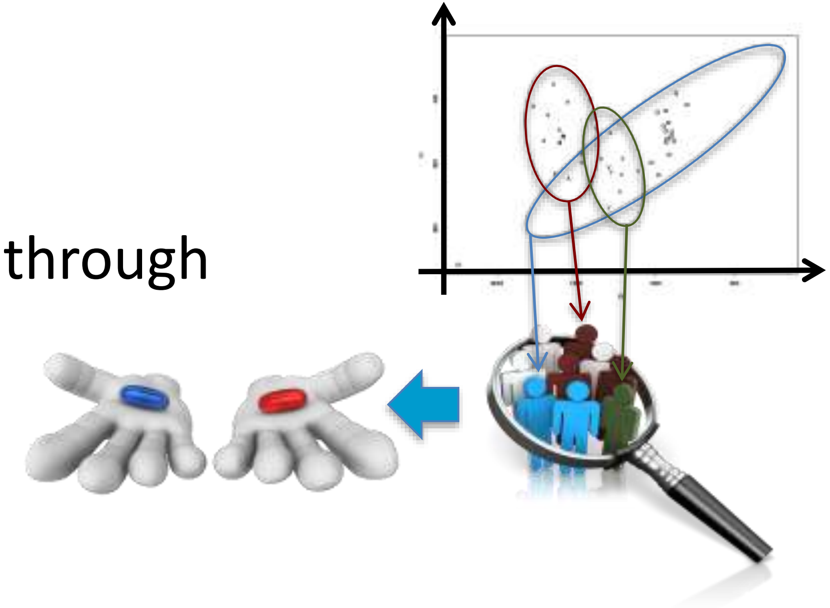
Priority Medicines for Europe and the World - Update Report, WHO 2013

Data Challenge

Stratified Medicine – Data Challenge

Stratified Medicine:

- Classification of pathologies
- Selection of appropriate therapy through classification of patient groups

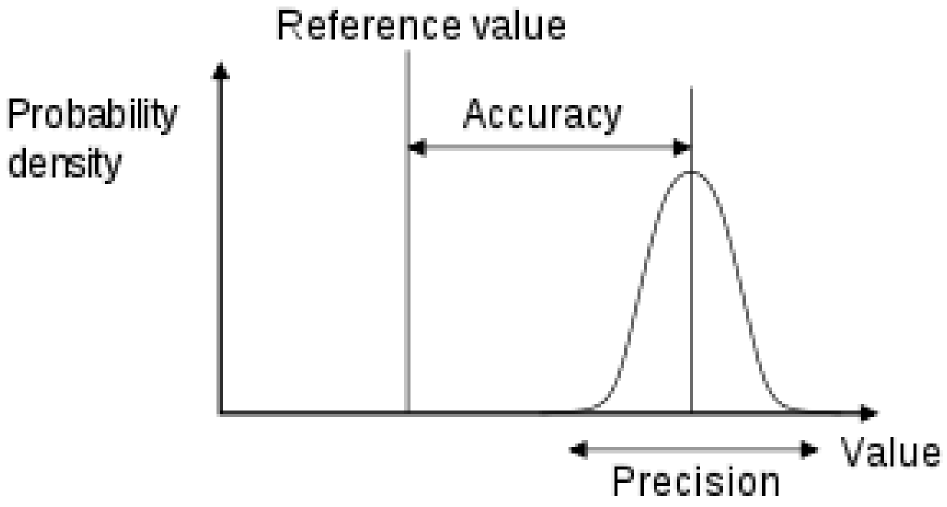


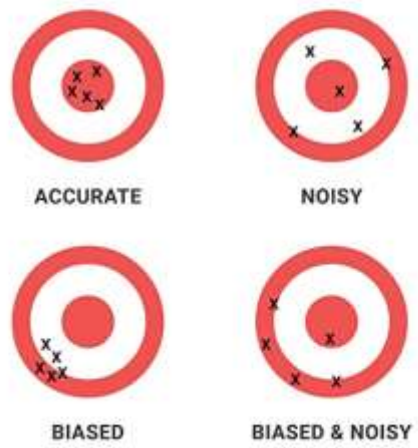
Data Challenge:

Proper patient-group selection requires "**accurate measurements**" for modern analysis

Metrology

- **Accuracy *(Bias)*** - closeness of measurement results to a reference;
- **Precision (Variability)** – closeness of measurements to each other



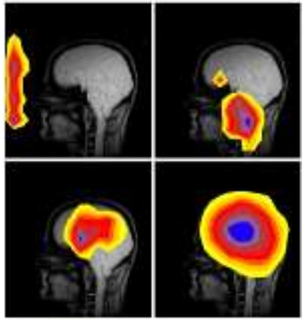


https://en.wikipedia.org/wiki/Accuracy_and_precision

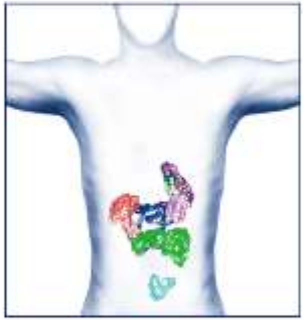
Division Medical Physics and Metrological IT

The aim of the division is the development of **new quantitative measurement techniques** and provision of **reference methods** for **precision medicine**.

The division provides mathematically sound approaches for **data analysis** and ensures **IT security** in legal metrology.

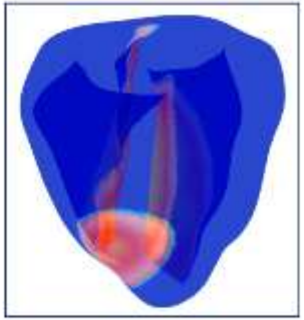


MR-Tomography



Biosignals

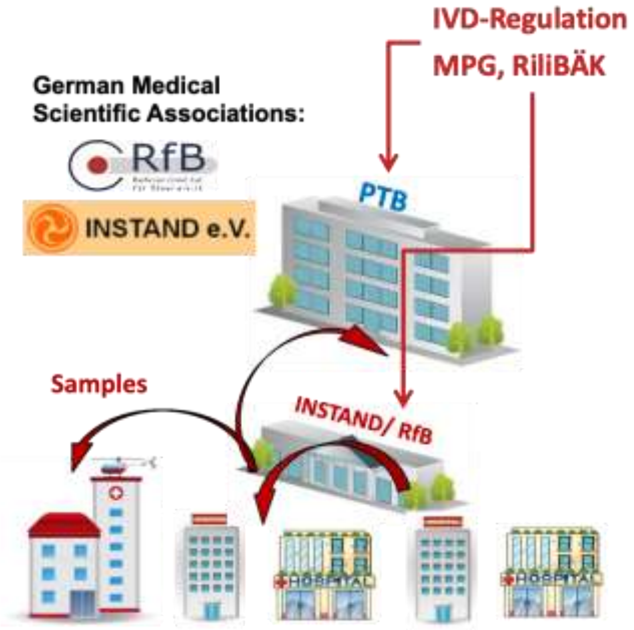
Biomedical Optics

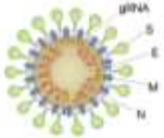


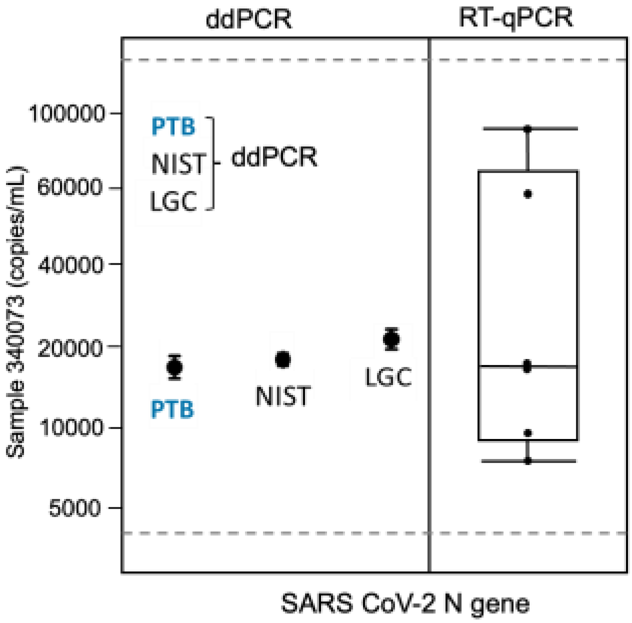
Modelling & Data Analysis

Metrological IT

Stratified medicine is based on precise data



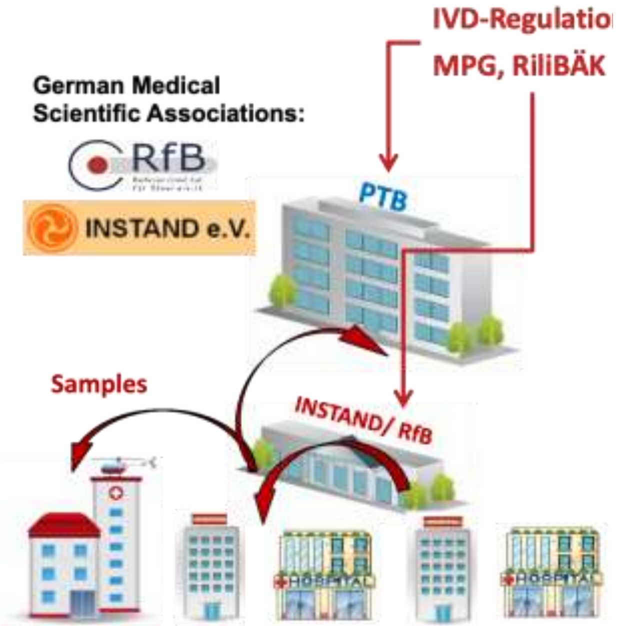


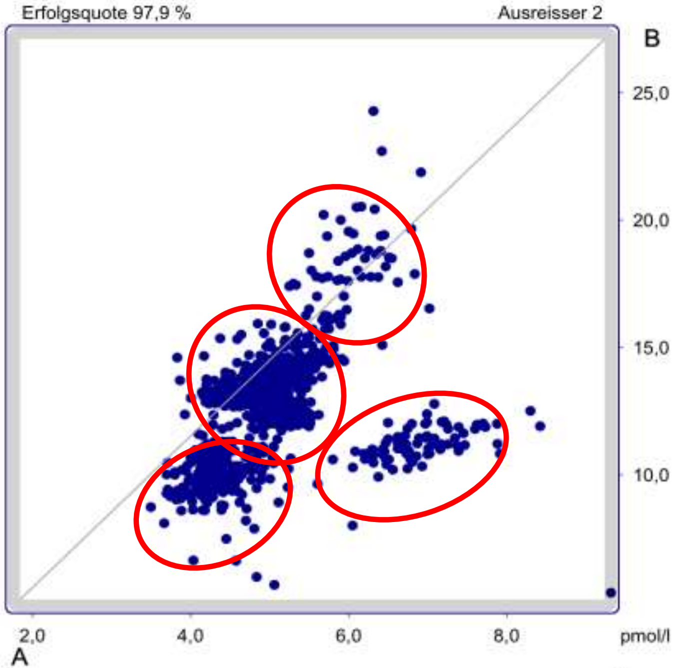


June 2020 Quantitative, copies/mL

INSTAND-RINGVERSUCHE
BEGLEITHEFT
INSTAND
Extra-INSTAND-Ringversuch -
Virusgenom-Nachweis (340)
Coronavirus SARS-CoV-2

Stratified medicine is based on precise and accurate data

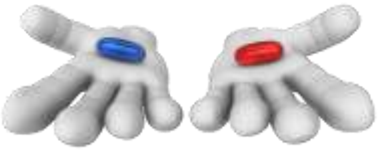




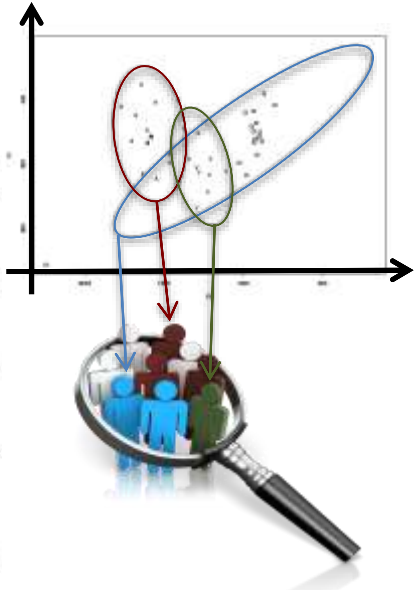
Teilnehmerzahl		**1113**	
Probe/Einheit	A	pmol/l	B
Mittelwert	5.08		12.8
Standardabweichung	0.732		2.15
Variationskoeffizient	14.4		16.8

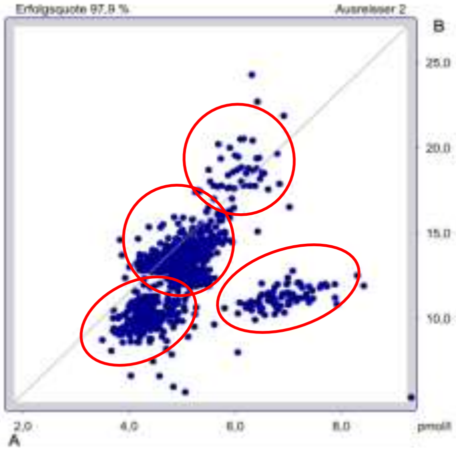
Stratified Medicine – Data Challenge

Stratified Medicine:



- Selection of appropriate therapy through classification of patient groups
- Difference due to pathology and **NOT due to different measurement techniques**



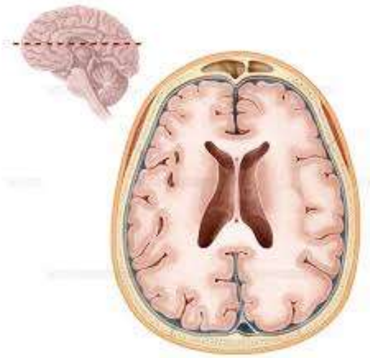


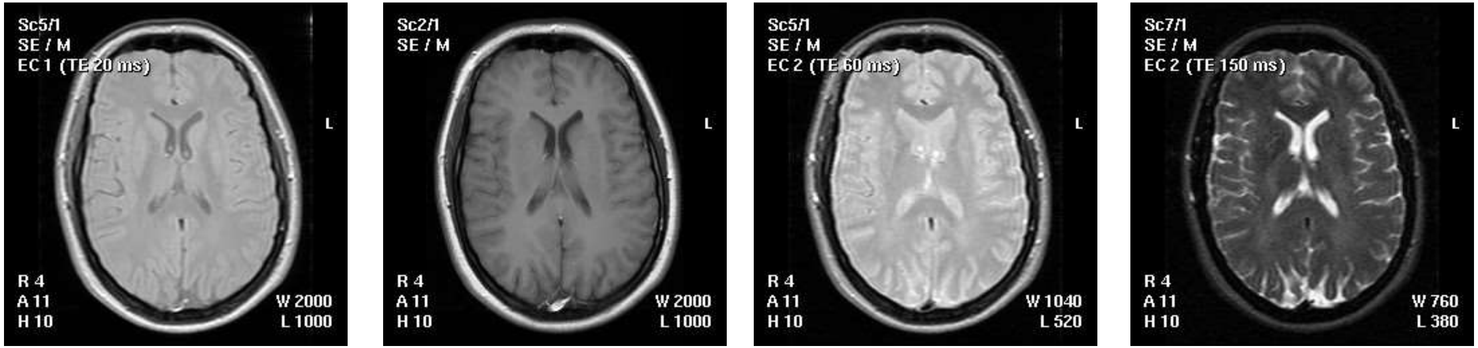
Patient-group selection requires **accurate measurements**

Accurate Measurements

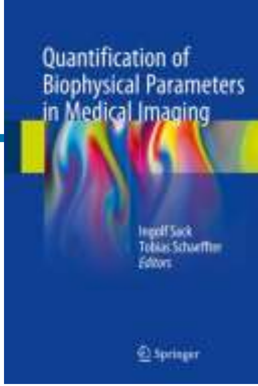
-

Imaging

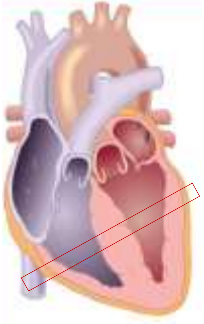




Quantitative MRI - Motivation



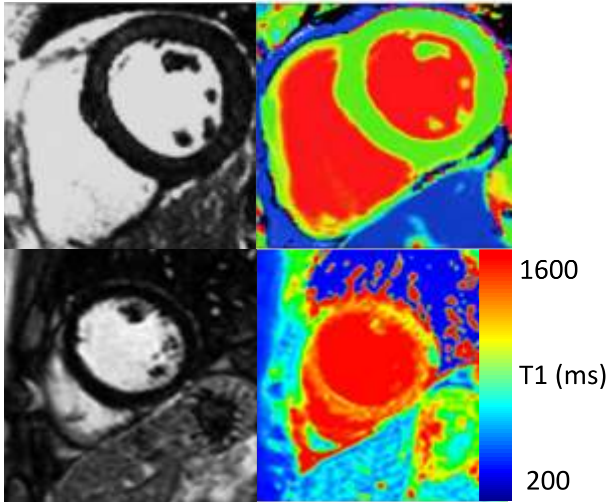
- Data consistency and comparability
- Detection of diffuse disease
- Contrast agent quantification
- ...



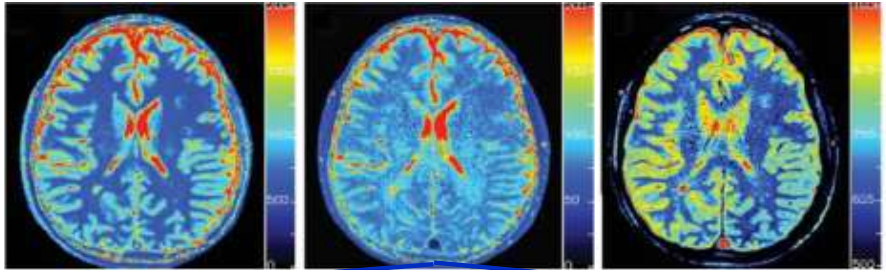
Qualitative MRI Quantitative MRI

Healthy

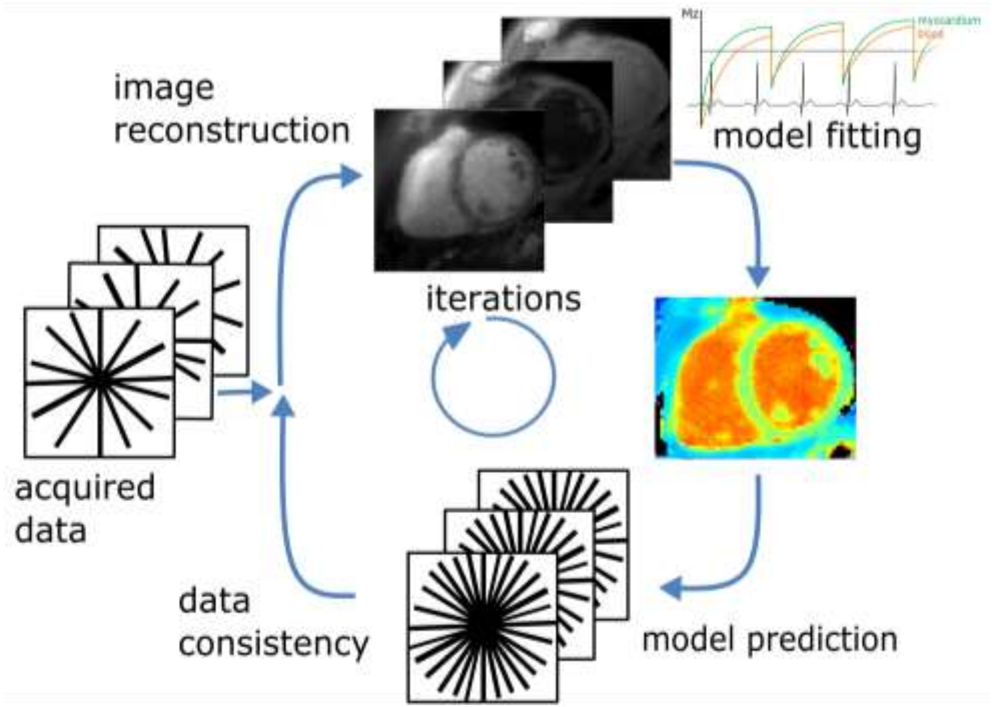
Diffuse Fibrosis

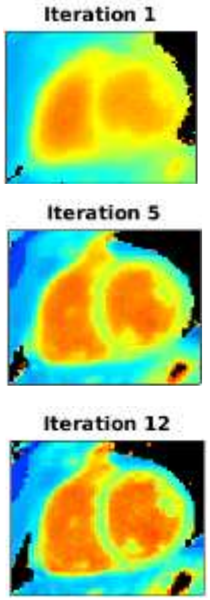


T_1 map T_2 map PD map



Model-based reconstruction for QMRI



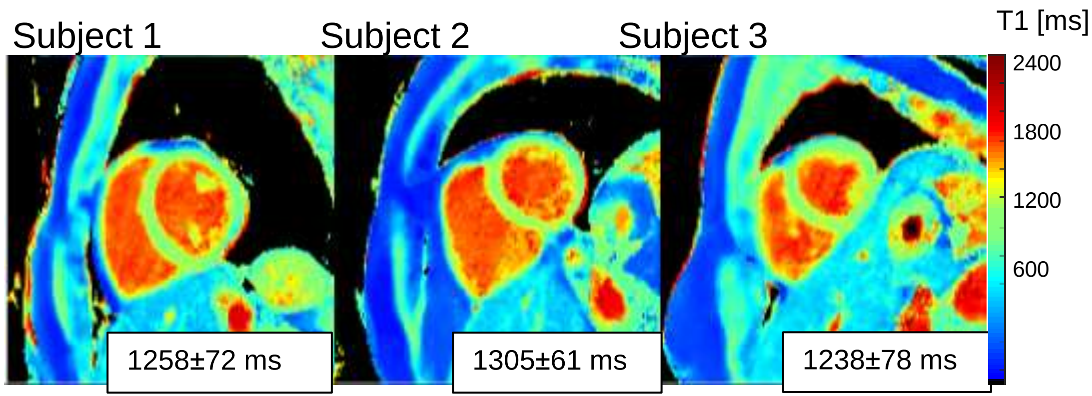


Becker K et al. MRM 2018

Results: T1 maps and cine-MRI

T1 maps diastole

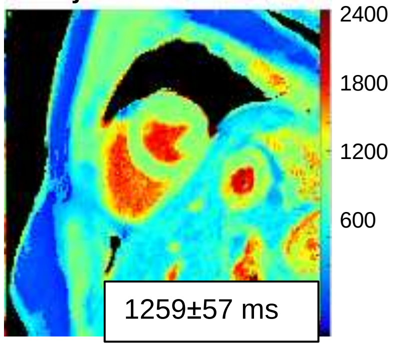
Subject 1 Subject 2 Subject 3



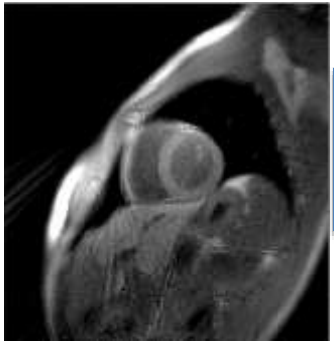
Resolution: 1.3 x 1.3 x 8 mm³

T1 map systole

Subject 3



Cine images

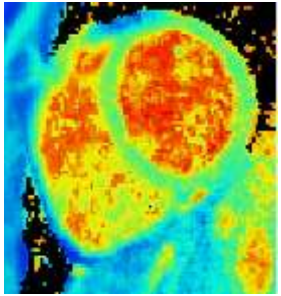


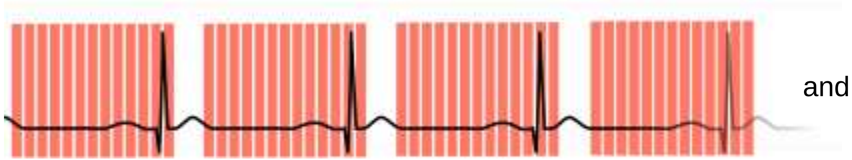
16 sec Breathold

Becker K et al. MRM 2018

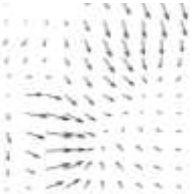
Motion Compensated T1 Mapping

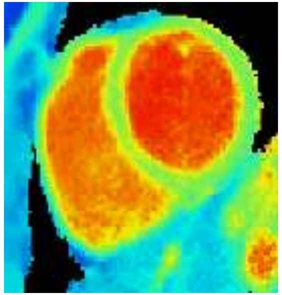
Acquisition → Motion estimation → Motion correction → T_1 Mapping

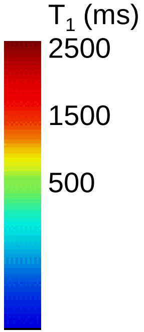




and



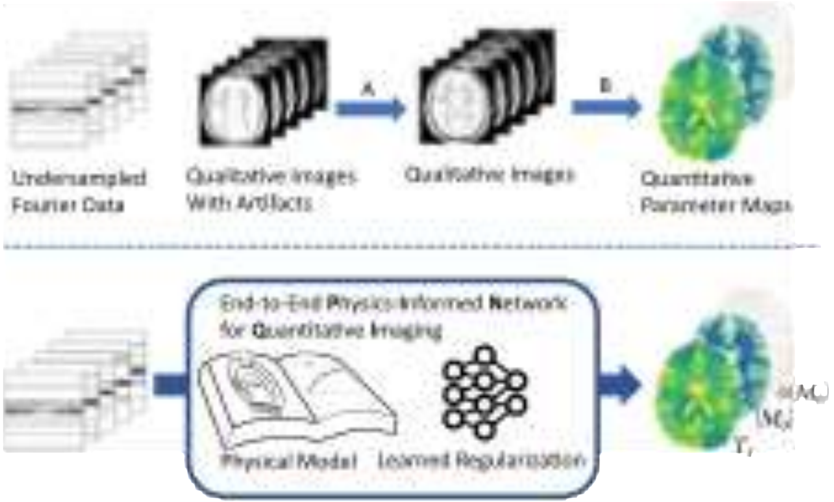




Becker K et al. MRM 2019

→ **Higher precision**

Physics-Informed Machine Learning for Quantitative MRI



$$p^* = \arg\min_{p} \|\mathbf{A}q(p) - k\|_2^2 + \lambda_{\tilde{p}} \|p - p_{\text{reg}}\|_2^2 ,$$

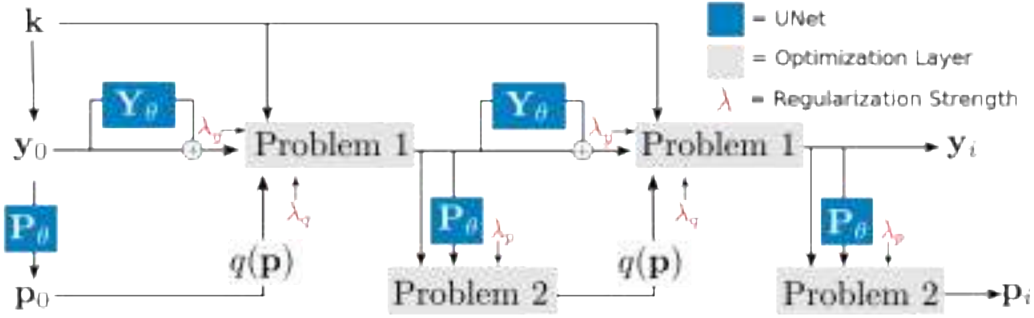
Subproblem 1:

$$y^* = \arg\min_{y} \|\mathbf{A}y - k\|_2^2 + \lambda_y \|y - y_{\text{reg}}\|_2^2 + \lambda_q \|q(p) - y\|_2^2$$

Subproblem 2:

$$p^* = \arg\min_{p} \|q(p) - y\|_2^2 + \lambda_p \|p - p_{\text{reg}}\|_2^2 .$$

Zimmermann et al.
IEEE Trans. on Comput. Imaging 2024



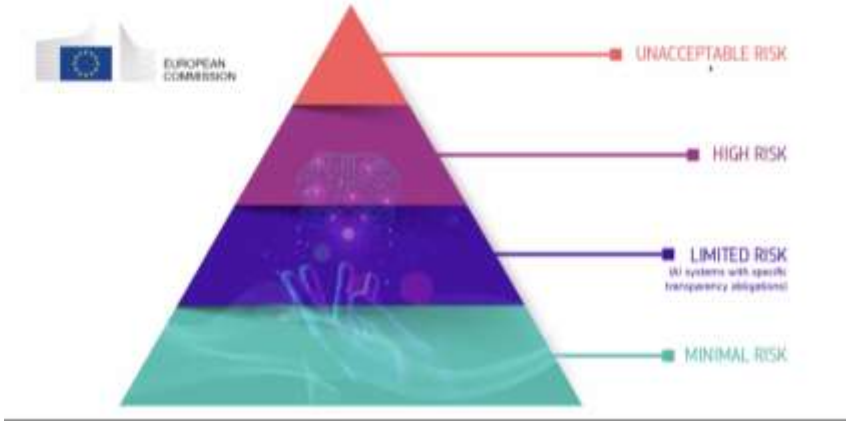
AI-Challenge

EU-AI Act for Trustworthy AI

Trust in "algorithms" that are not fully understood ("black box"), especially for high risk applications

Certification of AI requires:

- explainability,
- robustness,
- accuracy,
- security



Accuracy Robustness, Security

EU Artificial Intelligence Act

Article 15: Accuracy, Robustness and Cybersecurity

1. High-risk AI systems shall be designed and developed in such a way that they achieve an appropriate level of accuracy, robustness, and cybersecurity, and perform consistently in those respects throughout their lifecycle.
   1a. To address the technical aspects of how to measure the appropriate levels of accuracy and robustness set out in paragraph 1 of this Article and any other relevant performance metrics, the Commission shall, in **cooperation with relevant stakeholder and organisations such as metrology and benchmarking authorities**, encourage as appropriate, **the development of benchmarks and measurement methodologies.**

2. The levels of accuracy and the relevant **accuracy metrics** of high-risk AI systems shall be declared in the accompanying instructions of use.

3. High-risk AI systems shall be as resilient as possible regarding **errors, faults or inconsistencies** that may occur within the system ......

Data is the base of AI

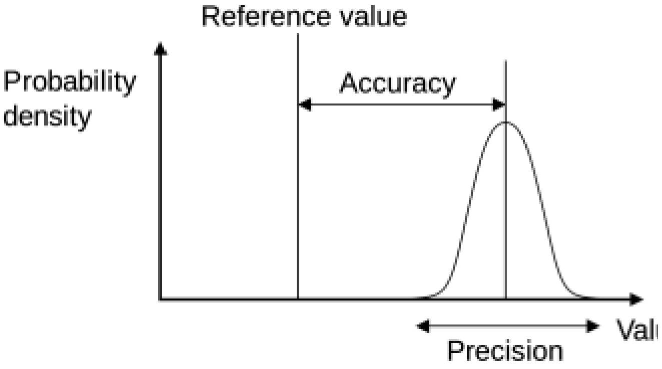
The quality of AI strongly depends on

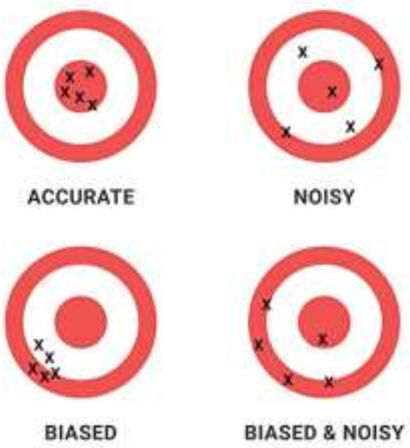
- Data uncertainty ("noise", "bias")
- Annotation inconsistencies ("label noise")

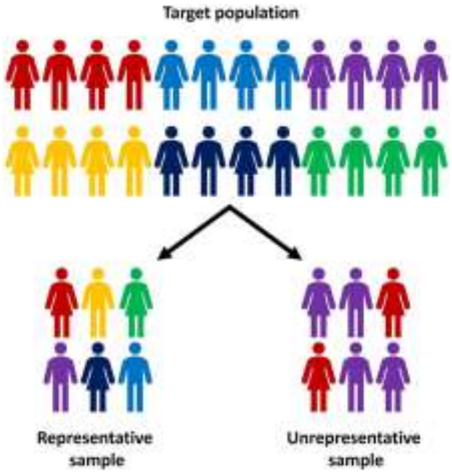
Standards for Data Quality

Uncertainty and Representativeness

- **Precision (Variability)** – closeness of measurements to each other
- **Accuracy *(Bias)*** - closeness of measurement results to a reference;
- **Representativeness** - accurate conclusions about a population from sample



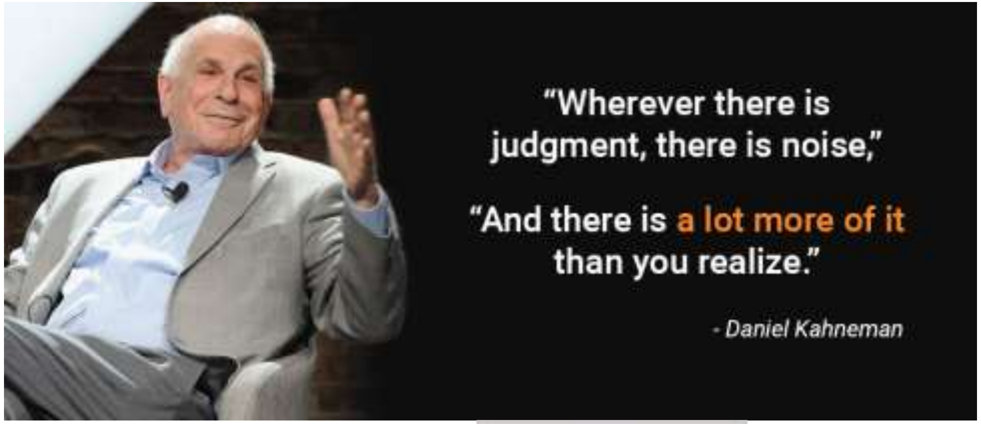


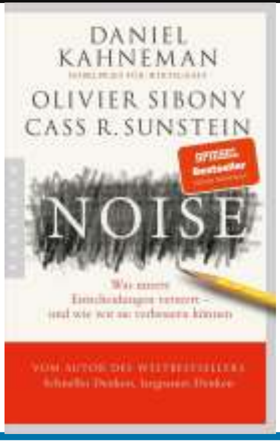


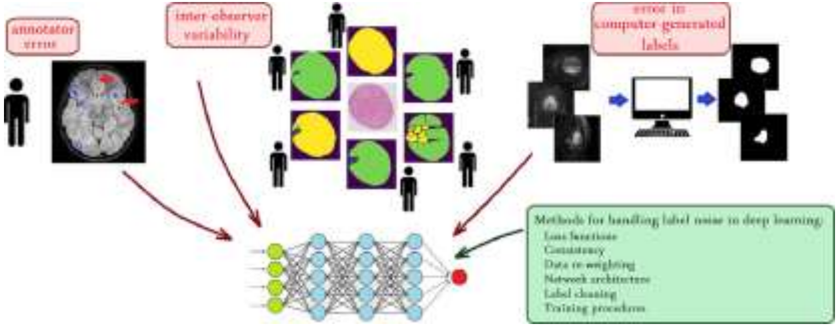
https://en.wikipedia.org/wiki/Accuracy_and_precision

Koçak B. DOI:10.5152/dir.2022.211297

Label Uncertainty

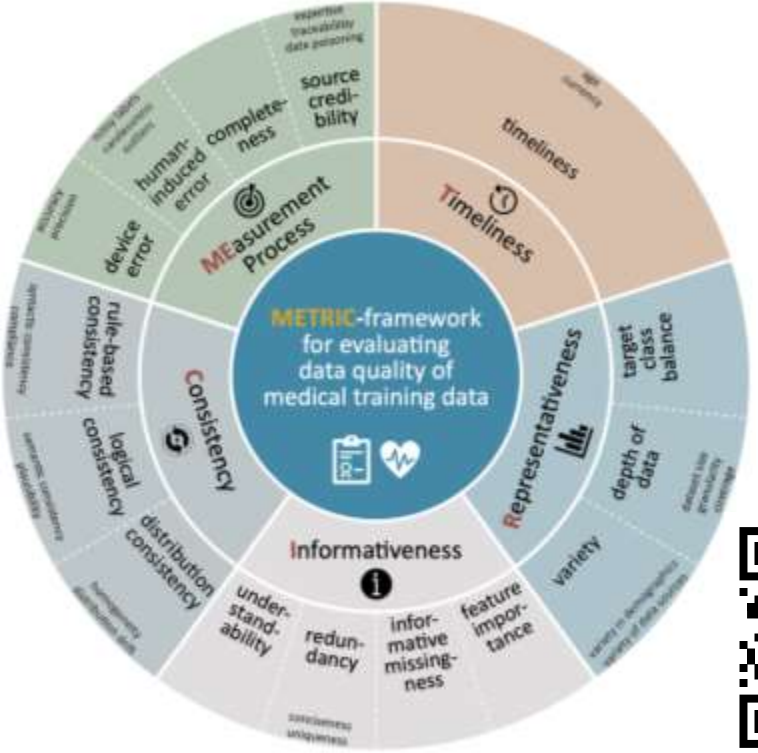




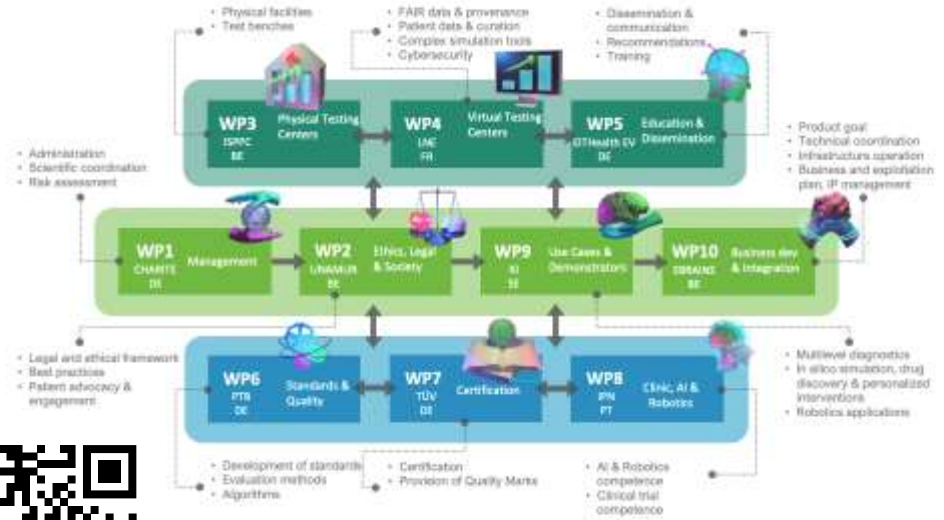


Karimi D et al. Deep learning with noisy labels: Exploring techniques and remedies in medical image analysis. Med Image Anal. 2020

Data Quality Dimensions – METRIC Framework



EU Testing und Experimentation Facilities (TEF) - Health

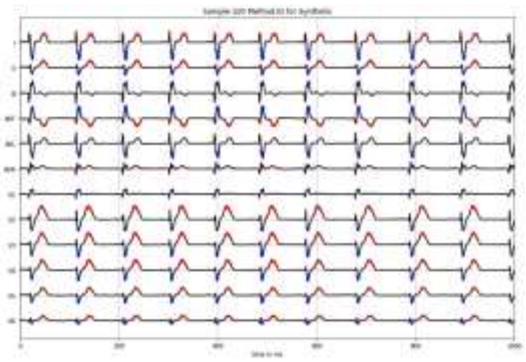


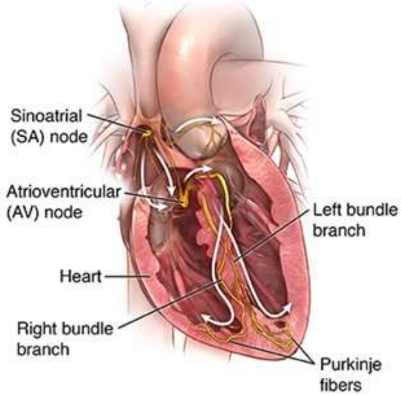
WP6 & 7: Agile Certification
(PTB, Fraunhofer, TÜV, LNE, KTH, Charité)

Use Cases

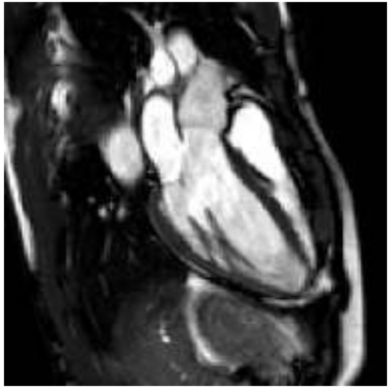
-

Vital Signs





Imaging



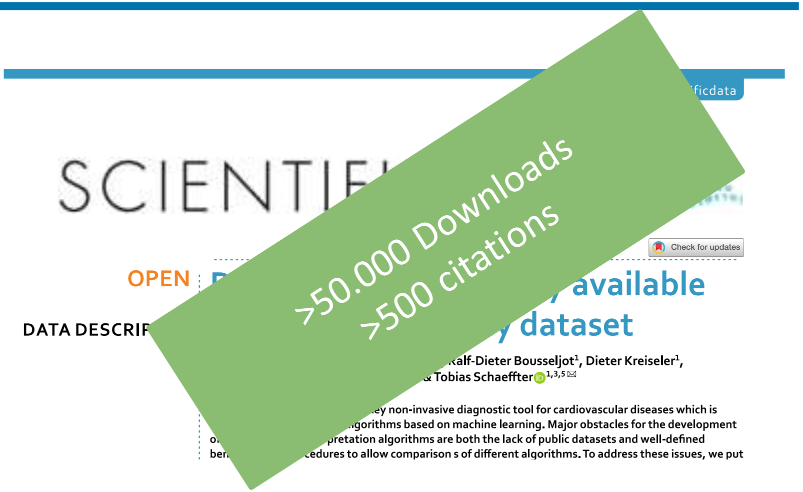
ECG-Reference-Dataset: PTB-XL

Application:

- Over 300 Mill. ECGs per year
- Strong application of AI for automatic analysis of ECG (arrhythmia, infarkt, hypertrophy,..)

Reference-Data

- EKG-Quality
- Defined Training-, Validation and Testdata
- Distribution within pathologies „taking representativeness into account“



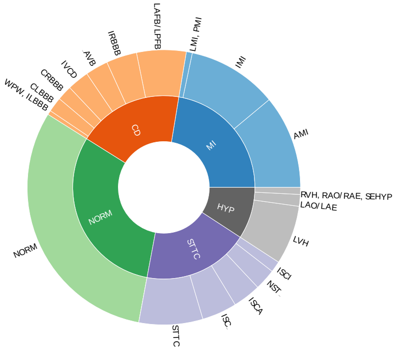


Figure 1: Graphical summary of the *PTB-XL* dataset in terms of diagnostic superclasses and subclasses, see Table 5 for a definition of the used acronyms.

Benchmarktests and Metrics

IEEE JOURNAL OF BIOMEDICAL AND HEALTH INFORMATICS, VOL. 25, NO. 5, MAY 2021 1519

Deep Learning for ECG Analysis: Benchmarks and Insights from PTB-XL

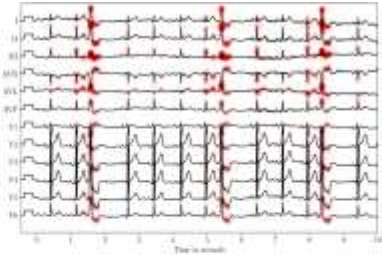
Nils Strodthoff, Patrick Wagner, Tobias Schaeffter, and Wojciech Samek, *Member, IEEE*

Metrics

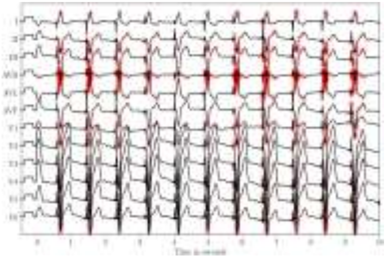
Method	all		diag.		sub-diag.		super-diag.		form		rhythm	
	AUC	Fmax	AUC	Fmax	AUC	Fmax	AUC	Fmax	AUC	Fmax	AUC	Fmax
lstm_bidir	.902(11)	.749(10)	.922(12)	.729(14)	.928(09)	.756(12)	.929(06)	.817(12)	.845(17)	.605(22)	.947(10)	.908(09)
lstm	.893(12)	.745(08)	.905(12)	.724(13)	.912(16)	.753(10)	.928(06)	.819(11)	.813(17)	.596(25)	.948(09)	.907(10)
fcn_wang	.911(10)	.754(08)	.922(10)	.731(14)	.920(14)	.752(11)	.927(07)	.815(12)	.875(18)	.625(23)	.928(10)	.899(11)
resnet1d_wang	.912(11)	.764(08)	.932(08)	.741(15)	**.932(09)**	.760(12)	**.932(06)**	**.825(12)**	.877(14)	.620(23)	.945(09)	.908(09)
xresnet1d101	**.920(08)**	**.765(08)**	**.935(08)**	**.743(13)**	.927(09)	.759(10)	.931(06)	.819(11)	**.885(13)**	**.629(20)**	**.957(20)**	.915(08)
Wavelet+NN	.811(14)	.678(10)	.823(19)	.627(15)	.845(17)	.654(14)	.870(10)	.731(13)	.798(21)	.526(22)	.857(52)	.866(13)
inception1d	.919(08)	**.765(07)**	.929(13)	.737(12)	**.932(08)**	**.763(10)**	.930(06)	.819(11)	**.885(14)**	.627(20)	**.957(14)**	**.917(09)**
ensemble	**.923(09)**	**.767(08)**	**.935(07)**	.740(12)	.928(11)	**.764(11)**	**.937(06)**	**.827(12)**	**.891(12)**	**.638(23)**	**.970(08)**	.916(08)
naive	*.500(00)*	*.557(11)*	*.500(00)*	*.440(18)*	*.500(00)*	*.440(18)*	*.500(00)*	*.448(09)*	*.500(00)*	*.365(19)*	*.500(00)*	*.797(13)*

Strodthoff et al. IEEE Journal of Biomedical and Health Informatics 2020.

Explainability

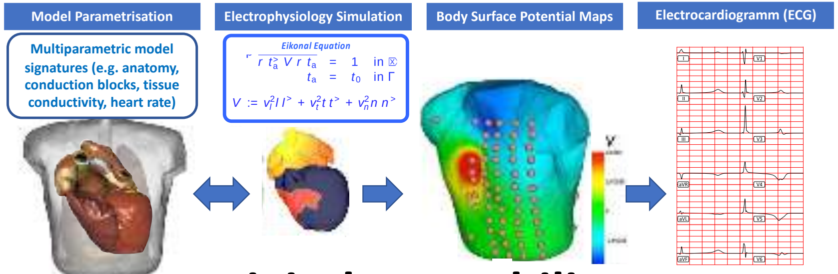


(a) PVC

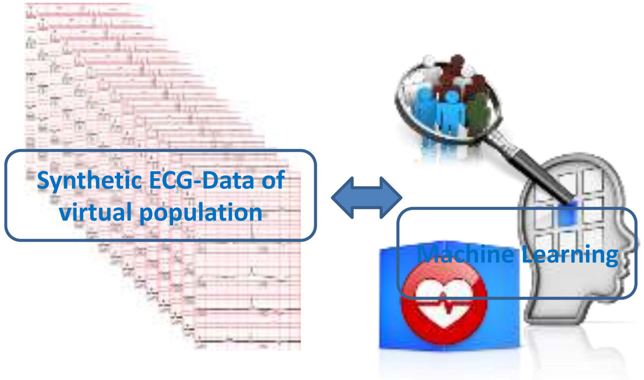


(b) PACE

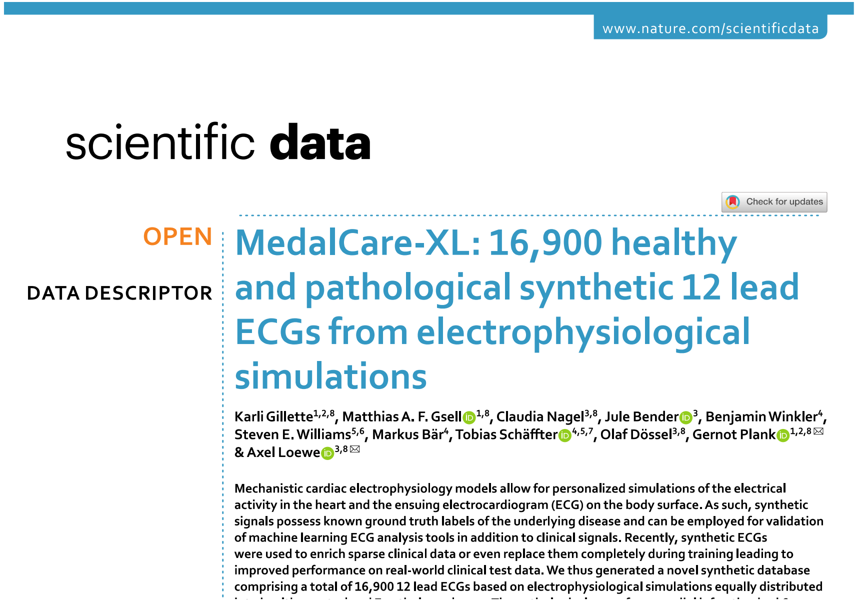
EU-Project: MedalCare Synthetic Reference Data



Digital Traceability



Uncertainty of ML


www.nature.com/scientificdata

scientific **data**

Check for updates

OPEN

DATA DESCRIPTOR

MedalCare-XL: 16,900 healthy and pathological synthetic 12 lead ECGs from electrophysiological simulations

Karli Gillette[1,2,8], Matthias A. F. Gsell[1,8], Claudia Nagel[3,8], Jule Bender[3], Benjamin Winkler[4], Steven E. Williams[5,6], Markus Bär[4], Tobias Schäffter[4,5,7], Olaf Dössel[3,8], Gernot Plank[1,2,8] & Axel Loewe[3,8]

Mechanistic cardiac electrophysiology models allow for personalized simulations of the electrical activity in the heart and the ensuing electrocardiogram (ECG) on the body surface. As such, synthetic signals possess known ground truth labels of the underlying disease and can be employed for validation of machine learning ECG analysis tools in addition to clinical signals. Recently, synthetic ECGs were used to enrich sparse clinical data or even replace them completely during training leading to improved performance on real-world clinical test data. We thus generated a novel synthetic database comprising a total of 16,900 12 lead ECGs based on electrophysiological simulations equally distributed

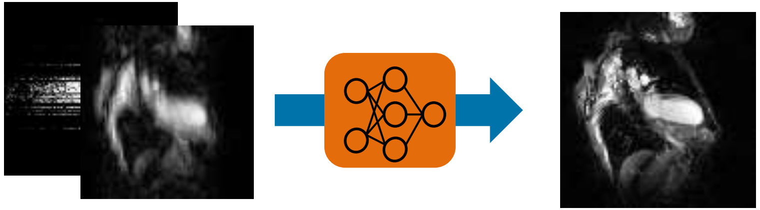
Fraunhofer HHI

PTB

KIT

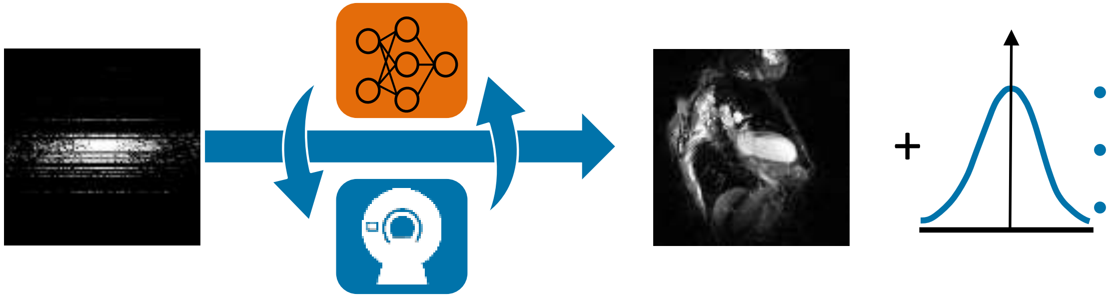
AI for Image Reconstruction

Deep Learning Reconstruktion



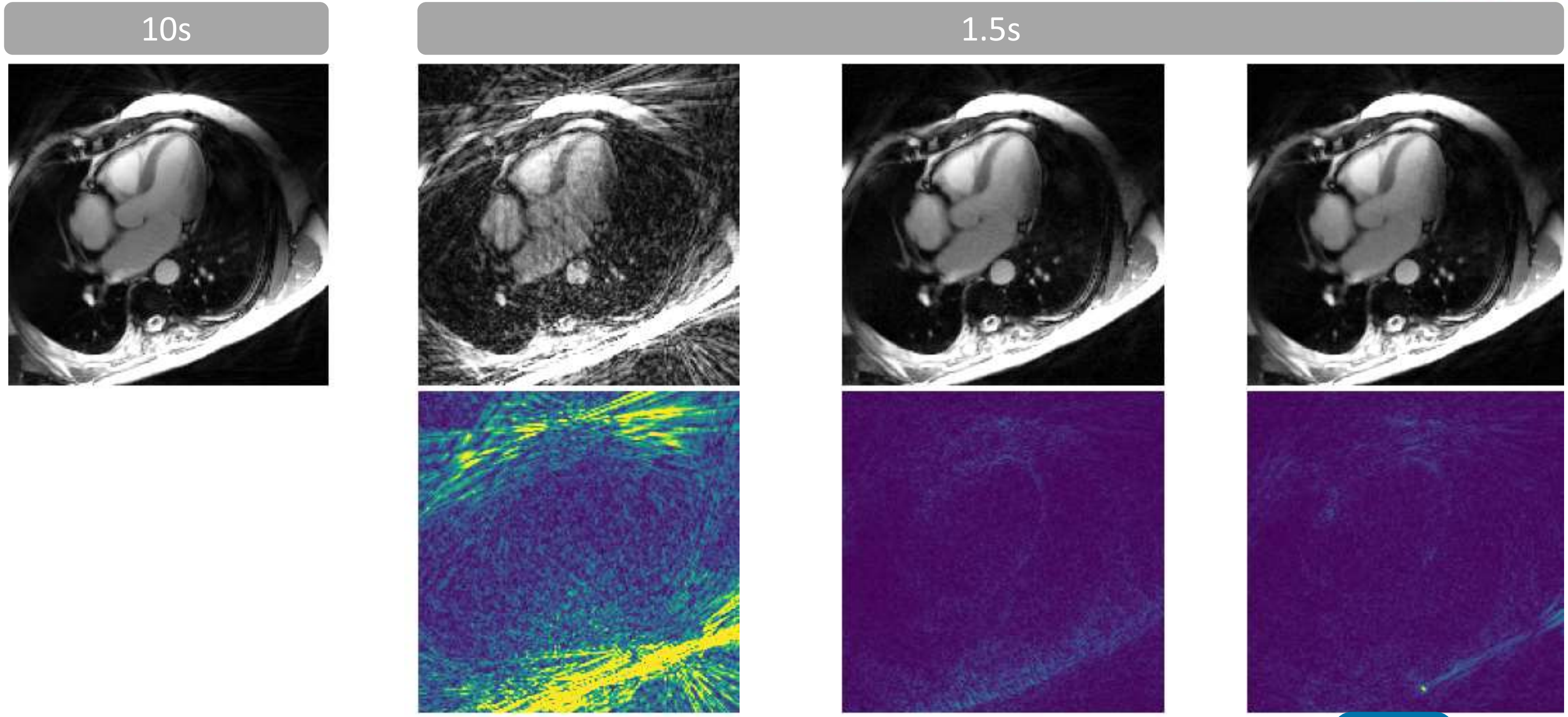
- complex CNN
- High number of parameter
- High amount on trainings data

Physics-informed deep learning



- efficient training
- **robustness**
- uncertainty

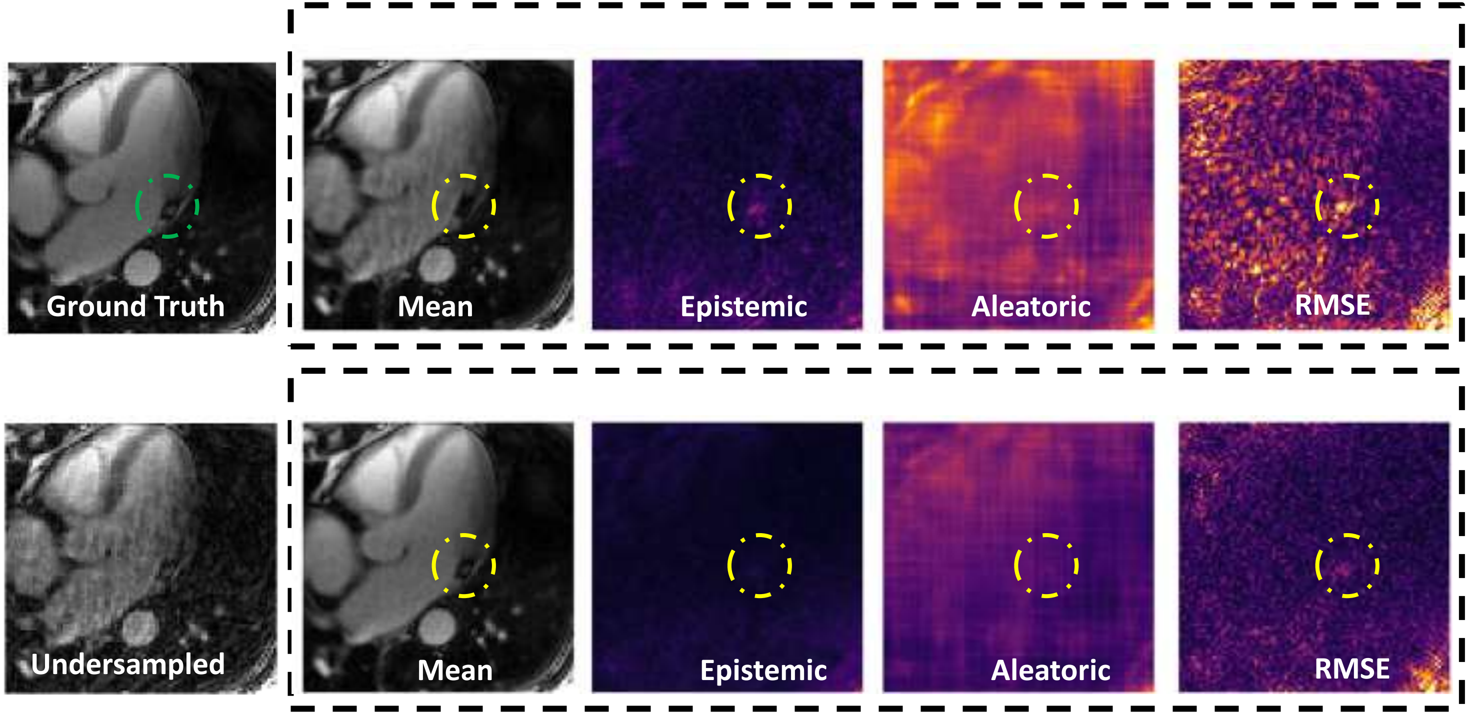
Robustness through physics input



Number of Parameter: >300000 <3000

Kofler et al., Med Phys, 2021; Kofler et al., ISBI, 2022

Reducing „ML-Uncertainty“: „Physics-Informed Learning“



Model-Agnostic

Physics-Informed

Brahma et al., Med Phys, 2023

Summary

- Modern medical diagnostics relies on multi-parametric data (in-vitro diagnostics, imaging, vital signs)
- Reliable, accurate and comparable measurements are required for selection of therapy and assessment of treatment effects
- Accurate and precise data provides the basis for the development of artificial intelligence in medicine
- Physics-informed learning increases robustness to data uncertainty

DFG Deutsche Forschungsgemeinschaft
EMPIR EURAMET
The EMPIR initiative is co-funded by the European Union's Horizon 2020 research and innovation programme and the EMPIR Participating States
DIGITAL EUROPE PROGRAMME
Bundesministerium für Wirtschaft und Klimaschutz