UNIVERSITÄT LEIPZIG

Medizinische Fakultät

Universitätsklinikum Leipzig

Medizin ist unsere Berufung.

Medical Data Science

Linking Medicine and Data Science

Leipzig, 02.12.2022

MDS

Medical Data Science

PROGRAM

Medical Data Science - An Overview (~ 20 min)
Prof. Toralf Kirsten

Distributed Analysis using the Personal Health Train (~ 5 min)
Maximilian Jugl

Generation of Synthetic Data (~ 5 min)
Dr. Sina Sadeghi

Privacy-preserving Diagnostic Support for Leukodystrophy Patients (~ 5 min)
Dr. Navid Shekarchizadeh

Predicting the Length of Stay in a Hospital (~ 5 min)
Lars Hempel

Early Detection of Sepsis: The AMPEL Project (~ 5 min)
Dr. Daniel Steinbach

Discussions and Poster Session

Medical Data Science - An Overview

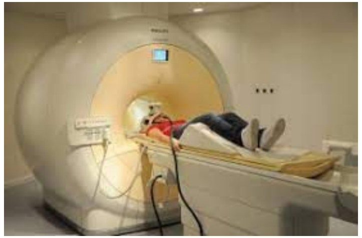
Toralf Kirsten

CHALLENGES IN MEDICINE

- Highly specialized (precision) medicine in an aging society
  - Stratified diagnostics and continued adaptation of care processes
  - Example: Support patients with rare diseases
- Development of new methods in diagnostics and therapies
  - Early diagnostic using available biomarkers
  - Prediction of continued health / care process
- Limited budget

⇒ **Need for efficient use of limited resources**

DATA IS EVERYWHERE



Diagnostics

Intensive Care

Wearables, PROMS, …

Medical Administration

Clinical Trials, Epi Studies

GOALS

- Capture and manage the right data with the right quality at the right time
- Privacy-preserving data integration and sharing
- Goal-driven data analysis
  - Patient-specific diagnostic support
  - Dedicated to specific medical research questions
  - Administrative reporting

REQUIREMENTS

Infrastructures

Organizational and technical Infrastructures → FAIR data

Methods

Spectrum allowing to

- capture and integrate data
- prepare and check data
- analyze data

ORGANIZATION AT LEIPZIG UNIVERSITY MEDICINE

Institute for Medical Informatics, Statistics & Epide-miology

Junior Research Group
Terminology and Ontology-based Phenotyping

Medical Departments, Clinics, and Institutions

MDS
Medical Data Science

Medical Data Science

Data Integration Center

Medical Informatics Center (Core Unit)

IT Department

OUR TEAM

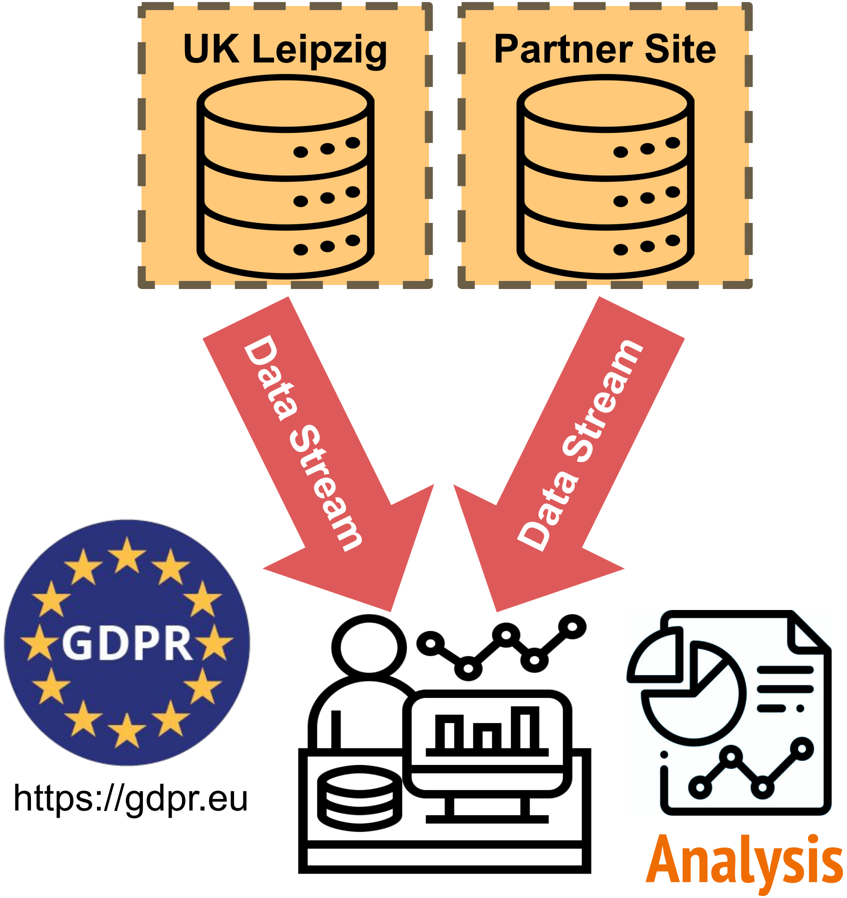
OUR EXPERTISE — OVERVIEW

- **Infrastructure**
  - Enhance data analysis in a distributed mode
  - Software applications: patient-specific guiding and diagnostic support

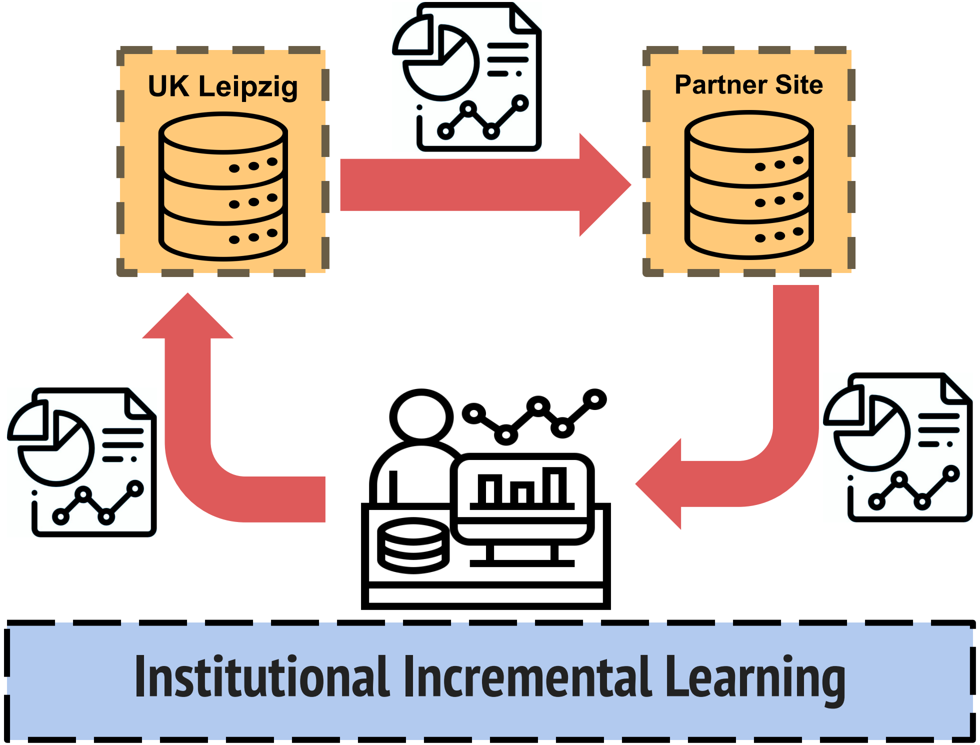
- **Methods**
  - Artificial-Intelligence-based methods for
    - Early detection, prediction, and diagnosis
    - Identification and segmentation of abnormalities in images
    - Data augmentation, especially for Rare Diseases
  - Privacy-preserving methods for data sharing

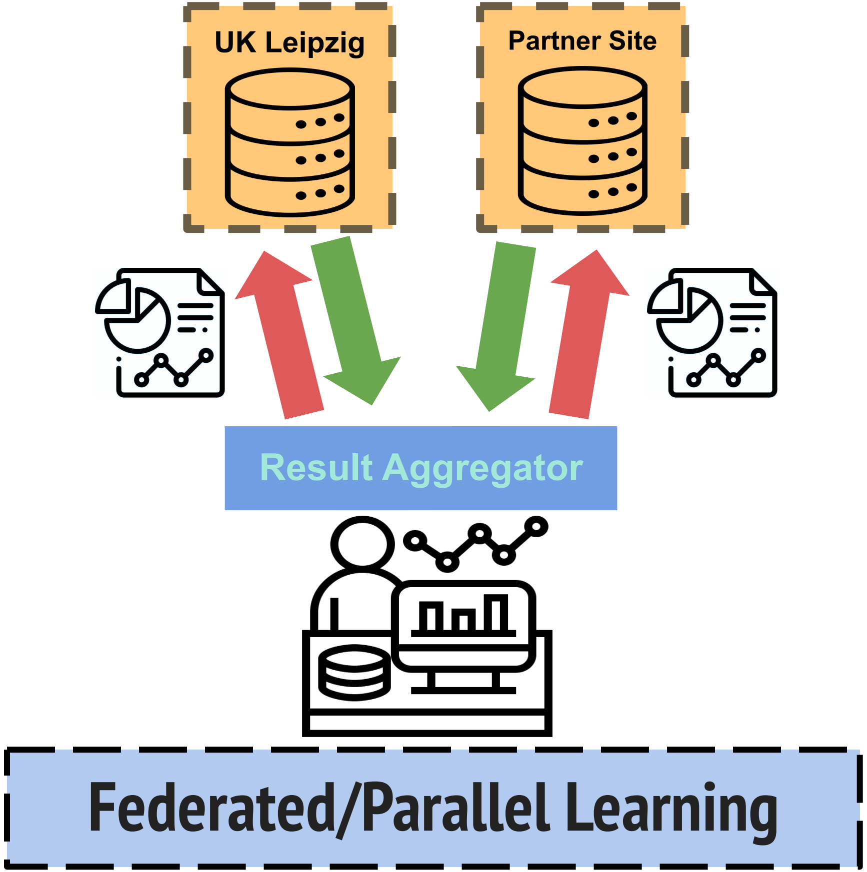
HOW TO RUN THE ANALYSIS?

- Centralized (pooled) vs. distributed data analysis
- Influencing factors
  - Number of sites
  - Trust in partner sites
  - Analysis goal and plan
  - Amount and type of data
  - Legal aspect: Consent available?
  - …



DISTRIBUTED ANALYSIS





PERSONAL HEALTH TRAIN

Publication:
Welten S, Mou Y, Neumann L, Jaberansary M, Ucer YY, Kirsten T, Decker S, Beyan O: A Privacy-Preserving Distributed Analytics Platform for Health Care Data. Methods of Information in Medicine. 61(S 01): e1-e11, DOI: 10.1055/s-0041-1740564

- **Key idea**: Bring the analysis to the data
- No direct data access for scientists: Upload the analysis script and run at each involved partner site
- Keep the control over own data at each location
- Strong cooperation with RWTH Aachen, Fraunhofer Institute for Applied Information Technology, Univ. of Cologne, Univ. of Tübingen

⇒ **Pitch Maximilian Jugl**

CHALLENGES IN DISTRIBUTED ANALYTICS

- Infrastructure
  - Computational resources depending on data volume, type and analysis tasks (algorithms) at each location

    ⇒ **Need to have physical resources or involve further partners**
- Analysis (methods)
  - Impact of routing on final result
  - Catastrophic forgetting: potential drop in accuracy with every new location
  - Unbalanced data: different cohort sizes & class imbalances at different sites, in particular for small data volumes (Rare Diseases)

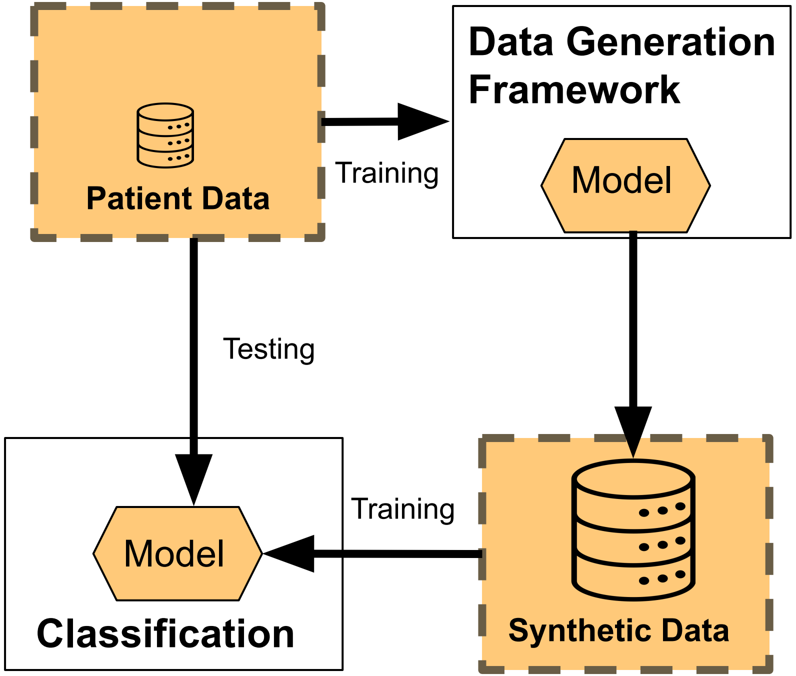
    ⇒ **Creation of synthetic data by learning from real world data**

TACKLING SMALL DATA VOLUME BY SYNTHETIC DATA

Publication:
Abedi, M.; Hempel, L.; Sadeghi, S.; Kirsten, T. **GAN-Based Approaches for Generating Structured Data in the Medical Domain**. Appl. Sci. 2022, 12, 7075. https://doi.org/10.3390/app12147075

- **Key Idea**: Use synthetic data to train AI-based models
  - Train a generative model
  - Create synthetic data
  - Train the intended (classification) model using synthetic data
  - Test the intended model by using real world data
- **Our result:** Framework for generating and testing synthetic data

⇒ **Pitch Sina Sadeghi**



CLASSIFICATION APPROACH FOR POTENTIAL LEUKODYSTROPHY PATIENTS

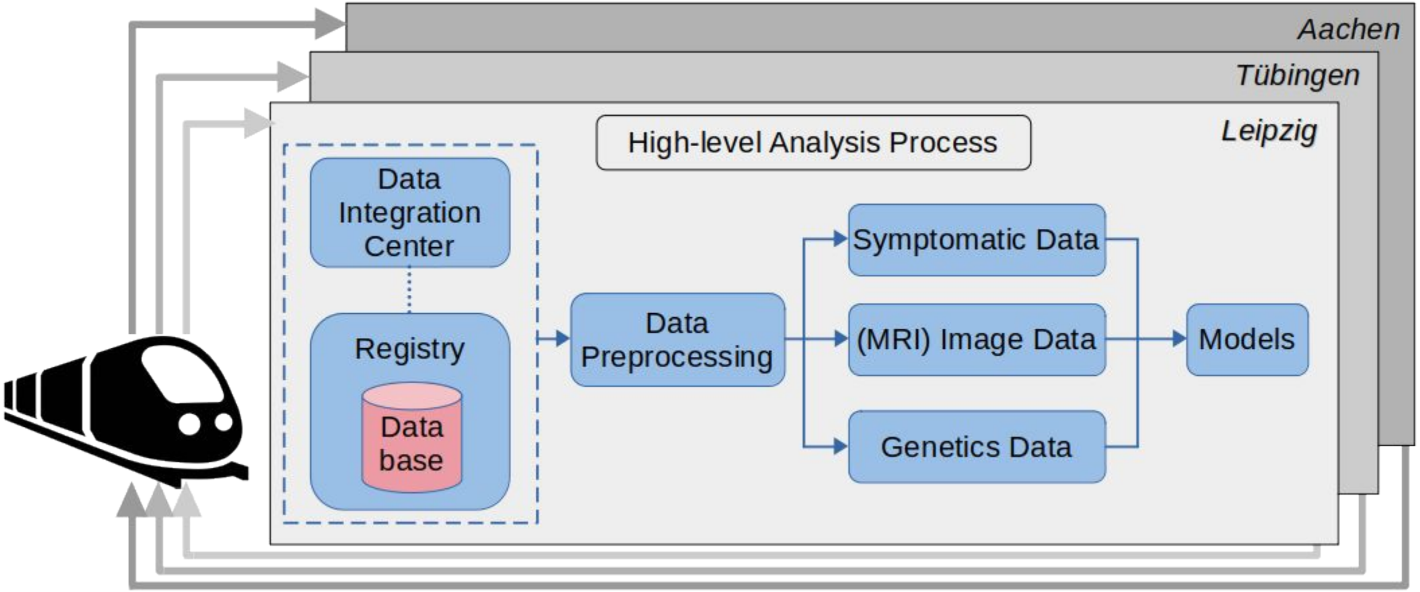
Close collaboration with local clinicians of the MZEB ❤

PERSONAL HEALTH TRAIN

PADME

https://padme-analytics.de

- ✔ Secure onsite data analysis and result shipping
- ✔ Community-based analysis code review
- ✔ Free of programming languages and data types



Creating, merging, and evaluating models

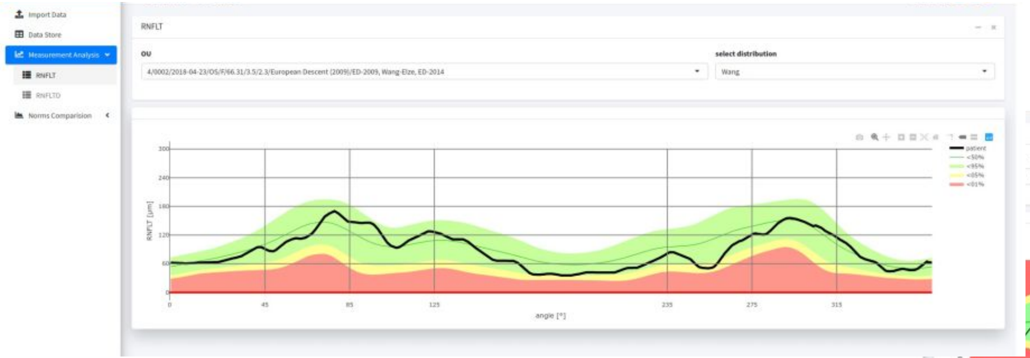
⇒ Pitch Navid Shekarchizadeh

Best Poster Award: BMG, Nov. 2022

Publication:
Thomas Peschel, Mengyu Wang, Toralf Kirsten, Franziska G Rauscher Tobias Elze. **A Cloud-based Infrastructure for Interactive Analysis of RNFLT Data.** Proc. E-Science-Tage. 54-67. 2021. https://doi.org/10.11588/heibooks.979

Designing Diagnostic Support Applications

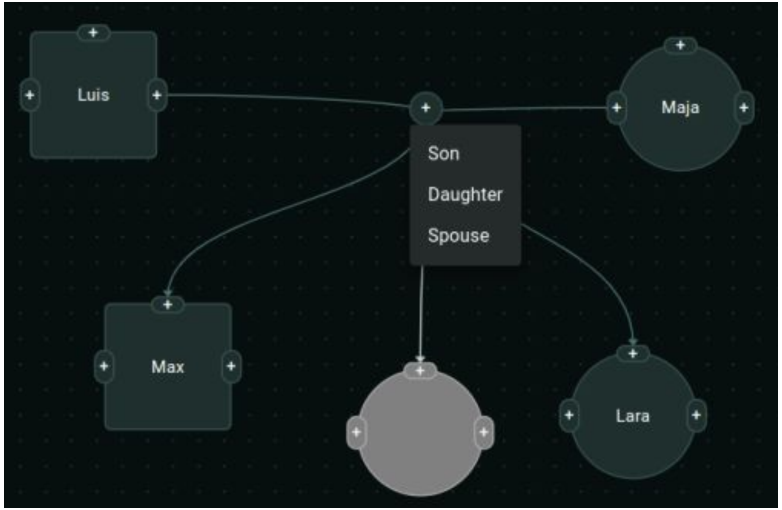
- **Goal**: Bring research results into practice
- Example: RNFLT-Visualizer (Retinal Nerve Fiber Layer Thickness)
  - Generate new normative data based on a large cohort study (LIFE-Adult)
  - Web-based software application importing and visualizing OCT measurements with normative data



⇒ **Poster available**

Designing Applications

- Example: Family Manager
  - Capture inheritances of study participants in a pseudonymized way
  - Join other scientific data or consider disease data within a pedigree



Family Manager application

SCIENTIFIC PROJECTS WITH FUNDING

- **SMITH (01’2018 - 12’2022, 01’2023 - 12’2026)**
  - Goal: Establishing data integration centers
  - Target: University hospitals
- **NFDI4Health (10’2020 - 09’2025)**
  - Goal: Establishing research infrastructure for study data
  - Target: Clinical trial centers, Large epidemiological studies
- **LEUKO-Expert (10’2020 - 09’2023)**
  - Goal: Guiding support for potential Leukodystrophy patients
  - Target: University Hospitals Leipzig, Tübingen, and Aachen
- **Tag-White (12’2021 - 11’2024)**
  - Goal: Diagnostic support for Leukodystrophy patients
  - Target: Leipzig University Medical Center
- **FAIR - Data Spaces (05’2021 - 04’2024)**
  - Goal: Distributed health data analytics using Gaia-X infrastructure
  - Target: Univ. Leipzig & several partners
- **NUM CODEX+ (01’2022 - 12’2022)**
- **ScaDS.AI (01’2022 - 12’2027)**

Upcoming Projects

- **PATH (12’2022 - 11’2025)**
  - Goal: Secured sharing of health and wellness data
  - Target: Univ. Leipzig & Dresden
- **PrivateAIM (04’2023 - 03’2027)**
  - Goal: Interoperable, secured distributed analytics infrastructure in Germany
  - Target: University Hospitals & other partner sites

OTHER PROJECTS

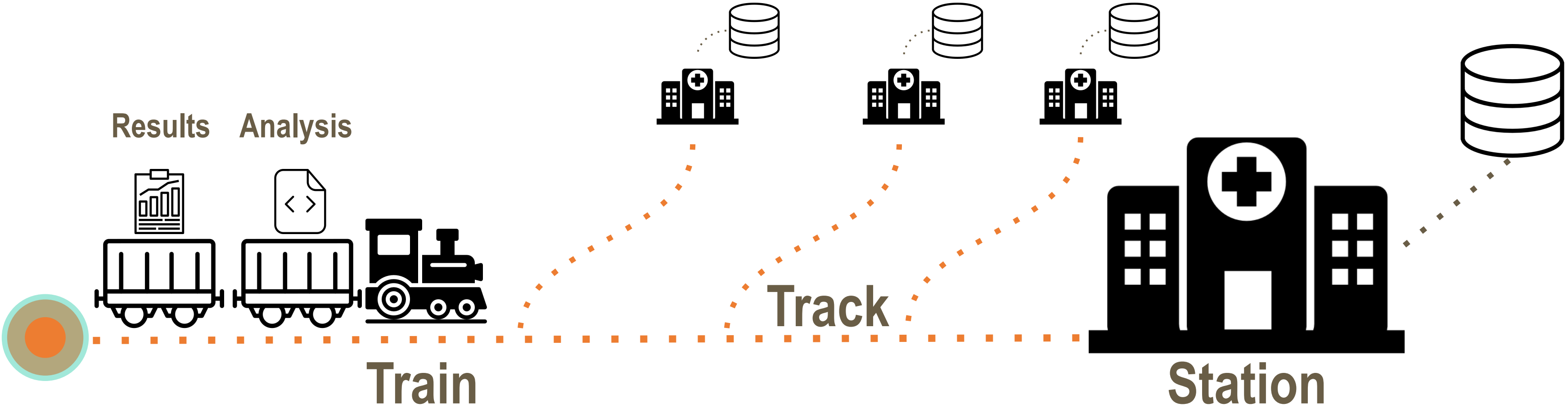
- **Predicting the length of stay in a hospital ⇒ Pitch Lars Hempel**
  - Goal: Planning support after hospital admission (example: ICU)
  - Target: Leipzig University Medical Center, but could be extended to an overarching project
- **Designing and analysing patient satisfaction (questionnaire-based)**
  - Goal: Analysis of strengths and weakness
  - Target: Leipzig University Medical Center - MZEB
- **DiaClusT**
  - Goal: Find patterns in Diabetes mellitus type 2 patients for a better and dedicated treatment
  - Target: Leipzig University Medical Center (extended to all University Hospitals planned)
- **Privacy-preserving record linkage**
  - Goal: Establishing an infrastructure for record linkage as part of the PHT
  - Target: All University Medical Centers and other partners (e.g., registries)

Distributed Analysis: The Personal Health Train

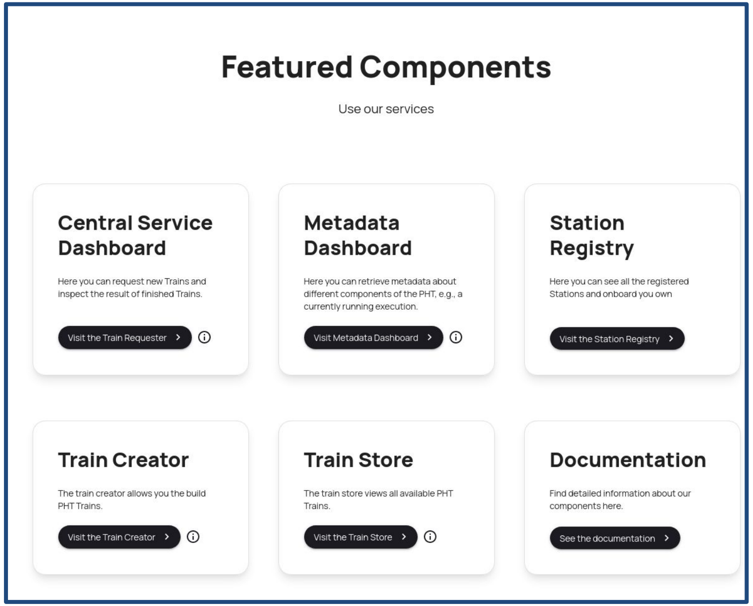
Maximilian Jugl

PERSONAL HEALTH TRAIN

- **Paradigm:** Bring the analysis to the data
- The data remains under the control of the data holders (**Station**)
- Analysis travels between stations to update the results (**Train**)



PERSONAL HEALTH TRAIN — PADME



- PHT implementations are in active development (e.g. PADME, VANTAGE6)
- Web portal with access to different components for running distributed analysis tasks

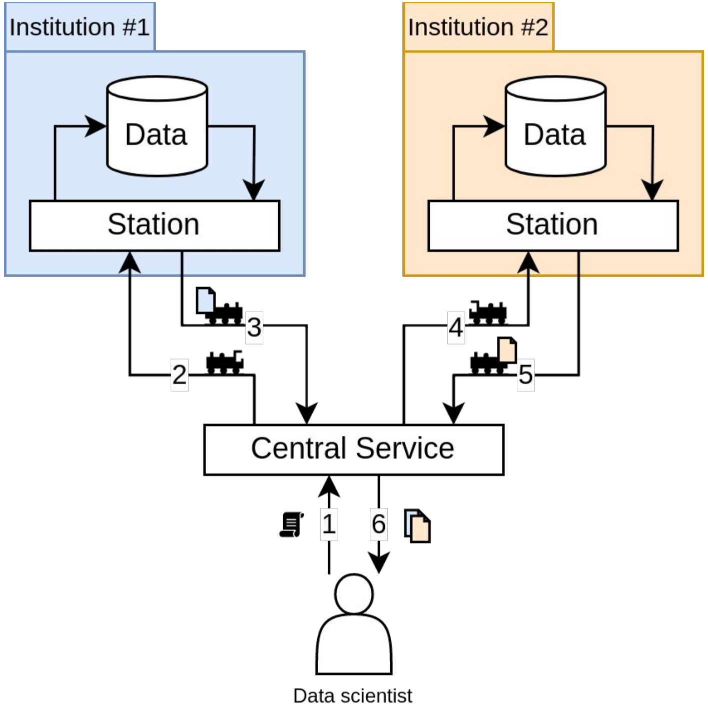
PADME

padme-analytics.de

Publication:
Welten, S., Mou, Y., Neumann, L., Jaberansary, M., Ucer, Y. Y., Kirsten, T., ... & Beyan, O. (2022). **A Privacy-Preserving Distributed Analytics Platform for Health Care Data**. Methods of Information in Medicine.

PERSONAL HEALTH TRAIN — Procedure

- Data scientist uploads analysis script to **central service** (CS)
- CS constructs **train class**
- Data scientist selects train class and target stations
- CS sends train to selected stations
- Station administrators provide access to relevant resources to train
- Train aggregates results and provides them to data scientist when back at CS

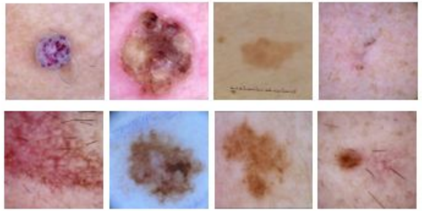


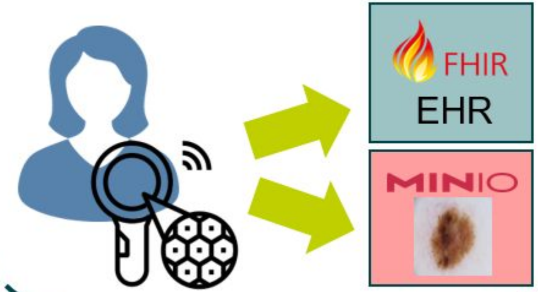
PHT APPLICATION SCENARIO

Publication:
Mou, Y.; Welten, S.; Jaberansary, M.; Ucer Yediel, Y.; Kirsten, T.; Decker, S.; Beyan, O.; **Distributed Skin Lesion Analysis Across Decentralised Data Sources**. Stud Health Technol Inform. 2021. https://doi.org/10.3233/SHTI210179

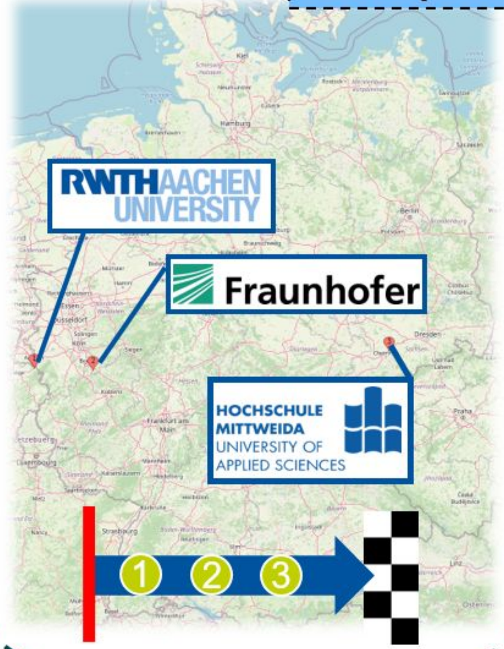
Dataset & Task

- ISIC 2019[1] Data set
- Classify dermoscopic images
- **Eight different diagnostic categories:**





Participants

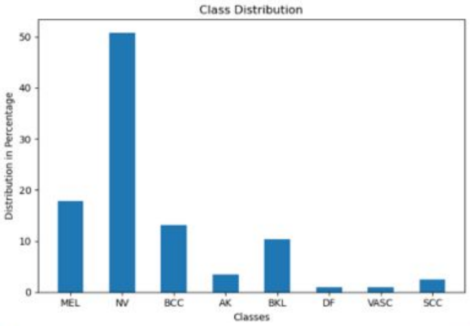


Results

- **General: Comp. to central model**
- **Some metrics:**

Central vs. PHT
75.40 vs. 71.83% (accuracy)
69.22 vs. 63.35% (recall)

- **In-depth statistical analysis:**



Study setup | Collaboration | Outcome

[1] https://challenge.isic-archive.com/landing/2019/ (Open-Source)

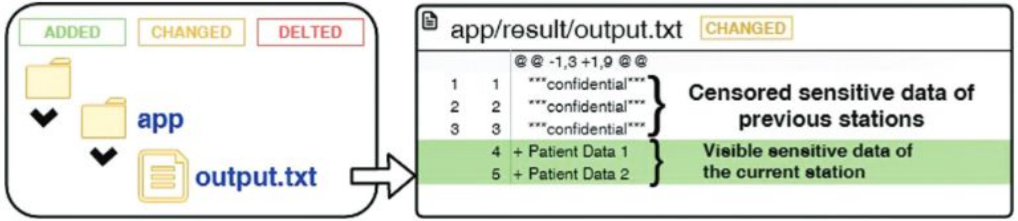
OTHER APPLICATION SCENARIOS

- Lesion detection for Leukodystrophies ⇒ **Poster available**
  - Design of an image synthesis and processing pipeline
  - Train is used for classification of annotated MRIs
- Privacy-preserving record linkage ⇒ **Poster available**
  - Integration of web services for PPRL into the PHT
  - Validation on synthetic data, real-world use case coming soon™
- Model training on skin lesions from ISIC dataset with 3 stations
- Model training on Breast Cancer Wisconsin dataset with 6 stations

Publications:
1) Mou, Y.; Welten, S.; Jaberansary, M.; Ucer Yediel, Y.; Kirsten, T.; Decker, S.; Beyan, O.; **Distributed Skin Lesion Analysis Across Decentralised Data Sources**. Stud Health Technol Inform. 2021.
2) Welten, Sascha; Hempel, Lars; Abedi, Masoud; Mou, Yongli; Jaberansary, Mehrshad; Neumann, Laurenz; Weber, Sven; Tahar, Kais; Yediel, Yeliz; Löbe, Matthias; Decker, Stefan; Beyan, Oya; Kirsten, Toralf: **Multi-Institutional Breast Cancer Detection Using a Secure On-Boarding Service for Distributed Analytics.** Applied Sciences 12. 4336. 2022.

SECURITY IN THE PERSONAL HEALTH TRAIN

- Trains are encrypted between stations using public-key cryptography
- Station administrators are in control of the analysis execution
  - Define access to shared resources with train
  - Reject malicious trains
  - Inspect results after analysis execution
- Only analysis results leave the station



Publication:
Welten, S., Mou, Y., Neumann, L., Jaberansary, M., Ucer, Y. Y., Kirsten, T., ... & Beyan, O. (2022). **A Privacy-Preserving Distributed Analytics Platform for Health Care Data**. Methods of Information in Medicine.

Generation of Synthetic Data

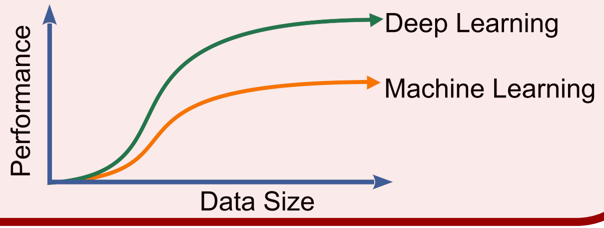
Sina Sadeghi

Medical Data Scarcity

- Patient Privacy
  (General Data Protection Regulation)
- Rare Diseases
  (with limited available data)

Artificial Intelligence Models

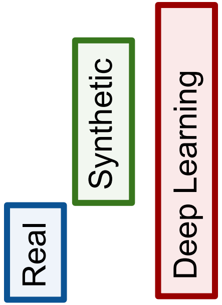
➢ require a large amount of data



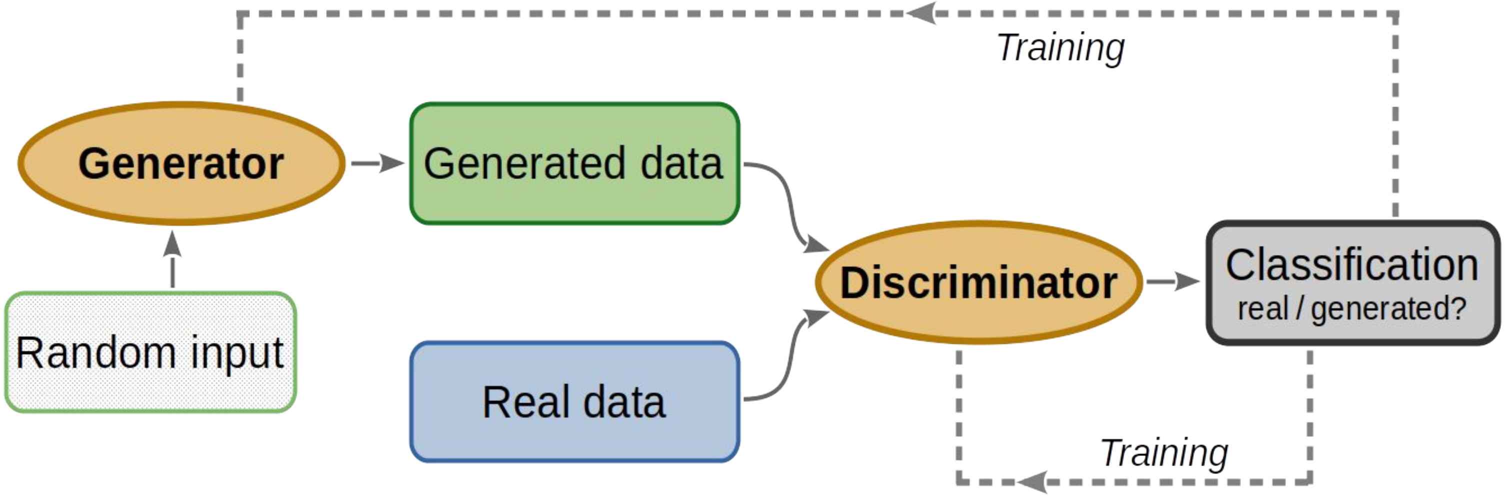
Generating Synthetic Data

- without reference to real patients
- improve data analysis performance

⇨ **Deep Generative Models**



Generative Adversarial Networks – GANs



Generating Tabular Data in the Medical Field

GANs

- Impressive results in generating images
- Challenging performance on tabular data especially in medical domain with limited data

GAN variants

➢ generate tabular data
➢ consider limited amount of data (small subsets of real data)

Breast Cancer Wisconsin Data

data patients with breast tumors

Primary task: Binary Classification (cancerous / non-cancerous tumors)

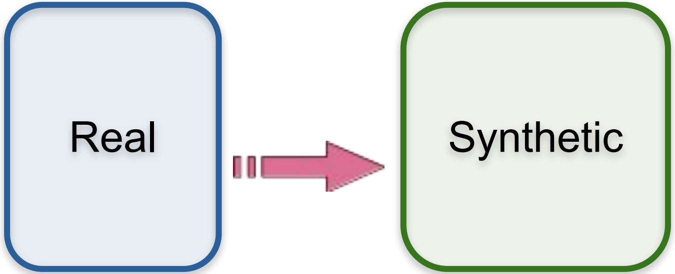
	Radius	Texture	Perimeter	Area	Smoothness	Compactness
0	17.99	10.38	122.80	1001.0	0.11840	0.27760
1	20.57	17.77	132.90	1326.0	0.08474	0.07864
2	19.69	21.25	130.00	1203.0	0.10960	0.15990
3	11.42	20.38	77.58	386.1	0.14250	0.28390
4	20.29	14.34	135.10	1297.0	0.10030	0.13280

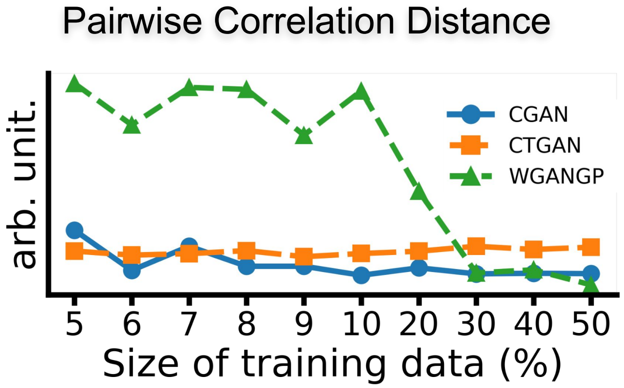
Wolberg, W.; Street, W.; Mangasarian, O. Breast Cancer Wisconsin (Diagnostic); UCI Machine Learning Repository. 1995.

Quality of Synthetic Data

- Conventional methods

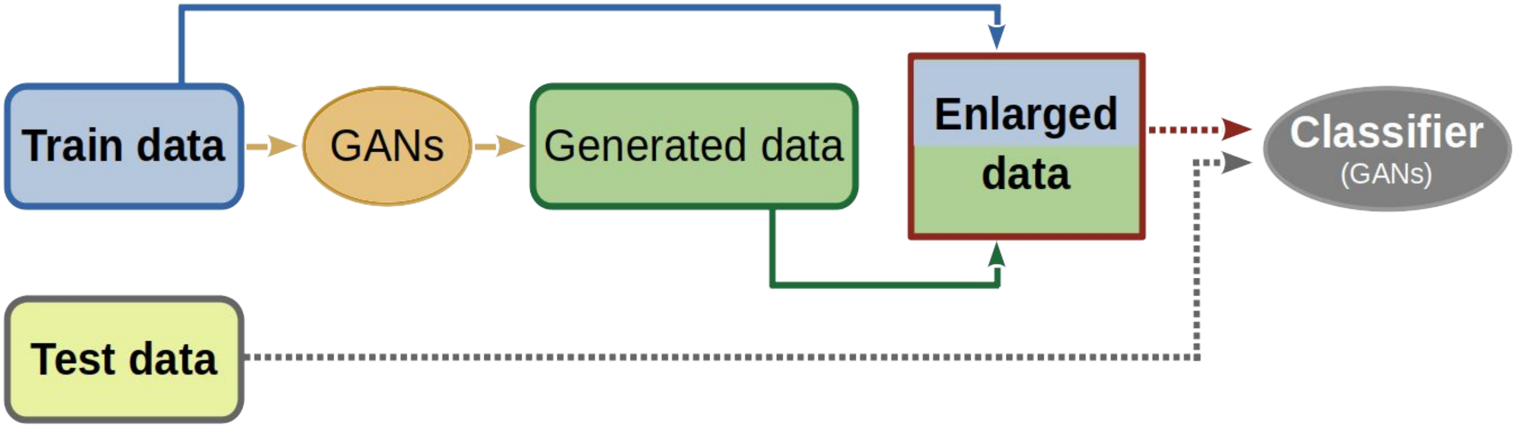
  characteristics of real data captured by synthetic data





- Target application: Binary Classification

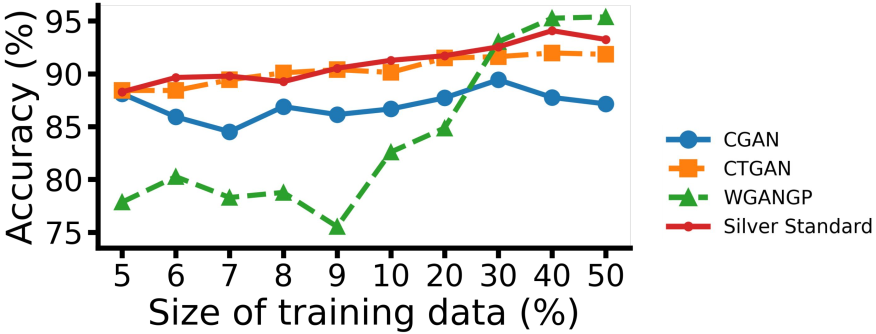
Evaluation Framework



Classification Accuracy

Abedi, M.; Hempel, L.; Sadeghi, S.; Kirsten, T. **GAN-Based Approaches for Generating Structured Data in the Medical Domain**. Appl. Sci. 2022, 12, 7075.

Accuracy of Classifiers using data from GANs



Improved accuracy by including synthetic data

- ❑ compared to Silver Standard (with only real data)
- ❑ even with small amount of available data

Abedi, M.; Hempel, L.; Sadeghi, S.; Kirsten, T. **GAN-Based Approaches for Generating Structured Data in the Medical Domain**. Appl. Sci. 2022, 12, 7075.

Privacy-preserving Diagnostic Support for Leukodystrophy Patients

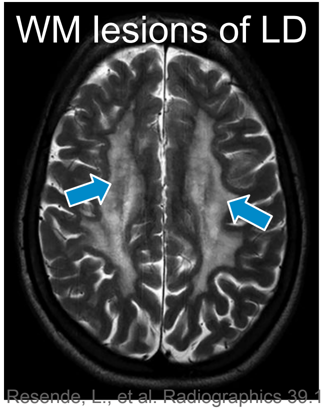
Navid Shekarchizadeh

Close collaboration with local clinicians of the MZEB ❤

LEUKODYSTROPHIES: Introduction

- Leukodystrophies (LD):
    - A group of genetic disorders
    - Neurodegenerative
    - Rare
    - Fatal
    - Differential diagnosis: Multiple Sclerosis (MS)

- **Goal:** Developing a diagnostic advisory system
    - Reusing the data of medical centers
    - Federated learning of artificial intelligence models
    - Preserving patients' privacy



Resende, L., et al. Radiographics 39.1(2019)

Best Poster Award: BMG, Nov. 2022

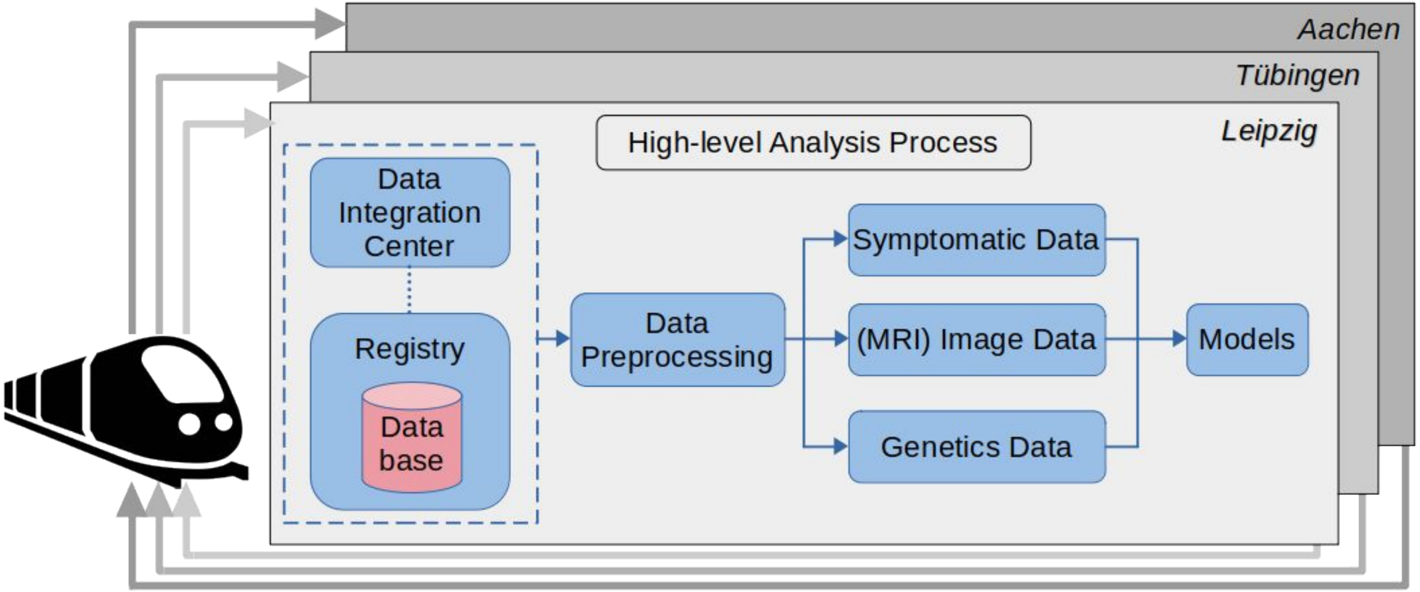
CLASSIFICATION APPROACH FOR POTENTIAL LEUKODYSTROPHY PATIENTS

PERSONAL HEALTH TRAIN

PADME

https://padme-analytics.de

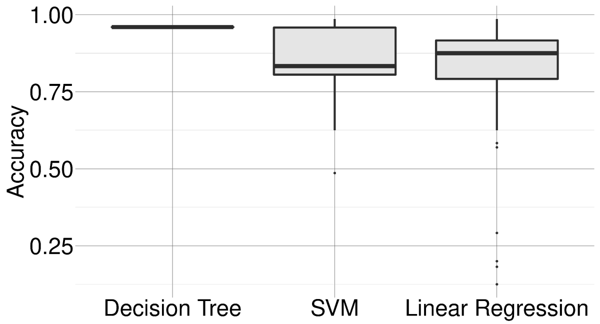
- ✔ Secure onsite data analysis and result shipping
- ✔ Community-based analysis code review
- ✔ Free of programming languages and data types



Creating, merging, and evaluating models

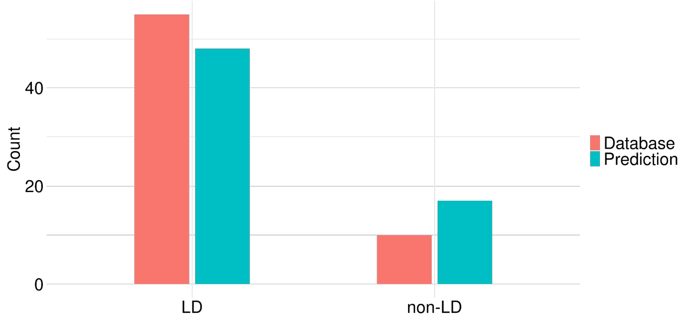
DIFFERENT DATA TYPES

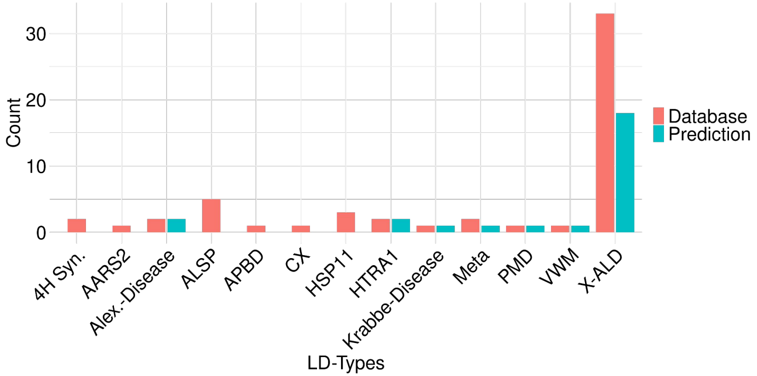
- Symptomatic data
  - Questionnaires and input forms
  - With HPO annotations in REDCap
  - Enlarged data volume by generating synthetic data
- Results
  - Stable accuracy for LD / non-LD classification



DIFFERENT DATA TYPES

- Genetics data
  - Rule-based approach using PhenoMan system (TOP JRG)
  - Reuse of publicly available genetic annotations (e.g., ClinVar)
- Results
  - High classification accuracy (91%)



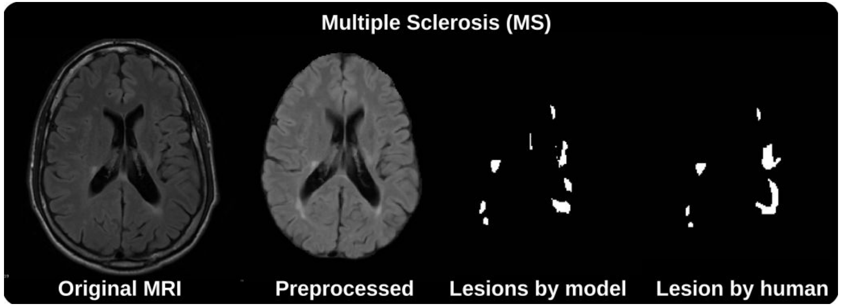


Publication:
Uciteli A, Beger C, Wagner J, Kiel A, Meineke FA, Stäubert S, Löbe M, Hänsel R, Schuster J, Kirsten T, Herre H. **Ontological Modelling and Execution of Phenotypic Queries in the Leipzig Health Atlas**. Stud Health Technol Inform. 2021 May 24;278:66-74. doi: 10.3233/SHTI210052. PMID: 34042877.

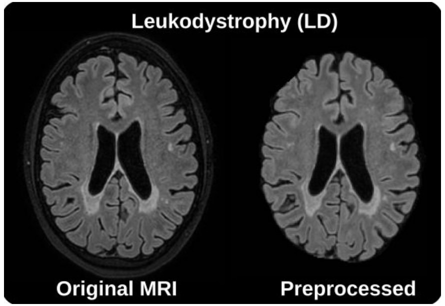
DIFFERENT DATA TYPES

- MRI data
  - Exported and managed by a XNAT system
  - Completed pre-processing pipeline: de-identification, skull stripping, and registration
  - Implemented a deep learning algorithm for MS lesion segmentation (similarity to LD)
  - Challenges:
    - Harmonizing the images from different MRI scanners and configurations
    - Applying the lesion segmentation algorithm to LD (rare disease)

- Results
  – Promising lesion segmentation of MS



  – Preprocessing LD images



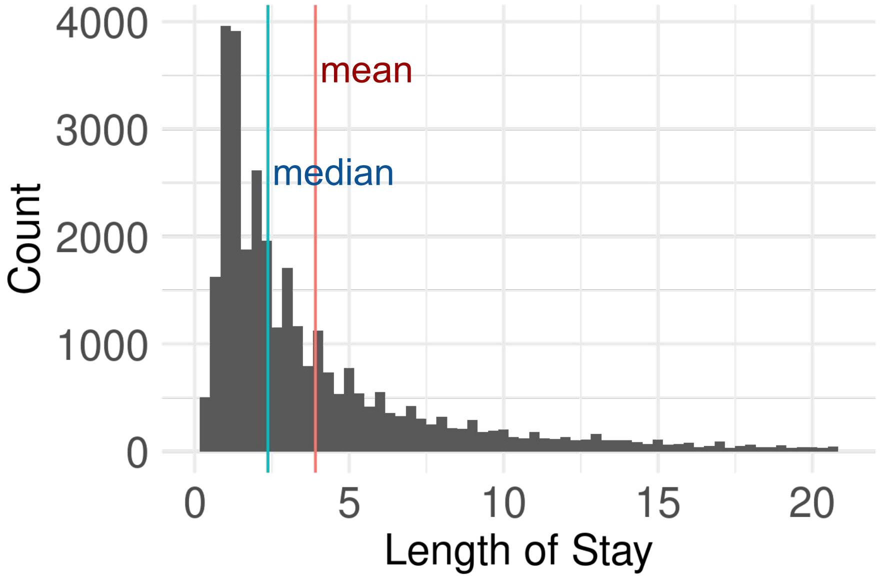
Predicting the Length of Stay in a Hospital

Lars Hempel

Introduction

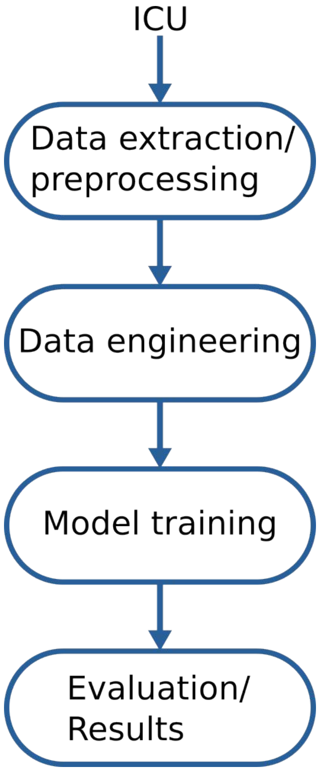
- Resource management
  - Indicator for hospital efficiency
  - Reducing costs

- Data: MIMIC IV
  - Demographics (e.g. Age, Sex)
  - Examination (e.g. height, weight)
  - Lab (e.g. Glucose/Potassium Concentration)



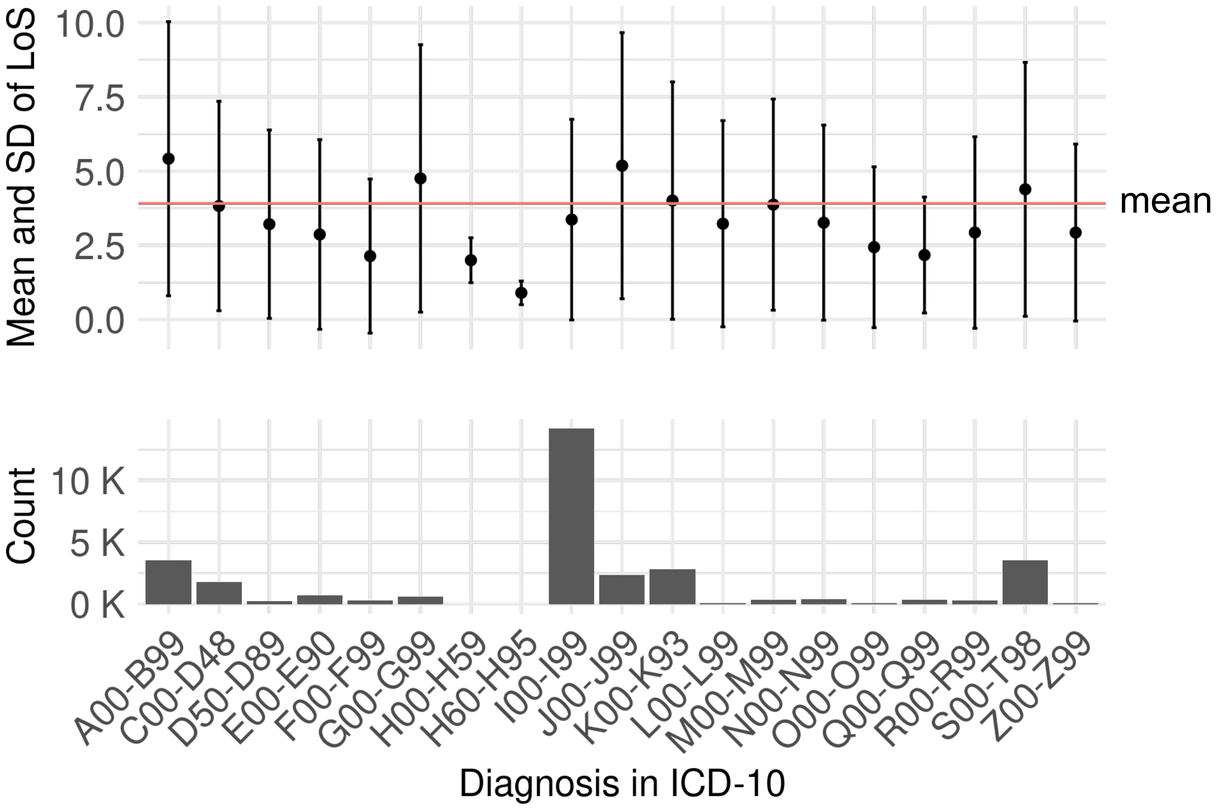
Preprocessing & Methods

- Extract from the database all ICU patients
- Remove missing data
- Filter on:
    - Only adults (age over 18)
    - Weight smaller than 30 Kg
    - Height bigger than 250 cm
    - Exclude patients who died within the stay
- Methods: mean, median, linear Regression, Support Vector Machine and Random Forest
- Split Train/Test (80/20) 10 times
- Evaluation metrics: RMSD, MAD and R^2



Data exploration Diagnosis

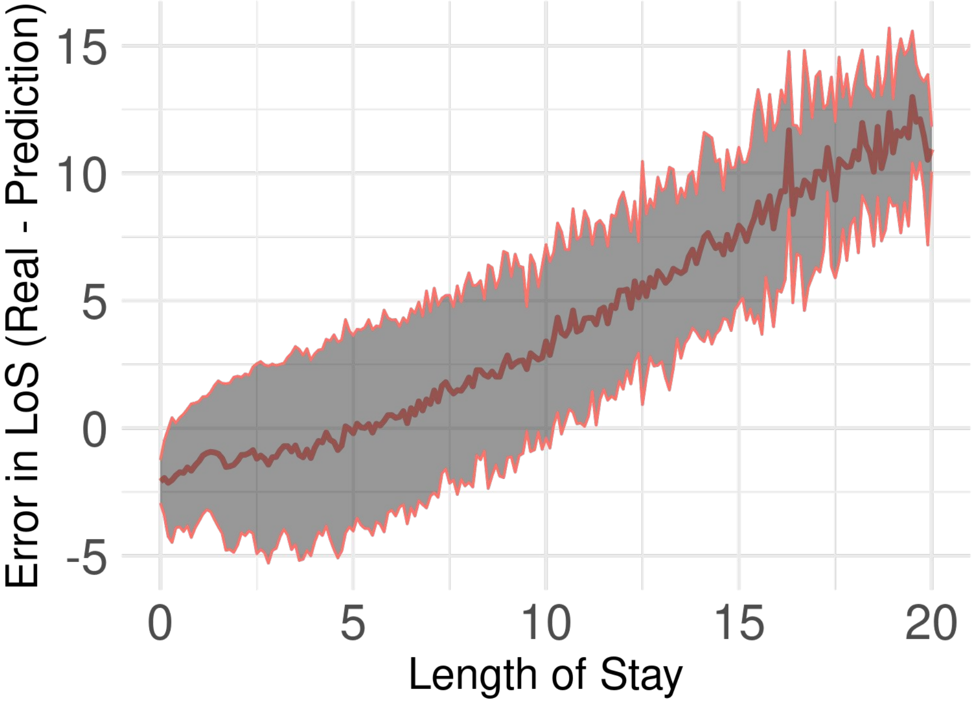
- #stay: 31916
- #features: 20
- mean: 3.91
- median: 2.37
- ICD-10 Coding System
- 14183 in I00 to I99 -> Cardiovascular Diseases



Results

- Best model: Random Forest
- Problems with larger Length of Stay in the Database
- RMSD: 2.88

- Future work:
  - Mixed models for long and short stays
  - Transfer knowledge to hospital data



Early Detection of Sepsis: The AMPEL Project

Daniel Steinbach

UNIVERSITÄT LEIPZIG
Medizinische Fakultät

Universitätsklinikum Leipzig
Medizin ist unsere Berufung.

Early Detection of Sepsis

Dr. med. Daniel Steinbach

AMPEL

Universität Leipzig
Medizinische Fakultät
Universitätsklinikum Leipzig
Medizin ist unsere Berufung.

AMPEL 2018 - 2022
eHealthSax

→

AMPEL CDSS 2022 - 2024
KHZG

Leipzig

Grimma

Wurzen

Belegung [illegible] mit 23 Patienten

Zimmer	Bett	Gsp.	Patient/Geschl./Alter	GebDatum	AMPEL	R	Studie	Abw.	PP
F4092	F04X1014								
F4090	F04X1010								
F4094	F04X1017		[illegible]	[illegible]					
F4084	F04X1002		[illegible]	[illegible]					
F4097	F04X1021		[illegible]	[illegible]					
F4092	F04X1013		[illegible]	[illegible]					
F4091	F04X1012		[illegible]	[illegible]					
F4098	F04X1024		[illegible]	[illegible]					
F4084	F04X1001		[illegible]	[illegible]					
F4086	F04X1003		[illegible]	[illegible]					
F4095	F04X1019		[illegible]	[illegible]					
F4088	F04X1005		[illegible]	[illegible]					
F4098	F04X1023		[illegible]	[illegible]					
F4095	F04X1020		[illegible]	[illegible]					

AMPEL Meldesystem

Automatisch generierter Hinweis aus dem AMPEL-Meldesystem

Es handelt sich um ein Forschungssystem zur Unterstützung der Patientensicherheit. Auf Grund des Forschungscharakters und der Randomisierung kann die Verfügbarkeit und Validität der Alarme nicht zugesichert werden. Es darf nicht allein aus der Alarmierung eine medizinische Konsequenz abgeleitet werden. Die zugrundeliegenden Befunde müssen geprüft werden.

Website: www.ampel.care | laboratoriumsmedizinischer Dienst: 22221

Name	Geschlecht	Alter	Ausgabe am
Einrichtung	Fall	Patient	Uhrzeit
1			

AMPEL

AM-ANÄM-L

AKIN

HYPERKALZIÄMIE

HYPOKALIÄMIE

LAKTAT

HYPONATRIÄMIE

REFEEDING SYNDROM

AMPEL Meldesystem

Automatisch generierter Hinweis aus dem AMPEL-Meldesystem

Es handelt sich um ein Forschungssystem zur Unterstützung der Patientensicherheit. Auf Grund des Forschungscharakters und der Randomisierung kann die Verfügbarkeit und Validität der Alarme nicht zugesichert werden. Es darf nicht allein aus der Alarmierung eine medizinische Konsequenz abgeleitet werden. Die zugrundeliegenden Befunde müssen geprüft werden.

Website: www.ampel.care | laboratoriumsmedizinischer Dienst: 22221

Name	Geschlecht	Alter	Ausgabe am
Einrichtung	Fall	Patient	Uhrzeit

AM-ANÄM-L

Status	Angelegt am	Angelegt um	Analyt	Ergebnis	Quittiert am	Quittiert um	Aktion
	10.11.2022	09:42:31	Hämoglobin	2,3 mmol/l			Quittieren
	09.11.2022	12:42:49	Hämoglobin	2,1 mmol/l			
	07.11.2022	10:44:53	Hämoglobin	2,6 mmol/l			
	04.11.2022	11:56:47	Hämoglobin	2,4 mmol/l			
	02.11.2022	12:14:30	Hämoglobin	2,6 mmol/l			

AKIN

HYPERKALZIÄMIE

AMPEL Meldesystem

Automatisch generierter Hinweis aus dem AMPEL-Meldesystem

Es handelt sich um ein Forschungssystem zur Unterstützung der Patientensicherheit. Auf Grund des Forschungscharakters und der Randomisierung kann die Verfügbarkeit und Validität der Alarme nicht zugesichert werden. Es darf nicht allein aus der Alarmierung eine medizinische Konsequenz abgeleitet werden. Die zugrundeliegenden Befunde müssen geprüft werden.

Website: www.ampel.care | laboratoriumsmedizinischer Dienst: 22221

Name	Geschlecht	Alter	Ausgabe am
Einrichtung	Fall	Patient	Uhrzeit
1			

AM-ANÄM-L

AKIN

HYPERKALZIÄMIE

HYPOKALIÄMIE

LAKTAT

HYPONATRIÄMIE

REFEEDING SYNDROM

Nutrition team

Infectiology

Nephrology

Endocrinology

AMPEL
CDSS

Cardiology

Transfusion medicine

Medical controlling

?

Nutrition team

Infectiology

Nephrology

Endocrinology

AMPEL
CDSS

Cardiology

Transfusion medicine

Medical controlling

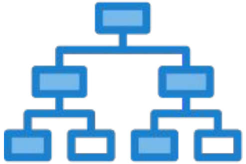
?

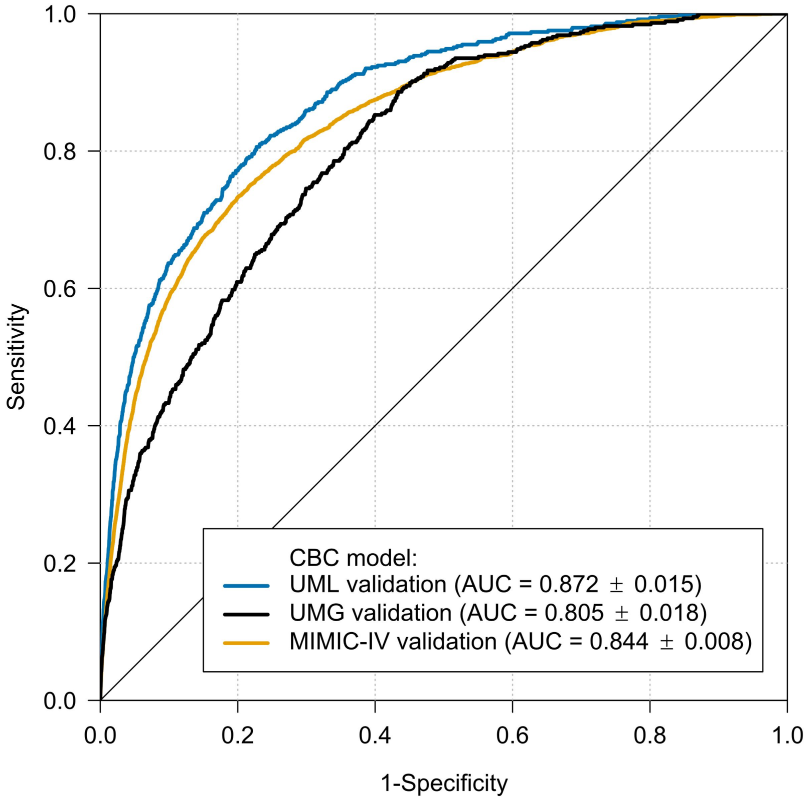
Blood Count (BC)

Label

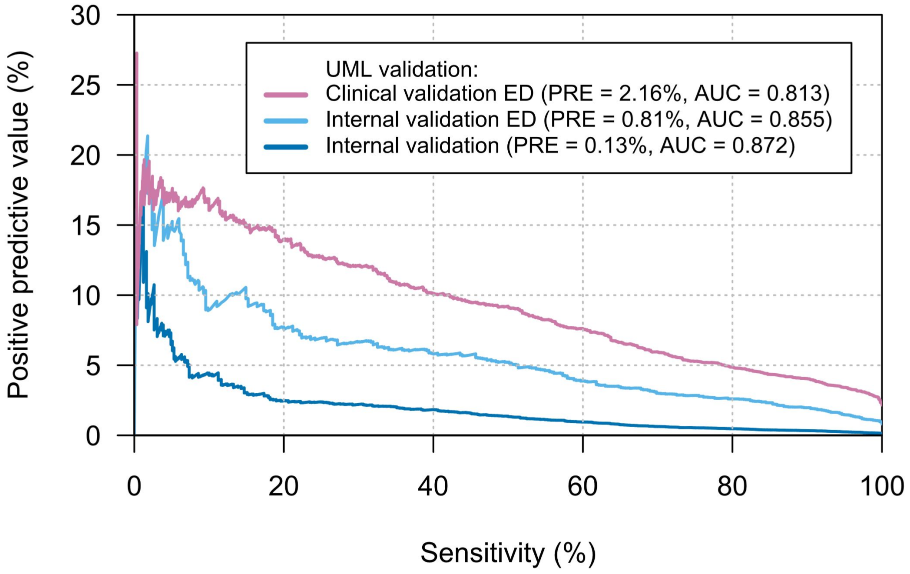
0-6h before ICU admission with sepsis

Random Forest





1.0
0.8
0.6
0.4
0.2
0.0
Sensitivity
0.0
0.2
0.4
0.6
0.8
1.0
1-Specificity
CBC model:
UML validation (AUC = 0.872 ± 0.015)
UMG validation (AUC = 0.805 ± 0.018)
MIMIC-IV validation (AUC = 0.844 ± 0.008)



medRxiv
THE PREPRINT SERVER FOR HEALTH SCIENCES
Cold Spring Harbor Laboratory BMJ Yale

Applying Machine Learning to Blood Count Data Predicts Sepsis with ICU Admission

Paul C. Ahrens, Daniel Steinbach, Maria Schmidt, Martin Federbusch, Lara Heuft, Christoph Lübbert, Matthias Nauck, Matthias Gründling, Berend Isermann, Sebastian Gibb, Thorsten Kaiser

doi: https://doi.org/10.1101/2022.10.21.22281348

Dataset published ✓
Code published ✓
R-Packages published ✓

UNIVERSITÄT LEIPZIG
Medizinische Fakultät

Universitätsklinikum Leipzig
Medizin ist unsere Berufung.

„Come Together“

- John Lennon

www.ampel.care

AMPEL

sepsis preprint

OPEN FOR COLLABORATIONS

- We are open for scientific medical research projects
- If you already have an idea, please get in contact with us

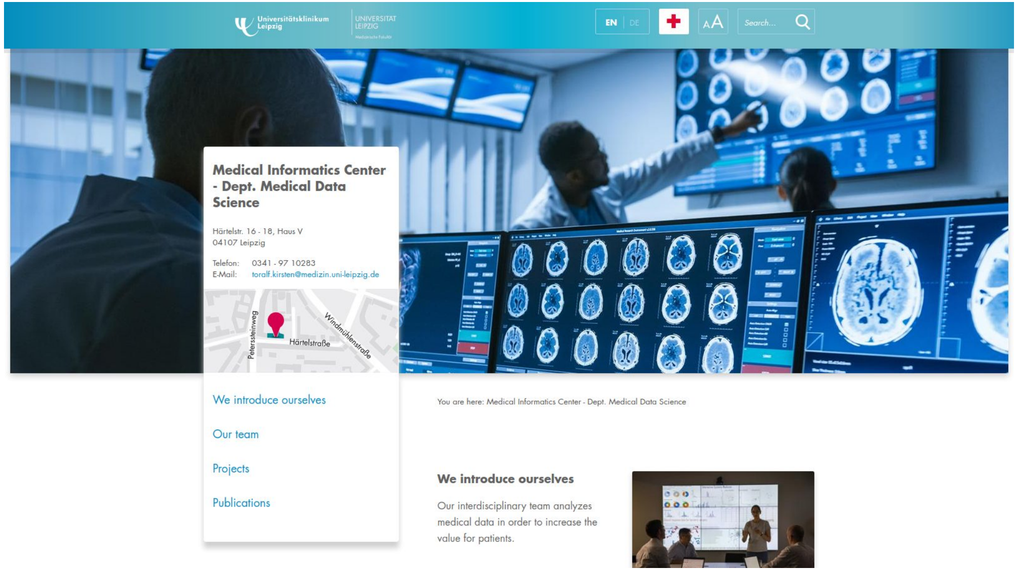
https://www.uniklinikum-leipzig.de/einrichtungen/medical-data-science

THANK YOU!

Toralf Kirsten
Medical Data Science

Universität Leipzig
T +49 341 97-10283

toralf.kirsten@medizin.uni-leipzig.de
https://www.uniklinikum-leipzig.de/einrichtungen/medical-data-science



There are refreshments and posters in front of this lecture hall. We are delighted to answer questions.

UNIVERSITÄT LEIPZIG
Medizinische Fakultät

Universitätsklinikum Leipzig
Medizin ist unsere Berufung.

Posters

Leipzig, 02.12.2022

MDS
Medical Data Science

SMITH.
Klinische Forschung und Patientenversorgung nachhaltig verbessern.

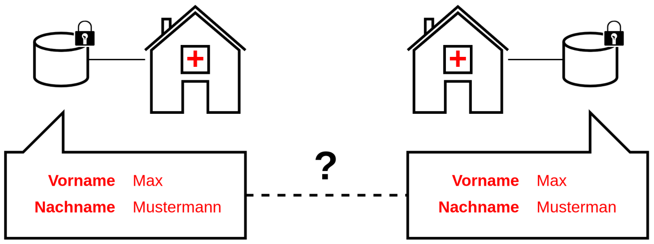
www.smith.care

Privatsphärenschützende Datenverknüpfung in verteilten Analysen mit dem PHT

Maximilian Jugl[1,2,3], Sascha Welten[4], Yongli Mou[4], Yeliz Ucer Yediel[5], Oya Beyan[5], Ulrich Sax[6], Toralf Kirsten[1,2,3]

Affiliations: 1) Abteilung Medical Data Science, Medizininformatikzentrum Leipzig 2) Institut für medizinische Informatik, Statistik und Epidemiologie, Universität Leipzig 3) Fakultät Angewandte Computer- und Biowissenschaften, Hochschule Mittweida 4) Lehrstuhl Informatik 5, RWTH Aachen 5) Abteilung Data Science und Künstliche Intelligenz, Fraunhofer FIT 6) Institut für Medizinische Informatik, Universitätsmedizin Göttingen

Hintergrund

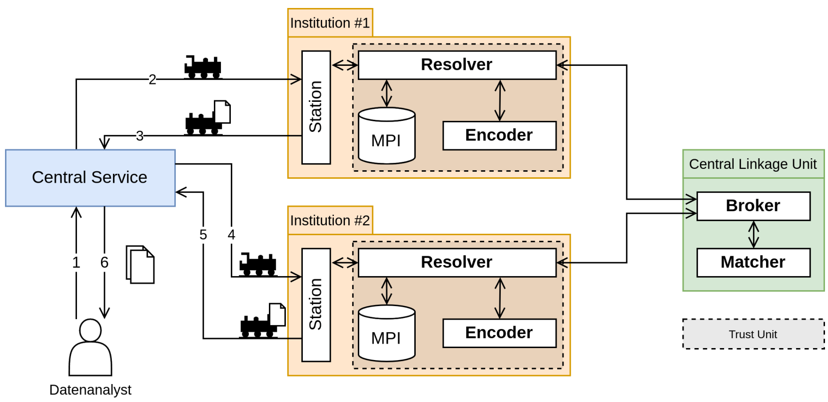


- Verteilte Datensätze und hohe Datenschutzanforderungen
- Personal Health Train (PHT) als Plattform für verteilte Analysen[1]
- Dubletten können statistische Ergebnisse verzerren
- Record Linkage (RL) zur Verbindung von Datenquellen
- **Ziel: Integration von RL-Methoden (PPRL) in den PHT**

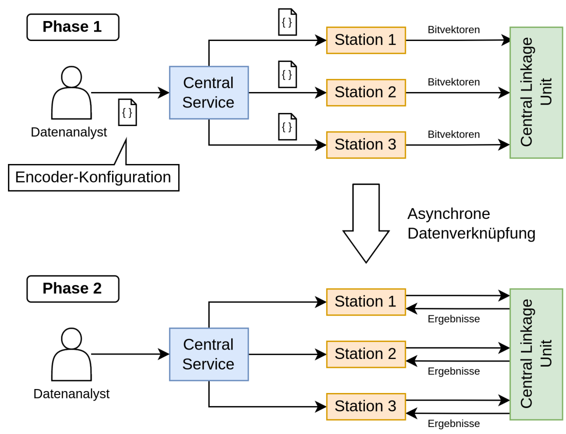
Methoden

- Erweiterung der PHT-Implementierung PADME[2]
- PPRL mit Bloomfilter für Datenverknüpfung[3]
- Durchführung in zwei Zugausführungen im PHT

Stationsdienste	Zentrale Dienste (TTP)
Encoder • Vorverarbeitung von IDATs • Maskierung von IDATs	**Broker** • Annahme von Bitvektoren • Verwaltung von Match-Sessions • Bereitstellung von Ergebnissen
Resolver • Auflösung von Pseudonymen • Übermittlung von Bitvektoren	**Matcher** • Abgleich von Bitvektoren



Architektur, Komponenten und Prozess



Ergebnisse

- Validierung des Verfahrens an zwei Datensätzen
- CORD „Private Set Intersection" → **100 % max. F1-Score**
- Synthetischer Datensatz → **99,7 % max. F1-Score**
  → IDATs aus öffentlichen Datenquellen
  → Typografische Fehler für Fehlertoleranz des Verfahrens

Diskussion

- Kein Unterschied zur zentralisierten Datenverknüpfung
- Anwendung von Strategien zur Behandlung von Dubletten
  → **Entity Consolidation und Duplicate Elimination**
- Sichere Verknüpfung durch Härtung des PPRL-Verfahrens
  → **Keine Rückschlüsse auf IDATs möglich**

Referenzen: [1] O. Beyan u. a., „Distributed Analytics on Sensitive Medical Data: The Personal Health Train", Data Intellegence, Bd. 2, Nr. 1–2, S. 96–107, Jan. 2020, doi: 10.1162/dint_a_00032. [2] S. Welten u. a., „A Privacy-Preserving Distributed Analytics Platform for Health Care Data", Methods Inf Med, Jan. 2022, doi: 10.1055/s-0041-1740564. [3] R. Schnell, T. Bachteler, und J. Reiher, „Privacy-preserving record linkage using Bloom filters", BMC Med Inform Decis Mak, Bd. 9, Nr. 1, S. 41, Aug. 2009, doi: 10.1186/1472-6947-9-41.

HOCHSCHULE MITTWEIDA University of Applied Sciences
nfdi 4 health
MDS Medical Data Science
PADME

Das Konsortium „Smart Medical Information Technology for Healthcare (SMITH)" wird mit Mitteln des Bundesministeriums für Bildung und Forschung (BMBF) unter dem Förderkennzeichen 01ZZ1803A-W gefördert.

MITGLIED DER
TMF

MEDIZIN INFORMATIK INITIATIVE

GEFÖRDERT VOM
Bundesministerium für Bildung und Forschung

SMITH.
Klinische Forschung und Patientenversorgung nachhaltig verbessern.
www.smith.care

Smart Medical Information Technology for Healthcare

Generating Structured Data in the Medical Domain using GANs

Sina Sadeghi, Lars Hempel, Masoud Abedi, Toralf Kirsten

Department for Medical Data Science, Leipzig University Medical Center
Institute for Medical Informatics, Statistics and Epidemiology, Leipzig University
Faculty Applied Computer and Bio Sciences, Mittweida University of Applied Sciences

MDS Medical Data Science

Introduction

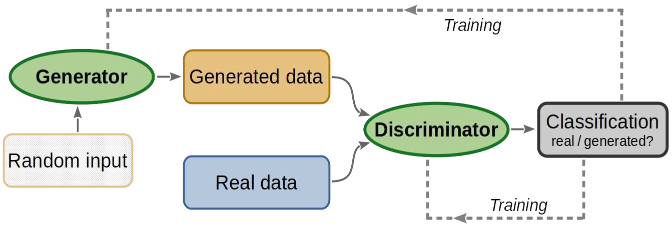
Medical data scarcity:

- Patient privacy (data protection regulations)
- Rare diseases with limited available data

Synthetic data generation:

- Artificial data resembling real data
  - not associated with real patients
  - improve data analysis performance

GANs - Generative Adversarial Networks



- Impressive results in generating images
- Challenging performance on tabular data
  - especially in medical domain with limited data

Methods

GAN variants for generating tabular data:

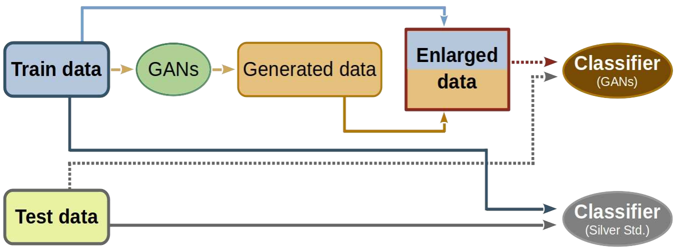
- CGAN - Conditional GANs
- CTGAN - Conditional Tabular GANs
- WGANGP - Wasserstein GANs with Gradient Penalty

BCW – Breast Cancer Wisconsin dataset

- Differentiating into small subsets
  - effect of size of training data on model accuracy
- Regular extension sampling
  - impact of newly available on model performance

Evaluation framework

Extended data consisting of real and synthetic data



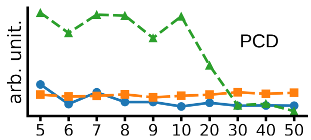
Binary (malignant / benign) classifiers:

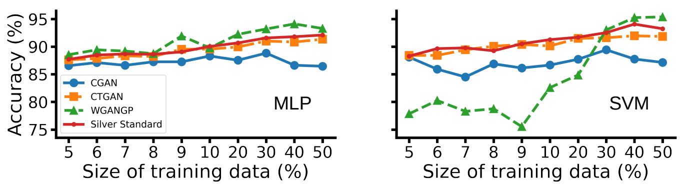
- SVM – Support Vector Machine
- MLP – Multi-Layer Perceptron

Results

PCD – Pairwise Correlation Difference

Correlation among features in real data captured by synthetic data





Improved accuracy by including synthetic data

- compared to Silver Standard (with only real data)
- even with small amount of available data

Other ratios for real and synthetic data in the extended dataset are be considered

Reference

Abedi, M.; Hempel, L.; Sadeghi, S.; Kirsten, T. GAN-Based Approaches for Generating Structured Data in the Medical Domain. Appl. Sci. 2022, 12, 7075.

Das Konsortium „Smart Medical Information Technology for Healthcare (SMITH)“ wird mit Mitteln des Bundesministeriums für Bildung und Forschung (BMBF) unter dem Förderkennzeichen 01ZZ1803A-W gefördert.

MITGLIED DER TMF

MEDIZIN INFORMATIK INITIATIVE

GEFÖRDERT VOM Bundesministerium für Bildung und Forschung

SMITH.
Klinische Forschung und Patientenversorgung nachhaltig verbessern.
www.smith.care

SMITH
Smart Medical Information Technology for Healthcare

An Image Processing Pipeline to Detect Potential Leukodystrophy Patients

Masoud Abedi[1,2,3] Navid Shekarchizadeh[1,2] Sina Sadeghi[1,2] Lars Hempel[1,2,3]
Christa-Caroline Bergner[4] Julia Lier[4] Wolfgang Köhler[4] Toralf Kirsten[1,2,3]
on behalf of the LEUKO-Expert Project Team

1) Department of Medical Data Science, Leipzig University Medical Center
2) Institute for Medical Informatics, Statistics and Epidemiology, Leipzig University
3) Faculty of Applied Computer and Bio Sciences, Mittweida University of Applied Sciences
4) Department of Neurology, University of Leipzig Medical Center

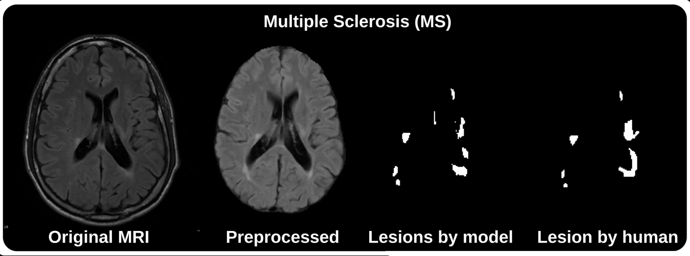
MDS Medical Data Science

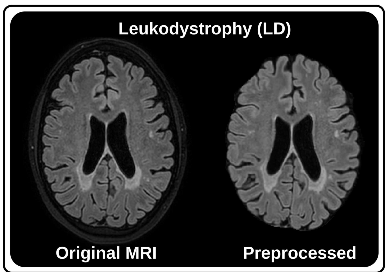
Introduction & Methods

- Leukodystrophies (LD):
  - A family of neurodegenerative diseases
    - ✓ Image-based diagnosis is crucial
  - Rare (1 in 40,000 people worldwide)
    - ✓ Synthetic data generation is necessary
  - MRI data available in 3 hospitals in Germany
    - ✓ Distributed & privacy-preserving data analysis
- Multi-site image processing pipeline:
  - Diagnostic support, education, and training
  - Personal Health Train (PHT) infrastructure [1]
  - White matter lesion segmentation modeling
    - ✓ Multiple Sclerosis (MS) is very similar to LD
    - ✓ Annotated data from ISBI challenge 2015 [2]
    - ✓ Preprocessing
    - ✓ Deep convolutional neural network

Results

- Promising performance of the model on MS
- Preprocessing MRI of LD (data not annotated)



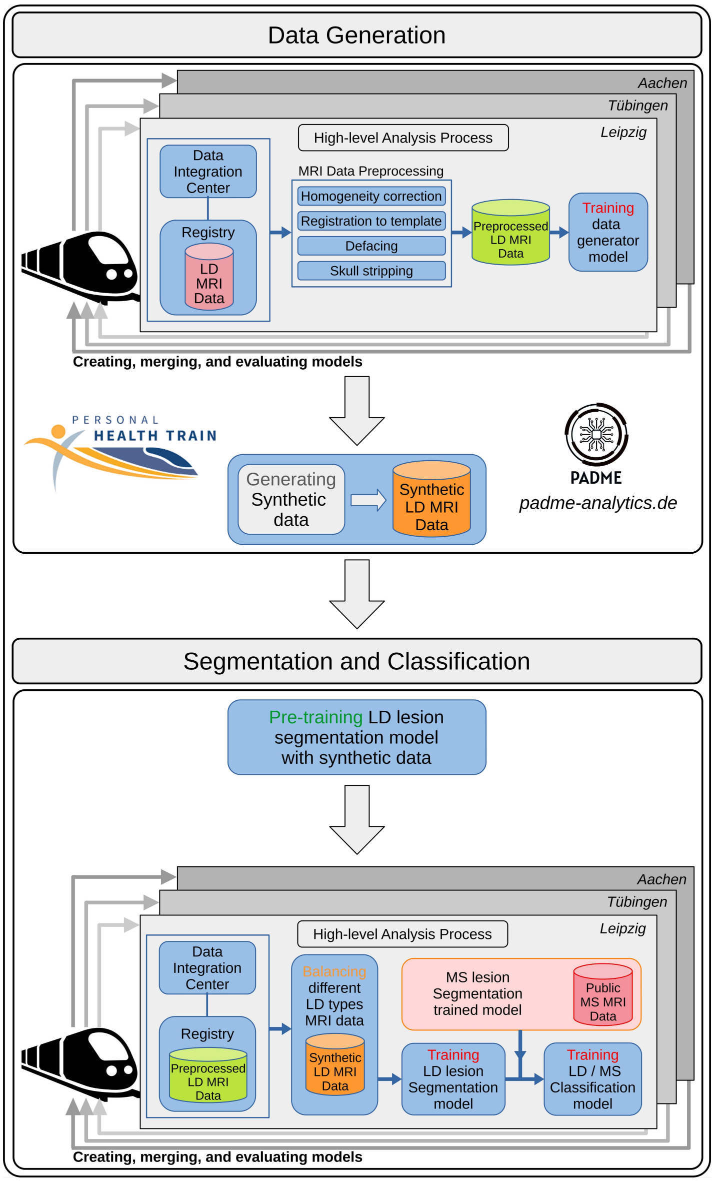


Metric	Value
Lesion false positive rate	0.29
Dice coefficient	0.67
Positive predictive value	0.71
Volume difference	0.32

Future work

- Transforming the model from MS to LD
- LD MRI data harmonization

Multi-Site Image Processing Pipeline



References

[1] Welten, S., et al. "A Privacy-Preserving Distributed Analytics Platform for Health Care Data." Methods Inf. Med. (2022).
[2] Carass, A., et al. "Longitudinal multiple sclerosis lesion segmentation: resource and challenge." NeuroImage 148 (2017): 77-102

Das Konsortium „Smart Medical Information Technology for Healthcare (SMITH)“ wird mit Mitteln des Bundesministeriums für Bildung und Forschung (BMBF) unter dem Förderkennzeichen 01ZZ1803A-W gefördert.

MITGLIED DER TMF

MEDIZIN INFORMATIK INITIATIVE

GEFÖRDERT VOM Bundesministerium für Bildung und Forschung

SMITH.
Klinische Forschung und Patientenversorgung nachhaltig verbessern.
www.smith.care

Smart Medical Information Technology for Healthcare

Vorhersage der Verweildauer auf einer ITS

Lars Hempel[1,3], Sven Bercker[2], Samira Zeynalova[3], Sina Sadeghi[1,3], Ulrike Klotz[1,3,4], Toralf Kirsten[1,3,4]

1) Abteilung Medical Data Science, Medizininformatikzentrum Leipzig
2) Klinik und Poliklinik für Anästesiologie und Intensivtherapie, Universitätsklinikum Leipzig
3) Institut für medizinische Informatik, Statistik und Epidemilogie, Universität Leipzig
4) Fakultät Angewandte Computer- und Biowissenschaften, Hochschule Mittweida

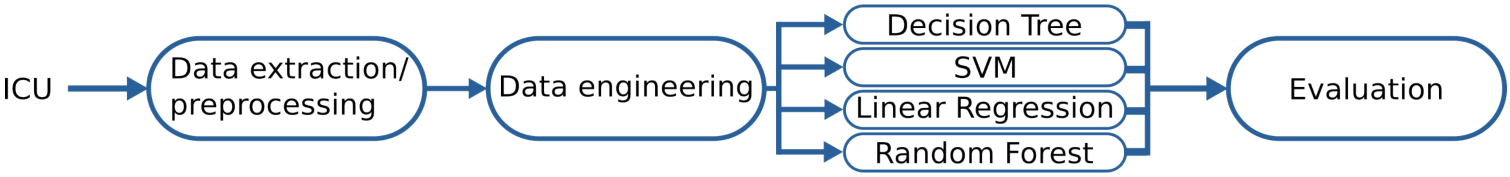
MDS Medical Data Science

Motivation

- Ressourcenmanagement als Indikator für Effizienz und Kostenverwaltung
- Aufenthaltsdauer eines Patienten (LoS) für die Bewertung des Managements und Pflege
- Aufenthaltsdauer abhängig von Patienten, dessen Symptomatik und Kondition
- **Ziel: Modellbasierte Prädiktion der LoS**

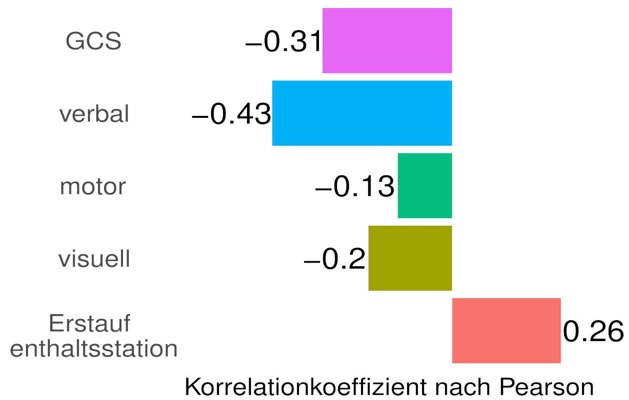
Methoden

- Datenquelle: ITS-Daten aus MIMIC IV
- Datenqualität: Konservative Kuration auf Basis von Literatur und med. Experten
- Methoden zur Prädiktion: Decision Tree, SVM, Linear Regression und Random Forest
- Evaluationsmetriken: RMSD, MAD, MAPE, R^2



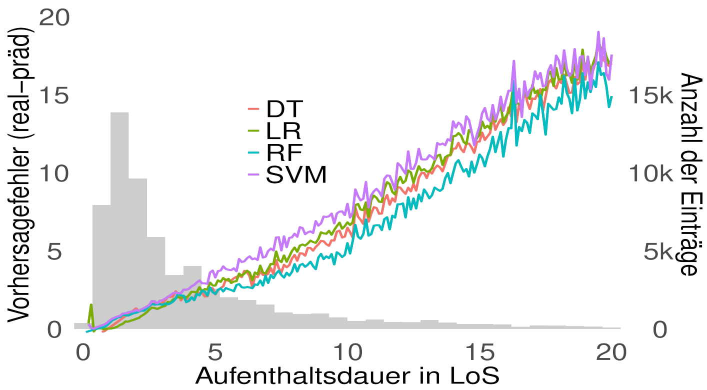
Ergebnisse

- Beste Modellprädiktion: **Random Forest**
- RMSD: 2,88 und R^2: 0,41
- Kohortengröße: 24 148, genutze Parameter: 20
- Wichtigste Einflussgrößen: Glasgow Coma Scale (GCS) und Erstaufenthaltsstation
- Viele Laborparameter mit geringem Einfluss
- Größerer Prädiktionsfehler bei steigenden LoS



Diskussion

- Ansatz: Prädiktion auf der Basis der Erstanamnese mit akzeptablen Ergebnissen
- Prädiktionfehler bei großen LoS auf Grund von kleinen Fallzahlen und ggfs. ungeplanten Ergebnissen
- Zukünftige Ausrichtung: Kontinuierliche Adaption der LoS Prädiktion unter Beachtung longitudinaler Diagnostik



Referenzen

1) Bacchi, S. et al. (2022), Machine learning in the prediction of medical inpatient length of stay. Intern Med J, 52: 176-185. https://doi.org/10.1111/imj.14962
2) *Rocheteau,E. et al. (2021)* Temporal pointwise convolutional networks for length of stay prediction in the intensive care unit. In Proceedings of the Conference on Health, Inference, and Learning (CHIL '21). Association for Computing Machinery, New York, NY, USA, 58–68. https://doi.org/10.1145/3450439.3451860

Das Konsortium „Smart Medical Information Technology for Healthcare (SMITH)“ wird mit Mitteln des Bundesministeriums für Bildung und Forschung (BMBF) unter dem Förderkennzeichen 01ZZ1803A-W gefördert.

MITGLIED DER TMF

MEDIZIN INFORMATIK INITIATIVE

GEFÖRDERT VOM Bundesministerium für Bildung und Forschung

UNIVERSITÄT LEIPZIG
Medizinische Fakultät

Universitätsklinikum Carl Gustav Carus
DIE DRESDNER.

Universitätsklinikum Leipzig
Medizin ist unsere Berufung.

Retinal layer thickness as a biomarker for monitoring the progression of diseases of the central nervous system – Medical Data Science using OCT and MRI imaging

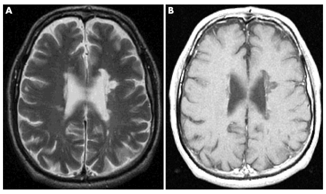
S.Sadeghi[1], T. Kirsten[1,2], S. Klumpe[3], M. Eulitz[3], A. Dillenseger[3], C. Beste[4], F. Ziemssen[5], T. Ziemssen[3], F.G.Rauscher[1,2]

1) Medical Data Science, Universitätsklinikum Leipzig, Leipzig, 2) Institut für Medizinische Informatik, Statistik und Epidemiologie, Universität Leipzig, Leipzig, 3) Zentrum für klinische Neurowissenschaften, Klinik und Poliklinik für Neurologie, Technische Universität Dresden, Dresden, 4) Klinik und Poliklinik für Kinder- und Jugendpsychiatrie und Jugendpsychotherapie, Technische Universität Dresden, Dresden, 5) Augenklinik, Universitätsklinikum Leipzig, Leipzig

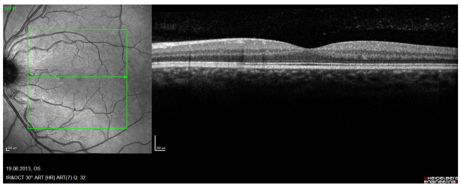
Background

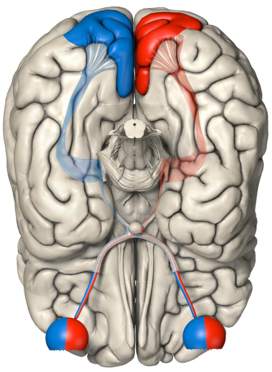
The eye is a window to the brain. There is promising evidence of correlations between retinal layer thickness from optical coherence tomography (OCT) and magnetic resonance imaging (MRI) markers.
→ This results in the potential for science-based clinical application of OCT and MRI in combination for diagnosis and follow-up of various diseases of the central nervous system.

- In the present study, retinal structures (measured with OCT) are compared with intracerebral structures and pathologies from available MRI examinations. In addition, their correlations to cognitive functions will also be investigated.
- For this purpose, existing OCT, MRI and data from neuropsychological and cognitive examinations (partly under EEG) of patients of the ZKN and test persons of the ZKN and KJP will be used (retrospective study design).



Axial MRI of a 46 year old man with secondary progressive MS: large left sided periventricular lesion A.) hyperintense with T2 weighted imaging and B.) hypointense with T1 weighted imaging





Methods

complete dataset: 316,681clinical images from existing documentation of clinical routine for 2,662 individuals (including data and information collected during clinical routine in connection with the OCT and/or MRI examinations)

removal of duplicates

study population: 85,167 clinical images for 2,659 patients of the ZKN as well as for subjects who received OCT and/or MRI examinations at the ZKN and KJP of the University Hospital Carl Gustav Carus Dresden, Germany

exclusion of 4,293 records of 1,047 patients with non-standardised image settings
(excluded scans belong to 9 patients who additionally also possess all standardised scans (=records not lost); or to 1,032 patients with only some of the required imaging;
and further to 6 patients not included in the standardised sample).

analysis subset: 64,925 clinical images for 1,659 patients imported into the analysis pipeline

- Cross-sectional and longitudinal analyses for three key OCT scan types of the Glaucoma Module by Spectralis spectral-domain Optical Coherence Tomography (SD-OCT), Heidelberg Engineering, Heidelberg, Germany
- Scan settings were:

Macular volume scan with horizontal scan location [PPoleH(APS)] in parallel to the disk-fovea-axis. 61 B-scans with a distance of 120µm between scans inside a 30° by 25° area. High-speed mode with 768 x 496 pixel.	Macular volume scan with vertical scan location [PPoleV(APS)] perpendicular to the disk-fovea-axis. 19 B-scans with a distance of 240µm between scans inside a 30° by 15° area. High-resolution mode with 1536 x 496 pixel.	Circumpapillary ring scan around the optic nerve head [ONH-RC(APS)] comprising of three concentric ring scans with 3.5mm, 4.1mm, and 4.7mm diameter (accompanied by 24 line scans). High-speed mode with 768 x 496 pixel

→ These three scan patterns were matched by regular volume and circular scans of the same resolution settings to increase the number of analyzed patients, when possible.

Results

Standardised SD-OCT imaging at the Multiple Sclerosis Centre Dresden, Centre of Clinical Neuroscience, Department of Neurology, Carl Gustav Carus Faculty of Medicine, Technical University Dresden, Germany was analyzed for fixed scan-settings.

- Horizontal volume scans with 61 B-Scans, vertical volume scans with 19 B-Scans and one version of circumpapillary Retinal Nerve Fibre Layer scan (either ONH-RC(APS) or one circular B-scan, sometimes combined with 24 radial scans) were present in combination in 1,659 patients for 1,640 right eyes and 1,600 left eyes.
- Of those, 1,379 (right eye) and 1,347 (left eye) presented with the ideal imaging by the Glaucoma Module (PPoleH(APS), PPoleV(APS) and ONH-RC(APS)). For 422 right eyes and 408 left eyes comparable scan combinations were obtained.
- In total, 64,925 records of a final sample of 1,659 patients were imported into the analysis pipeline. This clinical dataset consists of routine visits of 1,630 patients (60,233 records) as well as 53 specific study patients with 4,702 records (7% of records).

	Image Type	B-Scans
Set 1	OCT Volume Scan	61
Set 2	OCT Circle Scan	1
Set 3	OCT Volume Scan	19
Set 4	OCT Radial Scan	24
Set 5	OCT Radial Circle Scan	27

	Exam Date freq
Group 1	# <= 2
Group 2	2 < # <= 4
Group 3	4 < # <= 6
Group 4	6 < # <= 8
Group 5	8 < # <= 10
Group 6	10 < #

Legend:
R1: number of eyes with full set of records for Set 1, Set 3 and Set 5;
L1: number of eyes with full set of records for Set 1, Set 3 and Set 5

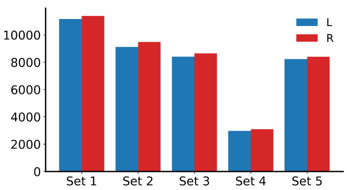


Figure 1 Frequency of desired image settings; Legend: R: right eye; L: left eye

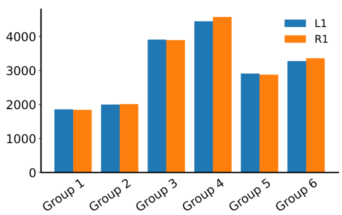
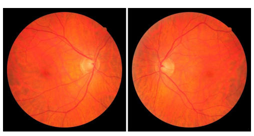


Figure 2 Frequency of examinations with the combination of desired settings: per eye

MDS
Medical Data Science



Contact: Dr. Sina Sadeghi : Sina.Sadeghi@medizin.uni-leipzig.de

Discussion

At ZKN, in combination with joint studies by ZKN and KJP, 80,874 images were obtained in the aforementioned volume and circular scan settings, Set 1, Set 3, and Set 5 or comparable (right eye: 40,985; left eye: 39,889)). This points to a very high level of standardisation of imaging in this clinical setting. This unprecedented dataset is of high value to the research questions of KZN and KJP and will serve as the ideal basis for cross-sectional and longitudinal analyses of diseases of the central nervous systems. A substantial number of patients had standardized imaging by the Spectralis Glaucoma Module, for the desired scan settings, 64,925 records of 1,659 patients were available. Inside this sample, 53 specific study IDs were counted, some of which could be the topic of a separate analysis.

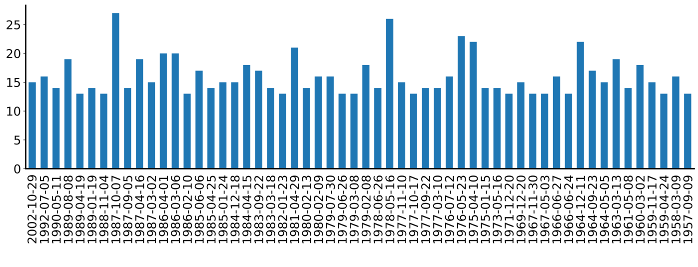
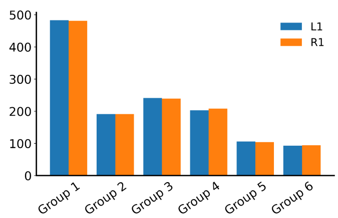


Figure 3 Frequency of exams per patient in Group 6, sorted by birthdate for longitudinal analysis;

Outlook: The 1,659 patients with available 64,925 OCT records will be stratified into clinical research questions to focus on cross-sectional and longitudinal analyses of diseases of the central nervous system, this will include data from MRI and other clinical information. The research will centre on Multiple Sclerosis (MS), with investigations on disease progression, and therapy, with a specific focus on relapsing and remitting MS.



Legend:
R1: number of patients with full set of records for Set 1, Set 3 and Set 5;
L1: number of patients with full Set of records for Set 1, Set 3 and Set 5

Figure 4 Number of patients with frequency of examinations with the combination of desired settings

UNIVERSITÄT LEIPZIG
Medizinische Fakultät

Universitätsklinikum Leipzig
Medizin ist unsere Berufung.

Medical Data Science: Interactive Analysis of Circumpapillary Retinal Nerve Fiber Layer Thickness

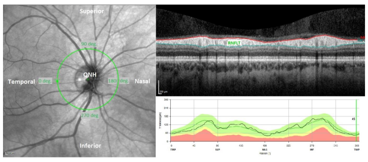
T. Peschel[1], T. Kirsten[2,3,1], M. Wang[4], T. Elze[3,4], F.G. Rauscher[1,3]

1) Institute for Medical Informatics, Statistics, and Epidemiology, Leipzig University, Leipzig, Germany; 2) Institute for Medical Data Science, Leipzig University Medical Center, Leipzig, Germany; 3) Leipzig Research Centre for Civilization Diseases (LIFE), Leipzig University, Leipzig, Germany; 4) Schepens Eye Research Institute, Harvard Medical School, Boston, Massachusetts, USA.

Motivation

- Early diagnosis of eye disease is of increased importance in an ageing population
- Normative data serve as the backbone for comparison with individual cases
- Such norms need to be accessible directly in clinics to support ophthalmological care
- Ocular optical coherence tomography (OCT) manufacturers aim to mark abnormalities in patient measurements based on their own, typically proprietary norms based on small samples

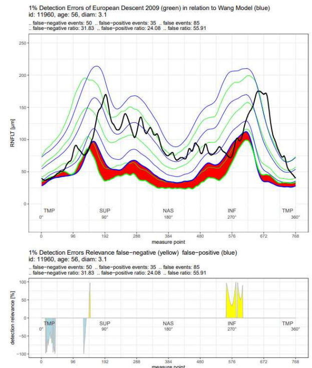
- Several different visualizations (e.g. sectors and continous plots) of normative data may help to improve decision-making for clinicians.

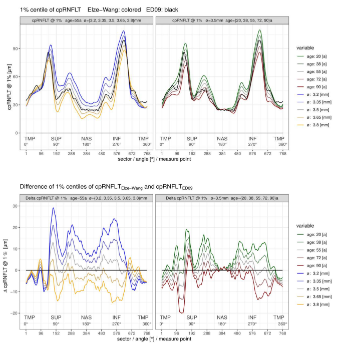


- Circumpapillary Retinal Nerve Fiber Layer Thickness (cpRNFLT) was measured by 768 A-scans on a 360° circle around the optic nerve head of the human eye
- Sectors are commonly used in the clinical context
- Sectors consist of 45° and 90° sections of the scan

Methods

- Normative data calculated from 5637 healthy eyes
- OCT-measurements of the LIFE-Adult-Study population
- Analysing detection errors as a way to judge the improvements of the model by Wang et al. in comparison to the so called „European Descent 2009“ (ED09) model used in the OCT device
- Counting false-positive and false-negative detection events assuming the higher accuracy of Elze-Wang's cpRNFLT model

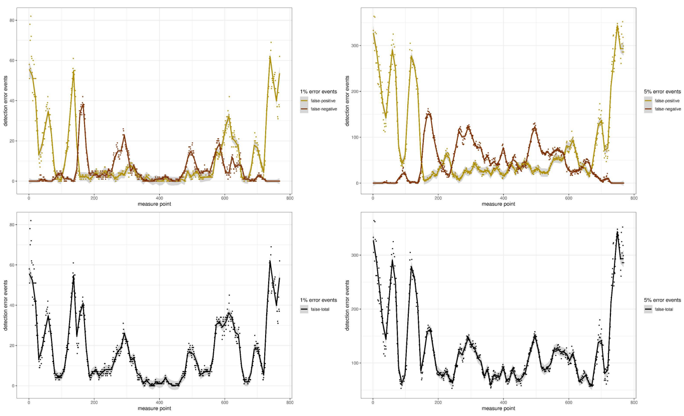




- cpRNFLT 768 measurement points are modelled by a Gaussian distribution
- cpRNFLT-Sectors are modelled by an averaged Gaussian distribution
- cpRNFLT(D) 768 measurement points are modelled by a Box-Cox-T distribution
- cpRNFLT(D)-Sectors-points are modelled by an averaged Box-Cox-T distribution
- Novel cpRNFLT-norms by Wang et al. are modelled as linear dependent on measure diameter and age

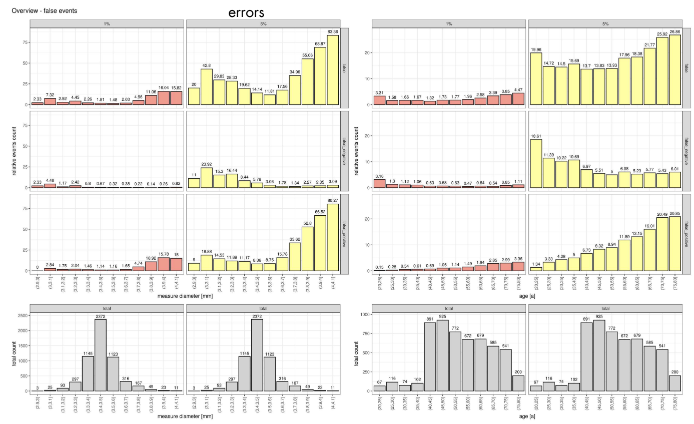
Results - Discussion - Outlook

- Influencing factors of the model (e.g. age, measurement diameter) need to be accounted for
- Adjusting for such factors in the model is of direct clinical significance for cpRNFLT as a diagnostic tool
- Dependency of the model on age and diameter are reflected by false-positive and false-negative events
- Thus the RNFLT(D)-Visualizer represents an additional value for clinical decision making in Ophthalmology



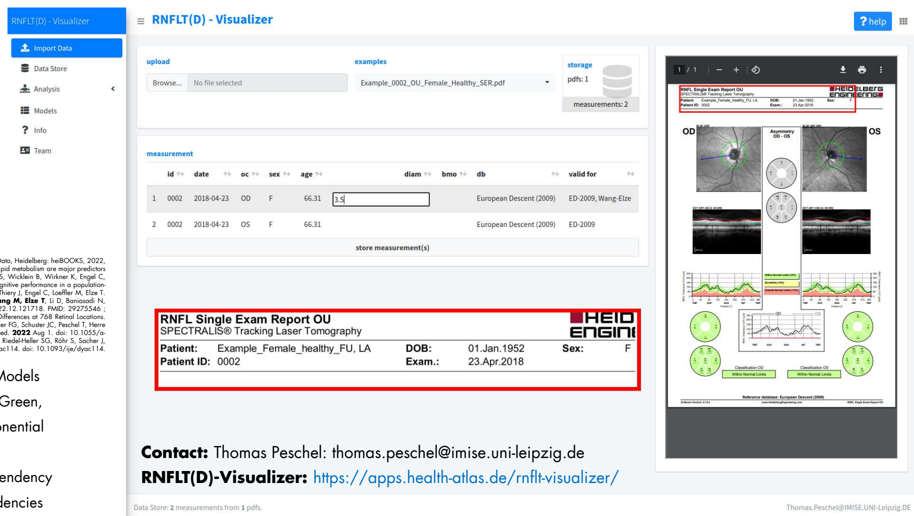
- Analysis of relative frequencies of false events for the 1st (red) and the 5th (yellow) percentile with reference to age and measure diameter

- Large diameters and high ages lead to small thicknesses
- Consequently, the rate of false positive errors produced by the built-in ED09 model increases
- Small diameters and low ages are more likely to produce false negative errors



Successful practical implementation as the RNFLT(D)-Visualizer Shiny-Web-App as part of the Leipzig Health Atlas:

- Opportunity to upload clinical data to enable analysis of this individual level data on the new Wang et al. norms, taking into account age and measurement diameter for the first time
- Importing and loading pdf documents of cpRNFLT measurements and its meta data
- Scanning of OCT print-out of patient measurement plots and read out meta data
- Various presentations and analyses of patient data with device implemented norms and norms from models by Wang et al. via continuous 360° circle plots or sector plots
- Opportunity to download the result graphs and tables in several common formats



References:

Peschel T, Wang M, Kirsten T, Rauscher FG, Elze E. A Cloud-based Infrastructure for Interactive Analysis of RNFLT Data, in: Heuveline, Vincent und Bisheh, Nina (Hrsg.): E-Science-Tage **2021**: Share Your Research Data, Heidelberg: heiBOOKS, 2022, p. 54-68. https://doi.org/10.11588/heibooks.979.c13718 ; **Rauscher FG, Wang M**, Francke M, Wirkner K, Tönjes A, Engel C, Thiery J, Stenvinkel P, Stumvoll M, Loeffler M, Elze T, Ebert T. Renal function and lipid metabolism are major predictors of circumpapillary retinal nerve fiber layer thickness-the LIFE-Adult Study. BMC Med. **2021** Sep 7;19(1):202. doi: 10.1186/s12916-021-02064-8. PMID: 34488766 ; **Girbardt J**, Luck T, Kynast J, Rodriguez FS, Wicklein B, Wirkner K, Engel C, Girbardt C, Wang M, Polyakova M, Witte AV, Loeffler M, Villringer A, Riedel-Heller SG, Schroeter ML, Elze T, Rauscher FG. Reading cognition from the eyes: association of retinal nerve fibre layer thickness with cognitive performance in a population-based study. Brain Commun. **2021** Nov 8;3(4):fcab258. doi: 10.1093/braincomms/fcab258. eCollection 2021. PMID: 35350710. ; **Li D, Rauscher FG**, Choi EY, Wang M, Baniasadi N, Wirkner K, Kirsten T, Thiery J, Engel C, Loeffler M, Elze T. Sex-Specific Differences in Circumpapillary Retinal Nerve Fiber Layer Thickness.. Ophthalmology. **2020** Mar;127(3):357-368. doi: 10.1016/j.ophtha.2019.09.019. Epub 2019 Sep 25. PMID: 31732228 ; **Wang M, Elze T**, Li D, Baniasadi N, Wirkner K, Kirsten T, Thiery J, Loeffler M, Engel C, Rauscher FG. Age, ocular magnification, and circumpapillary retinal nerve fiber layer thickness. J Biomed Opt. **2017** Dec;22(12):1-19. doi: 10.1117/1.JBO.22.12.121718. PMID: 29275546 ; **Baniasadi N, Rauscher FG**, Li D, Wang M, Choi EY, Wang H, Peschel T, Wirkner K, Kirsten T, Thiery J, Engel C, Loeffler M, Elze T. Norms of Interocular Circumpapillary Retinal Nerve Fiber Layer Thickness Differences at 768 Retinal Locations. Transl Vis Sci Technol. **2020** Aug 12;9(9):23. doi: 10.1167/tvst.9.9.23. eCollection 2020 Aug. PMID: 32879779. ; **Kirsten T**, Meineke F, Löffler-Wirth H, Uciteli A, Beger C, Stäubert S, Löbe M, Hänsel R, Rauscher FG, Schuster JC, Peschel T, Herre H, Wagner J, Zachariae S, Engel C, Scholz M, Rahm E, Binder H, Löffler M. The Leipzig Health Atlas - An open platform to present, archive and share bio-medical data, analyses and models online. Methods Inf Med. **2022** Aug 1. doi: 10.1055/a-1914-1985. Epub ahead of print. PMID: 35915977. ; **Engel C, Wirkner K, Zeynalova S**, Baber R, Binder H, Ceglarek U, Enzenbach C, Fuchs M, Hagendorff A, Henger S, Hinz A, Rauscher FG, Reusche M, Riedel-Heller SG, Röhr S, Sacher J, Sander C, Schroeter ML, Tarnok A, Treudler R, Villringer A, Wachter R, Witte AV, Thiery J, Scholz M, Loeffler M; LIFE-Adult-Study working group. Cohort Profile: The LIFE-Adult-Study. Int J Epidemiol. **2022** May 28:dyac114. doi: 10.1093/ije/dyac114. Epub ahead of print. PMID: 35640047.

MDS Medical Data Science

Outlook: Improving Models

- Box-Cox-T, Box-Cox-Green, Box-Cox-Power-Exponential
- sex dependency
- ocular laterality dependency
- higher order dependencies

Contact: Thomas Peschel: thomas.peschel@imise.uni-leipzig.de

RNFLT(D)-Visualizer: https://apps.health-atlas.de/rnflt-visualizer/

Retinal layer thickness as a biomarker for monitoring the progression of diseases of the central nervous system – Medical Data Science using OCT and MRI imaging

S. Klumpe[1], S.Heimann[1],T. Kirsten[2,3], M. Eulitz[1], A. Dillenseger[1], C. Beste[4], F. Ziemssen[5], T. Ziemssen[1], F.G.Rauscher[2,3]

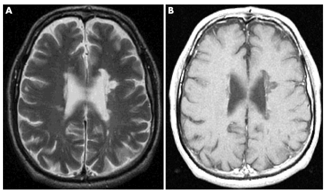
1) Zentrum für klinische Neurowissenschaften, Klinik und Poliklinik für Neurologie, Technische Universität Dresden, Dresden, 2) Institut für Medizinische Informatik, Statistik und Epidemiologie, Universität Leipzig, Leipzig, 3) Medical Data Science, Universitätsklinikum Leipzig, Leipzig, 4) Klinik und Poliklinik für Kinder- und Jugendpsychiatrie und Jugendpsychotherapie, Technische Universität Dresden, Dresden, 5) Augenklinik, Universitätsklinikum Leipzig, Leipzig

Background

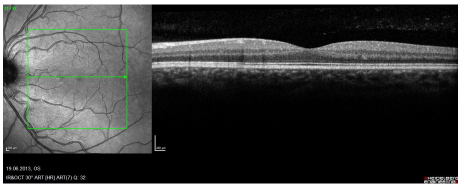
The eye is a window to the brain. There is promising evidence of correlations between retinal layer thickness from optical coherence tomography (OCT) and magnetic resonance imaging (MRI) markers.

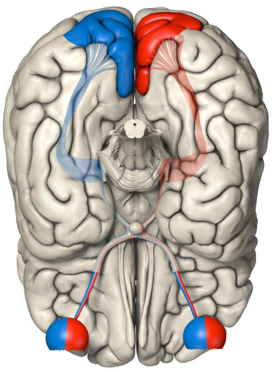
→ This results in the potential for science-based clinical application of OCT and MRI in combination for diagnosis and follow-up of various diseases of the central nervous system.

- In the present study, retinal structures (measured with OCT) are compared with intracerebral structures and pathologies from available MRI examinations. In addition, their correlations to cognitive functions will also be investigated.
- For this purpose, existing OCT, MRI and data from neuropsychological and cognitive examinations (partly under EEG) of patients of the ZKN and test persons of the ZKN and KJP will be used (retrospective study design).



Axial MRI of a 46 year old man with secondary progressive MS: large left sided periventricular lesion A.) hyperintense with T2 weighted imaging and B.) hypointense with T1 weighted imaging





Methods

complete dataset: 316,681clinical images from existing documentation of clinical routine for 2,662 individuals (including data and information collected during clinical routine in connection with the OCT and/or MRI examinations)

removal of duplicates

study population: 85,167 clinical images for 2,659 patients of the ZKN as well as for subjects who received OCT and/or MRI examinations at the ZKN and KJP of the University Hospital Carl Gustav Carus Dresden, Germany

exclusion of 4,293 records of 1,047 patients with non-standardised image settings (excluded scans belong to 9 patients who additionally also possess all standardised scans (=records not lost); or to 1,032 patients with only some of the required imaging; and further to 6 patients not included in the standardised sample).

analysis subset: 64,925 clinical images for 1,659 patients imported into the analysis pipeline

- Cross-sectional and longitudinal analyses for three key OCT scan types of the Glaucoma Module by Spectralis spectral-domain Optical Coherence Tomography (SD-OCT), Heidelberg Engineering, Heidelberg, Germany
- Scan settings were:

Macular volume scan with horizontal scan location [PPoleH(APS)] in parallel to the disk-fovea-axis. 61 B-scans with a distance of 120µm between scans inside a 30° by 25° area. High-speed mode with 768 x 496 pixel.	Macular volume scan with vertical scan location [PPoleV(APS)] perpendicular to the disk-fovea-axis. 19 B-scans with a distance of 240µm between scans inside a 30° by 15° area. High-resolution mode with 1536 x 496 pixel.	Circumpapillary ring scan around the optic nerve head [ONH-RC(APS)] comprising of three concentric ring scans with 3.5mm, 4.1mm, and 4.7mm diameter (accompanied by 24 line scans). High-speed mode with 768 x 496 pixel

→ These three scan patterns were matched by regular volume and circular scans of the same resolution settings to increase the number of analyzed patients, when possible.

Results

Standardised SD-OCT imaging at the Multiple Sclerosis Centre Dresden, Centre of Clinical Neuroscience, Department of Neurology, Carl Gustav Carus Faculty of Medicine, Technical University Dresden, Germany was analyzed for fixed scan-settings.

- Horizontal volume scans with 61 B-Scans, vertical volume scans with 19 B-Scans and one version of circumpapillary Retinal Nerve Fibre Layer scan (either ONH-RC(APS) or one circular B-scan, sometimes combined with 24 radial scans) were present in combination in 1,659 patients for 1,640 right eyes and 1,600 left eyes.
- Of those, 1,379 (right eye) and 1,347 (left eye) presented with the ideal imaging by the Glaucoma Module (PPoleH(APS), PPoleV(APS) and ONH-RC(APS)). For 422 right eyes and 408 left eyes comparable scan combinations were obtained.
- In total, 64,925 records of a final sample of 1,659 patients were imported into the analysis pipeline. This clinical dataset consists of routine visits of 1,630 patients (60,233 records) as well as 53 specific study patients with 4,702 records (7% of records).

	Image Type	B-Scans
Set 1	OCT Volume Scan	61
Set 2	OCT Circle Scan	1
Set 3	OCT Volume Scan	19
Set 4	OCT Radial Scan	24
Set 5	OCT Radial Circle Scan	27

	Exam Date freq
Group 1	# <= 2
Group 2	2 < # <= 4
Group 3	4 < # <= 6
Group 4	6 < # <= 8
Group 5	8 < # <= 10
Group 6	10 < #

Legend:
R1: number of eyes with full set of records for Set 1, Set 3 and Set 5;
L1: number of eyes with full set of records for Set 1, Set 3 and Set 5

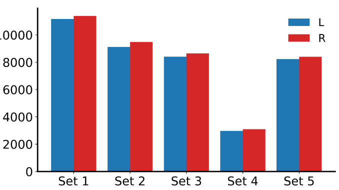


Figure 1 Frequency of desired image settings; Legend: R: right eye; L: left eye

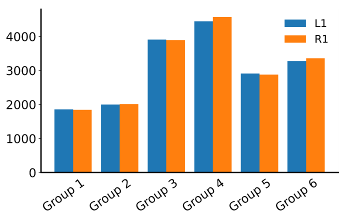
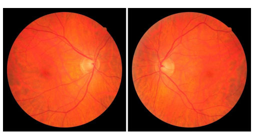


Figure 2 Frequency of examinations with the combination of desired settings: per eye



Contact: Professor Dr. med. Tjalf Ziemssen: Tjalf.Ziemssen@ukdd.de

Discussion

At ZKN, in combination with joint studies by ZKN and KJP, 80,874 images were obtained in the aforementioned volume and circular scan settings, Set 1, Set 3, and Set 5 or comparable (right eye: 40,985; left eye: 39,889)). This points to a very high level of standardisation of imaging in this clinical setting. This unprecedented dataset is of high value to the research questions of KZN and KJP and will serve as the ideal basis for cross-sectional and longitudinal analyses of diseases of the central nervous systems. A substantial number of patients had standardized imaging by the Spectralis Glaucoma Module, for the desired scan settings, 64,925 records of 1,659 patients were available. Inside this sample, 53 specific study IDs were counted, some of which could be the topic of a separate analysis.

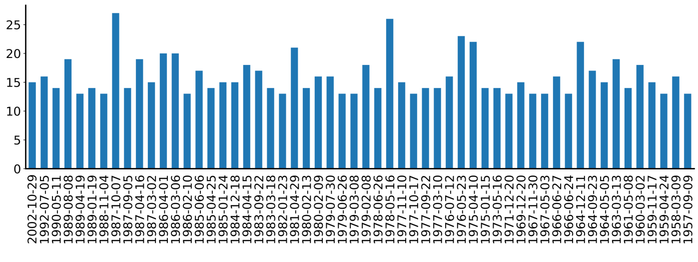
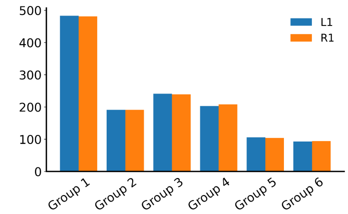


Figure 3 Frequency of exams per patient in Group 6, sorted by birthdate for longitudinal analysis;

Outlook: The 1,659 patients with available 64,925 OCT records will be stratified into clinical research questions to focus on cross-sectional and longitudinal analyses of diseases of the central nervous system, this will include data from MRI and other clinical information. The research will centre on Multiple Sclerosis (MS), with investigations on disease progression, and therapy, with a specific focus on relapsing and remitting MS.



Legend:
R1: number of patients with full set of records for Set 1, Set 3 and Set 5;
L1: number of patients with full Set of records for Set 1, Set 3 and Set 5

Figure 4 Number of patients with frequency of examinations with the combination of desired settings

UNIVERSITÄT LEIPZIG
Medizinische Fakultät

Universitätsklinikum Leipzig
Medizin ist unsere Berufung.

Photos & Impressions

Leipzig, 02.12.2022

MDS
Medical Data Science


Veranstaltung
“Medical Data Science – Linking Medicine and Data Sciences”
Prof. Dr. Toralf Kirsten
02.12.2022 14.30 - 15.30 Uhr Hörsaal, Liebigstraße 12
OUR EXPERTISE — OVERVIEW
Infrastructure
Enhance data analysis in a distributed mode
Software applications: patient-specific guiding and diagnostic support
Methods
Artificial-Intelligence-based methods for
Early detection, prediction, and diagnosis
Identification and segmentation of abnormalities in images
Data augmentation, especially for Rare Diseases
Privacy-preserving methods for data sharing
MDS
Medical Data Science
Privatsphärenschützende Datenverknüpfung in verteilten Analysen mit dem PHT

RESEARCH PROFILE

- Semantic integration of medical data
- Privacy-preserving distributed analysis of medical data
- Applications of novel artificial intelligence methods in medicine and healthcare
- Generation of synthetic data for medicine and healthcare applications
- Current application domains:
  - Leukodystrophies (rare diseases)
  - Ophthalmology
  - Diabetes

We are open to new projects, topics, and ideas in the field of medical data science. Please do not hesitate to contact us for more information if you are interested in collaborating with us or have any questions.

MEDICAL INFORMATICS CENTER

Department for Medical Data Science
Härtelstraße 16 – 18, Haus V
04107 Leipzig, Germany
Phone: +49 (0)341 9716159
E-Mail: toralf.kirsten@medizin.uni-leipzig.de

TRAVEL DIRECTIONS (BUILDING V)

Public Transport:
H (Station ÖPNV):
- Münzgasse, LVZ: Tram 10, 11, 16

by Car:
- Härtelstraße 16-18, 04107 Leipzig

Car Parking:
- Gaudigplatz

Visit our website

Imprint

Editor: University of Leipzig Medical Center
Editorial staff: Prof. Toralf Kirsten
Conception: Department for Medical Data Science
Photo credit: Sina Sadeghi

Revised: August 2022

Universitätsklinikum Leipzig
Medizin ist unsere Berufung.

MDS
Medical Data Science

MEDICAL INFORMATICS CENTER

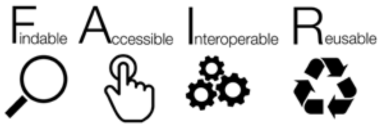
UNIVERSITÄT LEIPZIG
Faculty of Medicine

WE INTRODUCE OURSELVES

We are an interdisciplinary team of Medical Data Science at Leipzig University Medical Informatics Center. We conduct innovative research at the intersection of medical informatics and patient care using artificial intelligence and medical data science to increase benefits for patients, the healthcare system, and medical research. Our focus is twofold: First, we develop a secure computing infrastructure for the preparation, processing and analysis of medical data. This improves the accessibility of such data for research purposes, as medical data are often available from multiple locations, for example in distributed computing. Second, we perform medical data analysis and modeling to advance medical research and also support clinicians. For this, we employ both classical statistical methods and state-of-the-art machine and deep learning techniques.

FAIR DATA

Knowledge derived from captured and collected data has become increasingly important in research and development. In order to reduce difficulties in data sharing in this respect and maximize the value from data, data and metadata should comply with the FAIR principles - Findable, Accessible, Interoperable, and Reusable. To this end, we develop data structures and standardize the process for accessing the data to ensure that the basic requirements for reproducibility are met. We are part of the National Research Data Infrastructure (NFDI4Health), which is an initiative in this context.



INFRASTRUCTURES AND ALGORITHMS FOR DISTRIBUTED ANALYTICS

Personal medical data are often collected and available in multiple locations. Due to privacy concerns and the potentially large volume of medical records, the pooling concept for centralized analysis may not work. As part of the international Personal Health Train (PHT) network, we are working with our collaborators to develop an infrastructure for secure distributed computing. Such PHT infrastructure adheres to the principle of "bring-the-analysis-to-the-data". We have successfully implemented this infrastructure in several research projects within the medical field.

AI-BASED MODELING OF LEUKODYSTROPHIES

Leukodystrophies are a family of rare diseases with an estimated incidence of less than 1:40,000 worldwide. Patients with leukodystrophy typically experience an Odyssey and undergo a challenging, expensive and time-consuming journey. An important issue is that the disease is commonly misdiagnosed and therefore incorrectly treated. Furthermore, since only a few cases are known, there is a lack of validated treatment guidelines. In close collaboration with the Myelin Center at the Leipzig University Medical Center and other interdisciplinary international consortia, we frequently make use of a variety of artificial intelligence methods and analyze data from leukodystrophy patients to support earlier diagnosis, for example.

ABOUT TORALF KIRSTEN

Prof. Toralf Kirsten heads the Medical Data Science department. With his broad computer science background, which ranges from the conception and implementation of IT infrastructures to methods and algorithms of data analysis, he is involved with the department in various national and international research projects. He is actively involved in the Medical Informatics Initiative (MII), the Network University Medicine (NUM), the National Research Data Infrastructure (NFDI) and the international GOFAIR Initiative, which interacts closely with the European Open Science Cloud (EOSC).