Challenges and Risks of Digitalization and Machine Learning in Clinics
Potential Conflicts of Interest

• **Employment or Leadership Positions**
  German Center for Infection Research, University Hospital of Frankfurt, University of Cologne

• **Consultation**
  Astellas, Basilea, Gilead, Merck/MSD, Roche Diagnostics

• **Research Grants**
  Astellas Pharma, Basilea, German Center for Infection Research, German Federal Ministry of Education and Research, Medical University of Hannover, Merck / MSD Gilead, Pfizer

• **Speaker Fees and Honoraries**
  Ärztekammer (Medical Association) Nordrhein, Akademie für Infektionsmedizin, Astellas Pharma, Back Bay Strategies, Basilea, German Center for Infection Research, German Society of Infectious Diseases, German Society of Internal Medicine, Gilead, Merck / MSD, Onkologisch Pharmazeutischer Fachkongress NZW, Pfizer, Shire, University Hospital of Aachen, University Hospital of Freiburg / Kongress u. Kommunikation GmbH, University of Manchester

• **Ownership, stocks, equities, patents, shared rights licences, etc.**
  None

• **Other Relevant Conflicts of Interest**
  None

• **Immaterial Conflicts of Interest**
  German Society for Hematology and Oncology, German Society for Infectious Diseases, German Society for Epidemiology, German-Speaking Mycological Society
I will not talk about these topics (but we can discuss them)

• Heterogeneity in clinical data systems across Europe
• F.A.I.R. principle challenges
  • Findable
  • Accessible
  • Interoperable
  • Reusable
• Data privacy & consent challenges
• Inhibitors of data sharing
• Patient reported outcome measure related challenges
• Telemedical consultation challenges
• „Artificial Intelligence“ as a term
Advances in Self-optimizing Algorithms

Chess
- 8x8 fields
- Max. 35 possible moves
- Ca. 60 moves per game
- $10^{47}$ possible combination
- After 3 moves, 42,875 possible positions
- Goal: Conquer enemy king

Go
- 19x19 fields
- Max. 300 possible moves
- Ca. 200 moves per game
- $10^{170}$ possible combinations
- After 3 moves, 8,000,000 possible positions
- Surround enemy territory

Ca. $10^{140}$ years to calculate all possible game courses
(age of the universe = 1.4 * $10^9$ years)
Deep Blue

AlphaGo
1. 30 million Go-moves:
   - Neural network
   - Monte-Carlo Tree Search

2. Reinforcement Learning
   - 30,000,000 games against iterations of itself

Image Recognition:
- Agar Plates
- Cytology
- Histology
- Gram stains
- Radiology
- Skin Cancer

Phenotype Prediction:
- Antimicrobial resistance
- Antiviral resistance
- Chemotherapy resistance

Pattern Recognition:
- Outbreak detection
- Sepsis detection
Challenge 1: Medical Outcomes cannot be Simulated and Quality Outcome Data are Scarce

Example:
- Invasive pulmonary aspergillosis: \(< 10\) cases / year / hospital
- Diagnosis ill defined, treatment based on surrogate markers

- Data points background: \(\sim 500-1,000\)
- Data points per inpatient day: \(\sim 100-300\) (excluding molecular and imaging data)
- Total per disease episode: \(1,900 – 9,400\)
- Degrees of freedom: \(2 - >1,000\)
- Possible case constellations: \(1,1 \times 10^{1,328} – 2,1 \times 10^{6,570}\)
- Available data worldwide: \(< 10,000\)
Challenge 2: Algorithms Have No Sense of Right and Wrong

„Hey, Siri, how is the clinical response rate of *Paecilomyces* spp. to isavuconazole?“

„I wondered about that myself.“
Challenge 2: Algorithms Have No Sense of Right and Wrong

- „Gack gack!“
- „Gack gack!“
- „Gack gack!“
- „Ente!“
Challenge 2: Algorithms Have No Sense of Right and Wrong
Challenge 2: Algorithms Have No Sense of Right and Wrong: Supervised Learning

Data lake → Algorithm

- Treatment Success: 92%
- Treatment Failure: 8%

Good bot!

Challenge 2: Algorithms Have No Sense of Right and Wrong: Actual Decision Making

- Urgency
- Symptomatic Burden
- Anamnestic Details
- Medical History
- Non-medical History

- Drug Label
- Insurance Status
- Social Background
- Estimated Compliance
- Demographic Factors

- Drug Interactions
- Resistance Pressure
- Short-term vs. Long-term Effectiveness
- Short-term vs. Long-term Toxicity
- Patient Preferences

- Estimated Compliance
- Social Background
- Medical History
- Non-medical History
Challenge 2: Algorithms Have No Sense of Right and Wrong: Actual Decision Making

- Overall Survival
- Disease Remission
- Inpatient vs. outpatient
- Sequelae
- Adverse Effects
- Organ Toxicities
- Resistance development
- Cost
- Quality of Life
Challenge 3: We Must Understand (and Check) the Reasoning of Learning Algorithms

Horse-picture from Pascal VOC data set

Lapuschkin S et al., Nature Communications 2019, doi: 10.1038/s41467-019-08987-4
Challenge 3: We Must Understand (and Check) the Reasoning of Learning Algorithms

Training data of pictures with horses vs. other animals

Lapuschkin S et al., Nature Communications 2019, doi: 10.1038/s41467-019-08987-4
Challenge 3: We Must Understand (and Check) the Reasoning of Learning Algorithms

Smart Hans with Wilhelm von Osten
Presentation to audience

Smart Hans "counting" on the board

Prof. Carl Stumpf, Berlin,
Prussian Academy of Science
Example: Machine Learning in Neutropenic Fever

Laboratory data (145)
- Body temperature (19)
- Microbiological findings (23)
- Virological findings (8)
- Radiological findings (3)
- Patient characteristics (8)
- Underlying disease
- Neutropenia (6)
- Medication (9)
- Comorbidities (4)

Value for day 1 to day 4 (febrile neutropenia (FN))
Value for day -4 to day 0 (time before FN)
Minimum/maximum value
Days from minimum/maximum value to day 1
Differences of values

Choice of class of features/categorisation
Missing data handling/imputation
Visualise interaction of features

Febrile neutropenia

Observational period
Admission
Day -4
Day 1
Day 4
Classification adverse outcome

Jakob C, ..., Vehreschild J, ECCMID 2019
Example: Machine Learning in Neutropenic Fever

- 65 selected features
- Internal validation AUC = 0.75
- Out-of-sample validation AUC = 0.68

Jakob C, ..., Vehreschild J, ECCMID 2019
Look, I am 99% sure it’s not *Pseudomonas*! Let’s use ampicillin!

You’re so funny. I agree, but you know what the guideline says. We must use pip/tazo.

…..99% likelihood of no *Pseudomonas*... ...give ampicillin...
Challenge 4: Evaluating and Maintaining Self-Optimizing Algorithms

Doctor
- Adheres to guidelines and lable information
- Is expected not to act primarily on private opinion and wild guesses

Drug
- Undergoes thorough investigation through many stages of research
- Must prove added value for patient health and/or health system
- Cannot be changed after marketing without losing approval

Self-Learning Algorithm
- Is supposed to do whatever it calculates?
- Is supposed to keep learning and changing with what oversight?
- May need to retrain algorithm in changing patient populations?
Clinical Decision Support

www.treatsystems.com
Challenge 5: Self-Learning Algorithms Sacrifice Detail for the Sake of Overall Correctness
Query:
Identify incident treatment naive HIV+ patients with at least five years follow-up

N= 247
N= 340
Challenge 7: Strongly Regulated in the European Union

REGULATION (EU) 2017/745 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL
of 5 April 2017


6.3. Rule 11

Software intended to provide information which is used to take decisions with diagnosis or therapeutic purposes is classified as class IIa, except if such decisions have an impact that may cause:

— death or an irreversible deterioration of a person's state of health, in which case it is in class III; or

— a serious deterioration of a person's state of health or a surgical intervention, in which case it is classified as class IIb.

Software intended to monitor physiological processes is classified as class IIa, except if it is intended for monitoring of vital physiological parameters, where the nature of variations of those parameters is such that it could result in immediate danger to the patient, in which case it is classified as class IIb.

All other software is classified as class I.
Tree of AI Decision Making
Tree of AI Decision Making

- Single Outcome Inference
- Simple Pattern Recognition
Tree of AI Decision Making

Highly specialized inference systems, e.g. insulin dosing, warfarin dosing, hypertension management etc. → Many low hanging fruits taken

- Single Outcome Inference
- Simple Pattern Recognition
Tree of AI Decision Making

- Single Outcome Inference
- Treatment Decisions
- Simple Pattern Recognition
- AI-guided Disease Management
- Complete Diagnostic Pathways
Universal, comprehensive systems with comprehensive assessment of all relevant information and autonomous decision making ➔ multi-billion dollar project
Tree of AI Decision Making

Integration of multiple layers of information and inference systems for individualized risk prediction and communication

- Single Outcome Inference
- Treatment Decisions
- Individual Risk Management
- Simple Pattern Recognition
- Plausibility Checks, Patient Safety
- Complete Diagnostic Pathways
- Al-guided Disease Management