

Overview of Achievements in Computational Medicine at the Sano Centre

M. Bubak, M. Malawski, A. Crimi, A.A. Hamed,
P. Korzeniowski, P. Nowakowski, and J. Sousa

m.bubak@sanoscience.org

CompBioMed, 12-14 September 2023, Garching, DE

Outline



- Funding and partnership
- Main objectives of Sano
- Sano research teams
- Examples of research at Sano
- Sano in numbers
- Summary - collaboration opportunities

Funding and partners

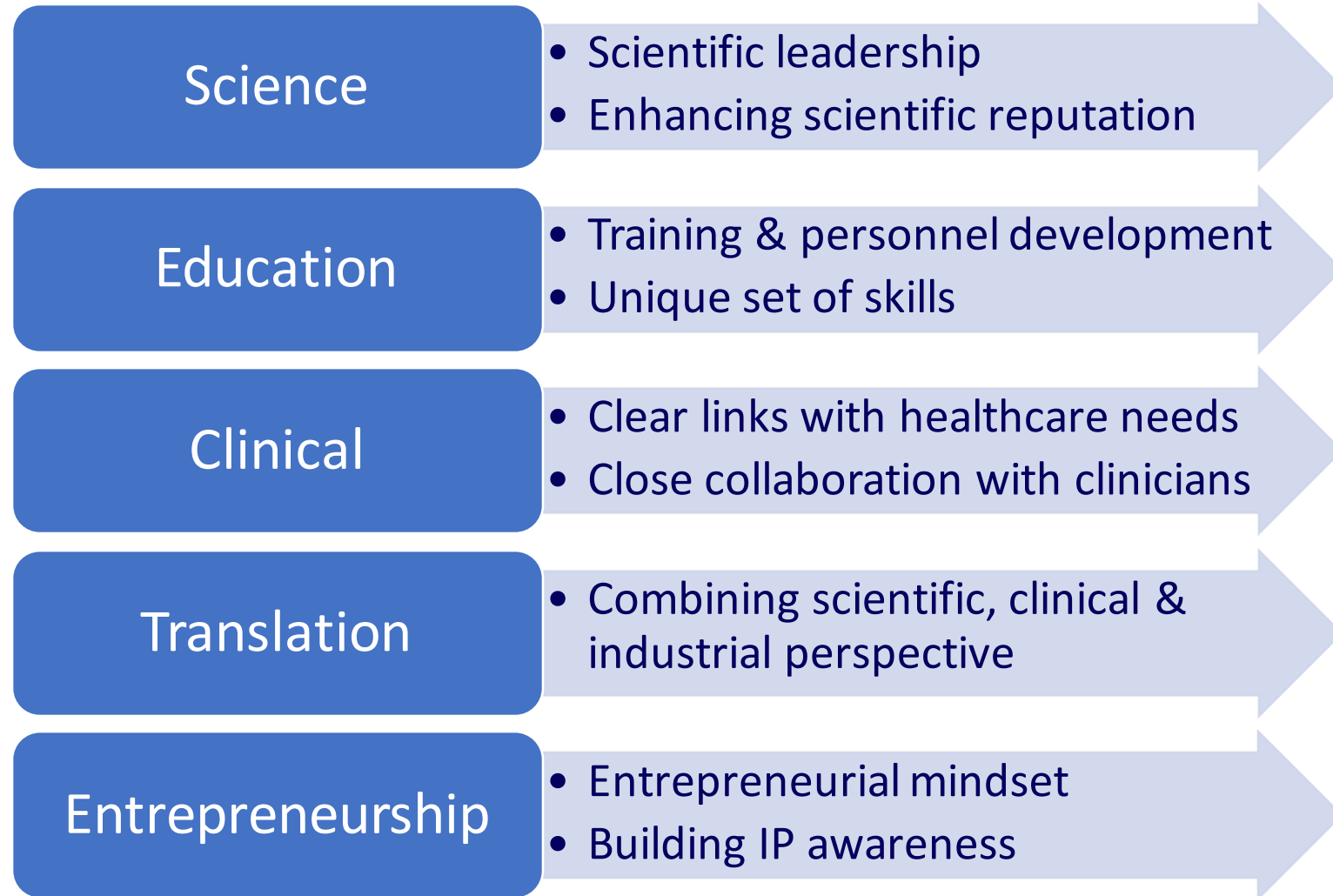


- H2020-WIDESPREAD-2016-2017 TEAMING PHASE 2 programme (grant 857533) (15 M €)
- International Research Agendas Programme (IRAP) of the Foundation for Polish Science, co-funded by the EU in the scope of the European Regional Development Fund (10 M €)
Strategic Partner: University of Sheffield
Applicants: **Marco Viceconti**, University of Bologna (ISC Chair), **Marian Bubak**, AGH University Krakow (Scientific Affairs Director)
Director of Sano and IRAP Part 2 Manager: Arkadiusz Sitek (since 15.02.2021), **Maciej Malawski** (since 1.08.2022)
- Polish Ministry of Science and Higher Education (after 2023, 5 M €)

Teaming Partners:



Main objectives of Sano



Sano Research Teams



Extreme-Scale Data & Computing



Maciej Malawski



Clinical Data



Ahmed Abdeen Hamed



Personal Health



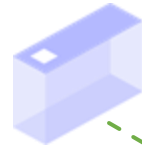
Jose Sousa



Computer Vision



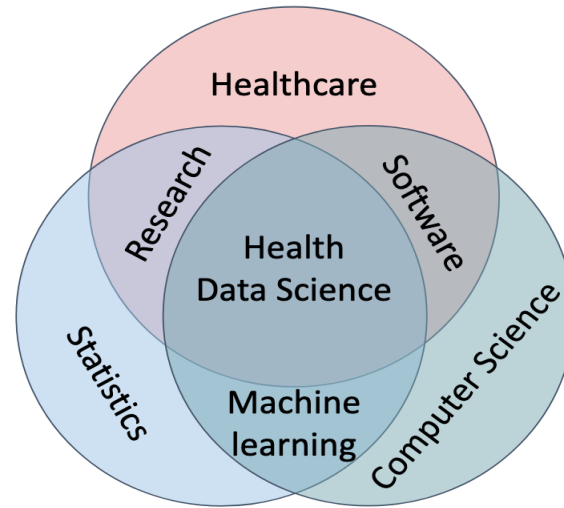
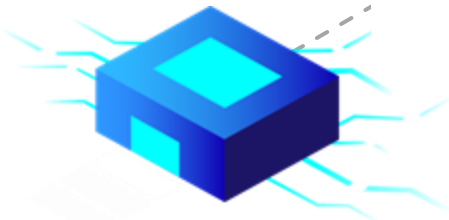
Alessandro Crimi



Health Informatics



Przemysław Korzeniowski



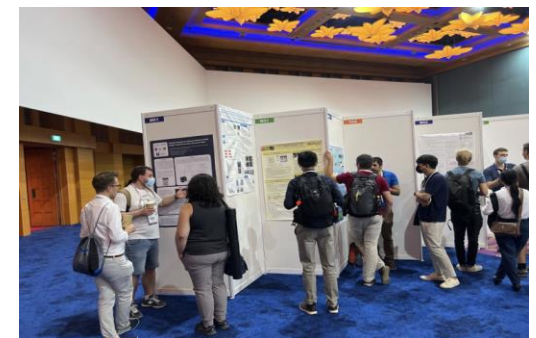
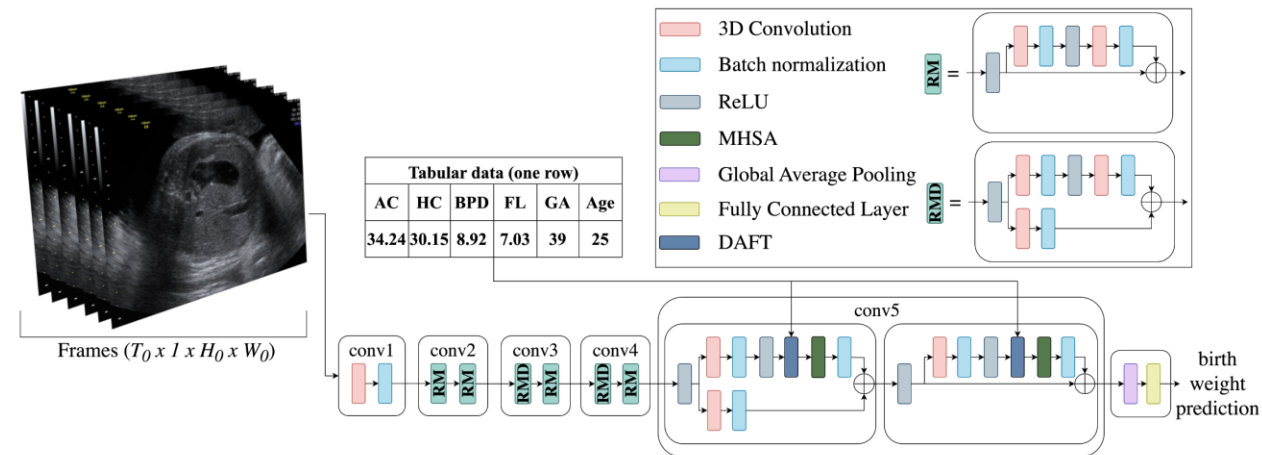
Healthcare Informatics Team

Leader: Przemyslaw Korzeniowski

- ✪ We develop new diagnostic and therapeutic processes based on patient data to extract insights that can be **used to improve patient care and prevention**
- ✪ We work on a new generation of medical communication and incorporation **of artificial intelligence (AI), virtual and augmented reality (VR/AR) and robotics** in existing and new medical workflows

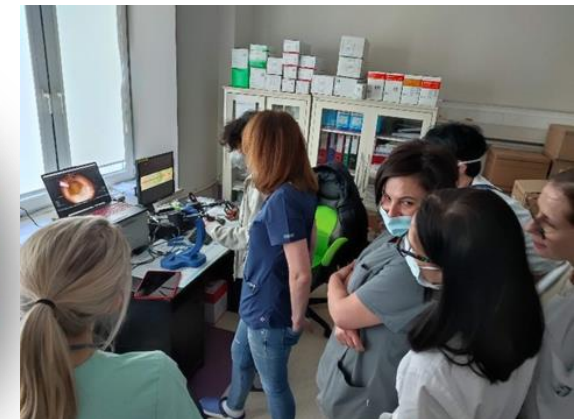
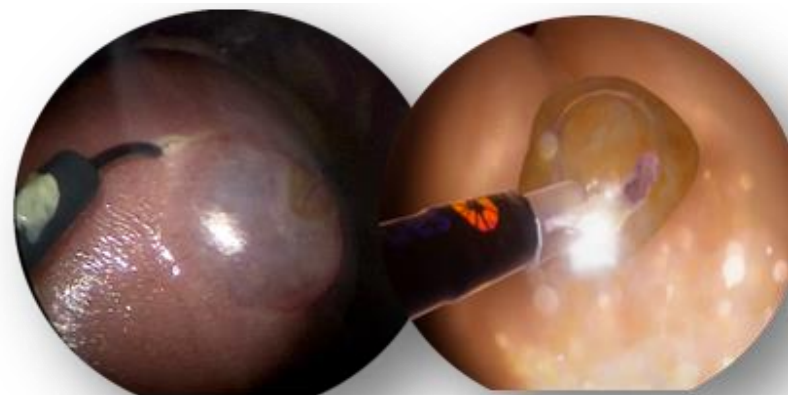
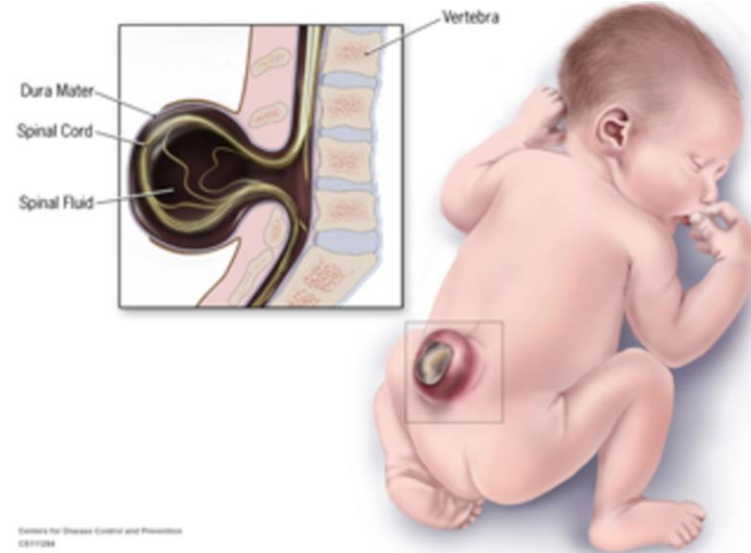
Fetal Birth Weight Prediction using Multimodal Data

- Goal: Accurate prediction of fetal weight at birth is essential for determining the timing and mode of delivery.
- Performing a prenatal ultrasound 24 hours prior to delivery presents challenges as it requires acquiring high-quality images, which becomes difficult during advanced pregnancy due to the lack of amniotic fluid.
- We present a novel method that automatically predicts fetal birth weight by using fetal ultrasound video scans and clinical data.
- Our results show that our method outperforms several state-of-the-art methods and estimates fetal birth weight with an accuracy comparable to human experts and can minimize the risk of errors inherent in manual measurements.
- Next steps: Add more data from multi-centers and deploy them as support tool for clinicians.



Virtual Reality Surgical Simulators

- Goal: Spina Bifida (SB) is a defect during the early stage of pregnancy - an incomplete closing of the spine around the spinal cord. The learning curve is steep and requires excellent procedural skills.
- Computer-based virtual reality (VR) simulation systems offer a safe, cost-effective, and configurable training environment free from ethical and safety issues.
- An initial simulation realism validation study was carried out by obtaining subjective feedback (face and content validity) from 14 clinicians. The overall simulation realism was on average marked 4.07 on a 5-point Likert scale (1 -'very unrealistic', 5 -'very realistic'). Its usefulness as a training tool for SB-repair as well as in learning fundamental laparoscopic skills was marked 4.63 and 4.80, respectively.
- Next steps: to develop the simulator further based on clinicians' feedback and follow it up with wider and more through clinical studies such as construct validity and/or transfer of skills.



Sim2Real with Generative AI

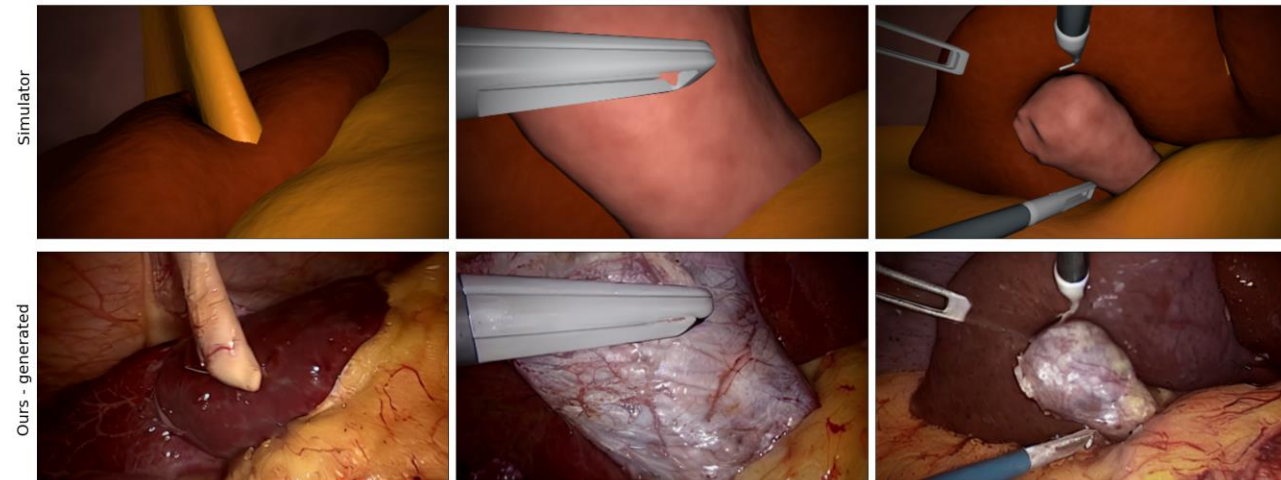
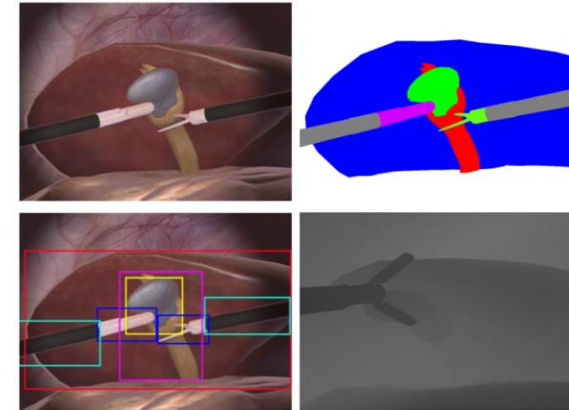
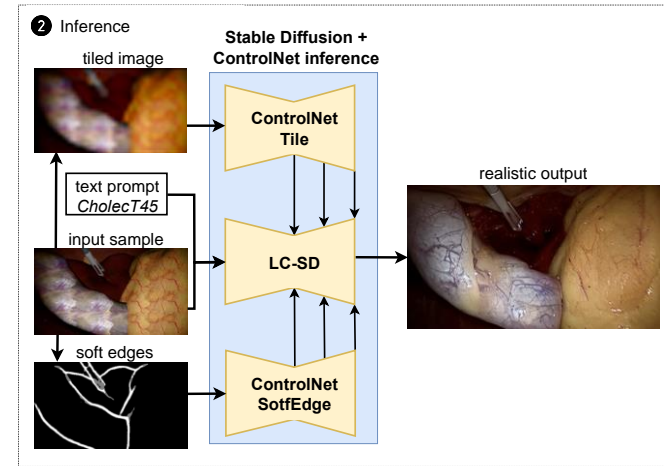
Goal: Developing an unsupervised Image-to-Image translation approach that could generate fully labeled and realistic-looking datasets based on surgical simulators. This is particularly effective with minimal real data requirements.

Research Results:

- Method which uses fewer than 100 actual images of laparoscopic cholecystectomy for training was created and evaluated on simplified abdomen simulator data.
- The approach consists of Stable Diffusion for image generation, ControlNets to maintain the consistency of generated samples, and Dreambooth for a minimal data training scheme. All techniques are novel and not widely employed in the medical Sim2Real field.
- The quality of the generated dataset was tested and compared with a GAN-based baseline.

Possible extensions of this approach:

- apply to a more realistic laparoscopic cholecystectomy simulation, including real tool-tissue interactions and temporal consistency between frames,
- reduce the dependency on input images and exploring multi-region approaches,
- explore lightweight techniques for training.



Computer Vision Team / Brain and More Lab

Leader: Alessandro Crimi

We focus on three main areas:

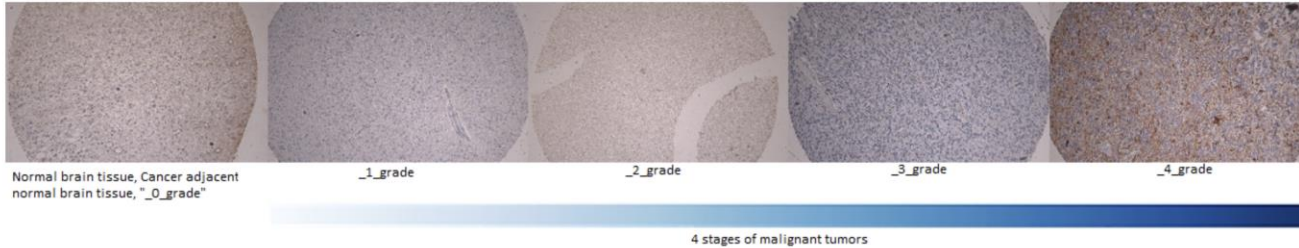
- Medical imaging (from histology/microscopy)
- Medical imaging (Neuroimaging)
- Brain Wearables (EEG, fNIRS, MEG, etc)

We are interested in studies spanning all dimensions of two disease processes (but open to others):

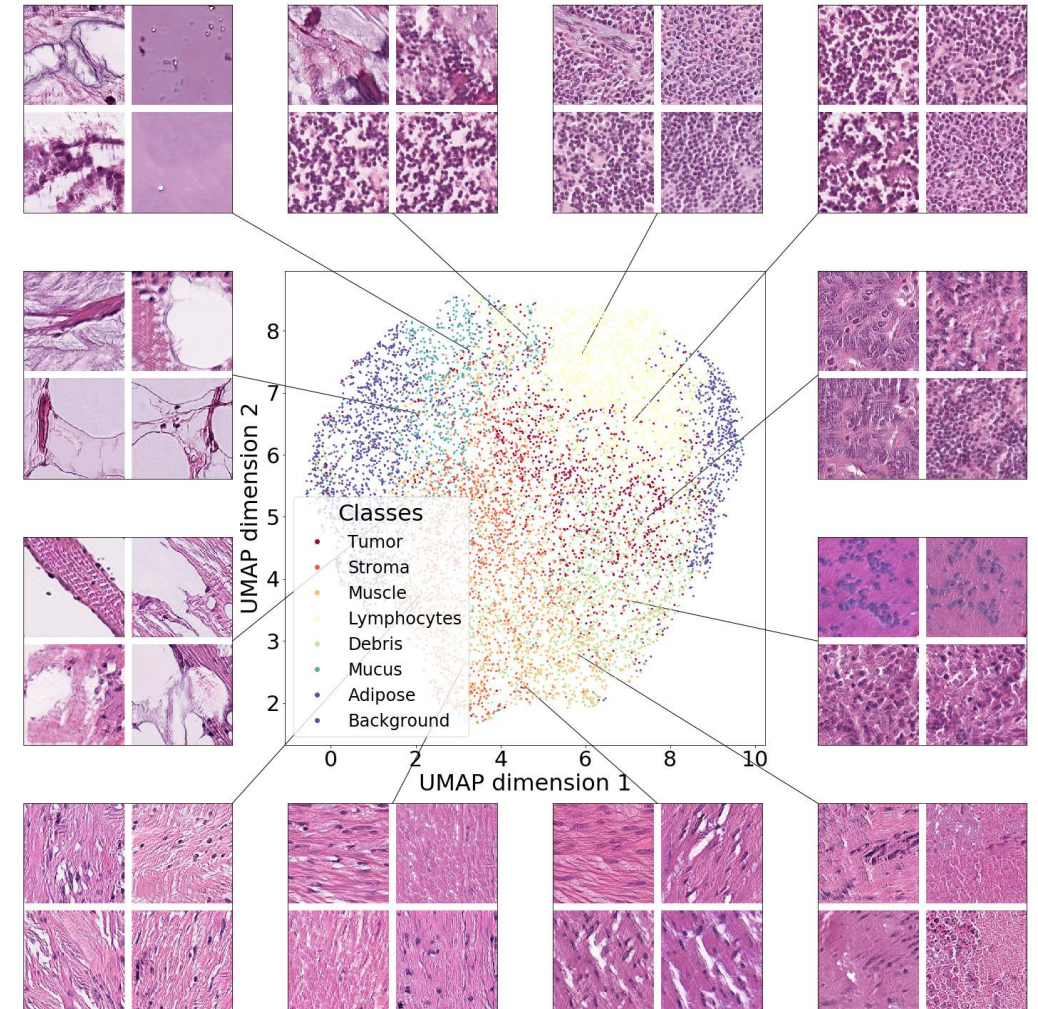
- Alzheimer/Dementia
- Brain tumors

Brain tumor classification and image translation

Monika Pytlarz in collaboration with Nencki Institute



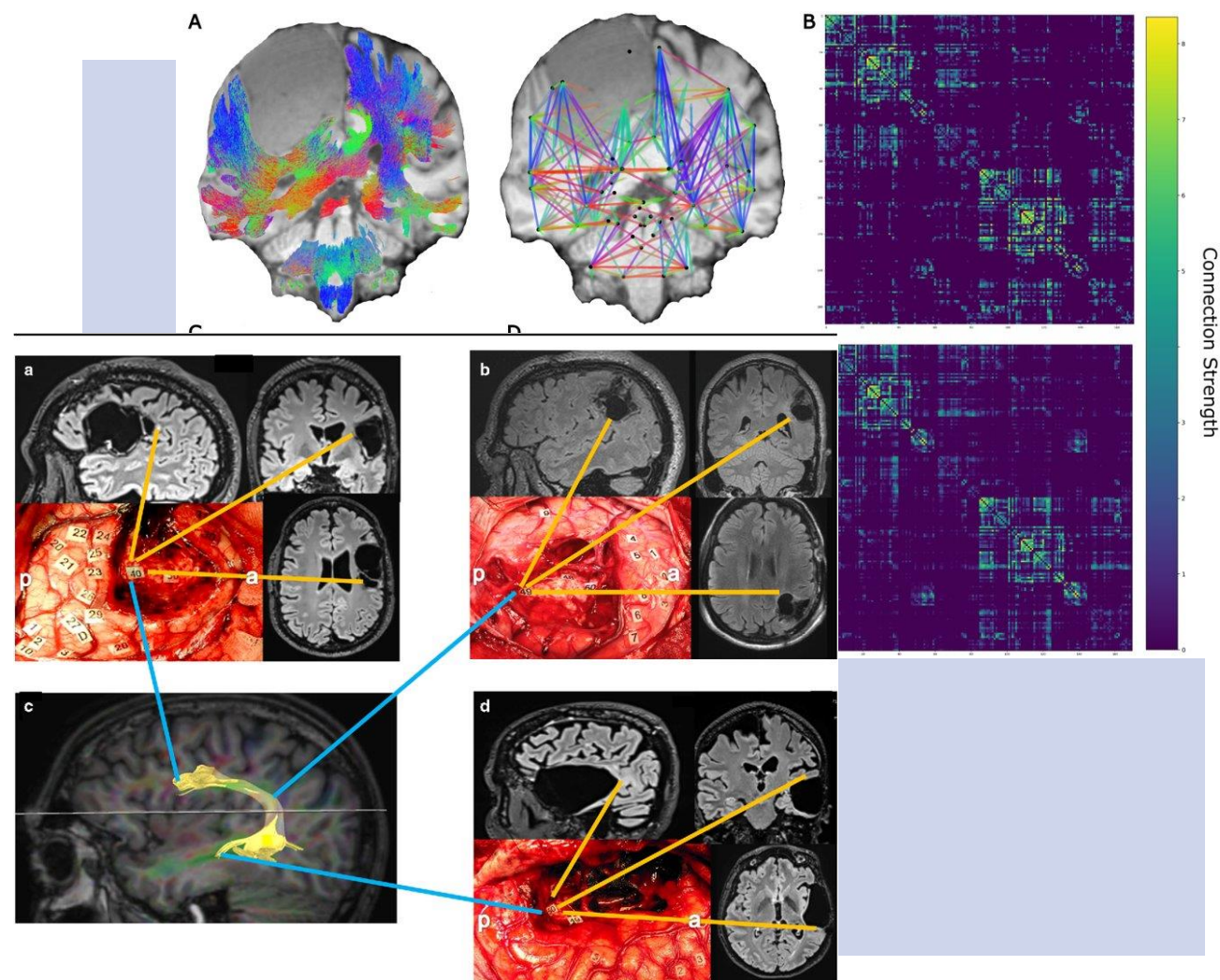
- A classification model is able to detect and classify the grade of a brain tumor within a biopsy
- Not just classification, but explain interaction of clusters of necrotic cells: e.g. in Grade 3 the AI algorithm found more discriminant cells clusters appearing to be more myelinating.



Brain tumor surgical planning and prediction

Joan Falco-Roget, Bartlomiej Ksiazek

- A trained model is able to predict the brain connectivity of patients post-surgery and recovery
- Collaborations: University of Ghent for data, University of Messina and Padua for clinical support.
- Follow-up of the project includes the roll out in clinical practice at the University hospital of Palermo



Bayesian Filtered Generation of Post-surgical Brain Connectomes on Tumor Patients” J. Roget Falco, F. Sambataro, A. Cacciola, A. Crimi MICCAI 2022, Nature Communication

https://link.springer.com/chapter/10.1007/978-3-031-21083-9_8

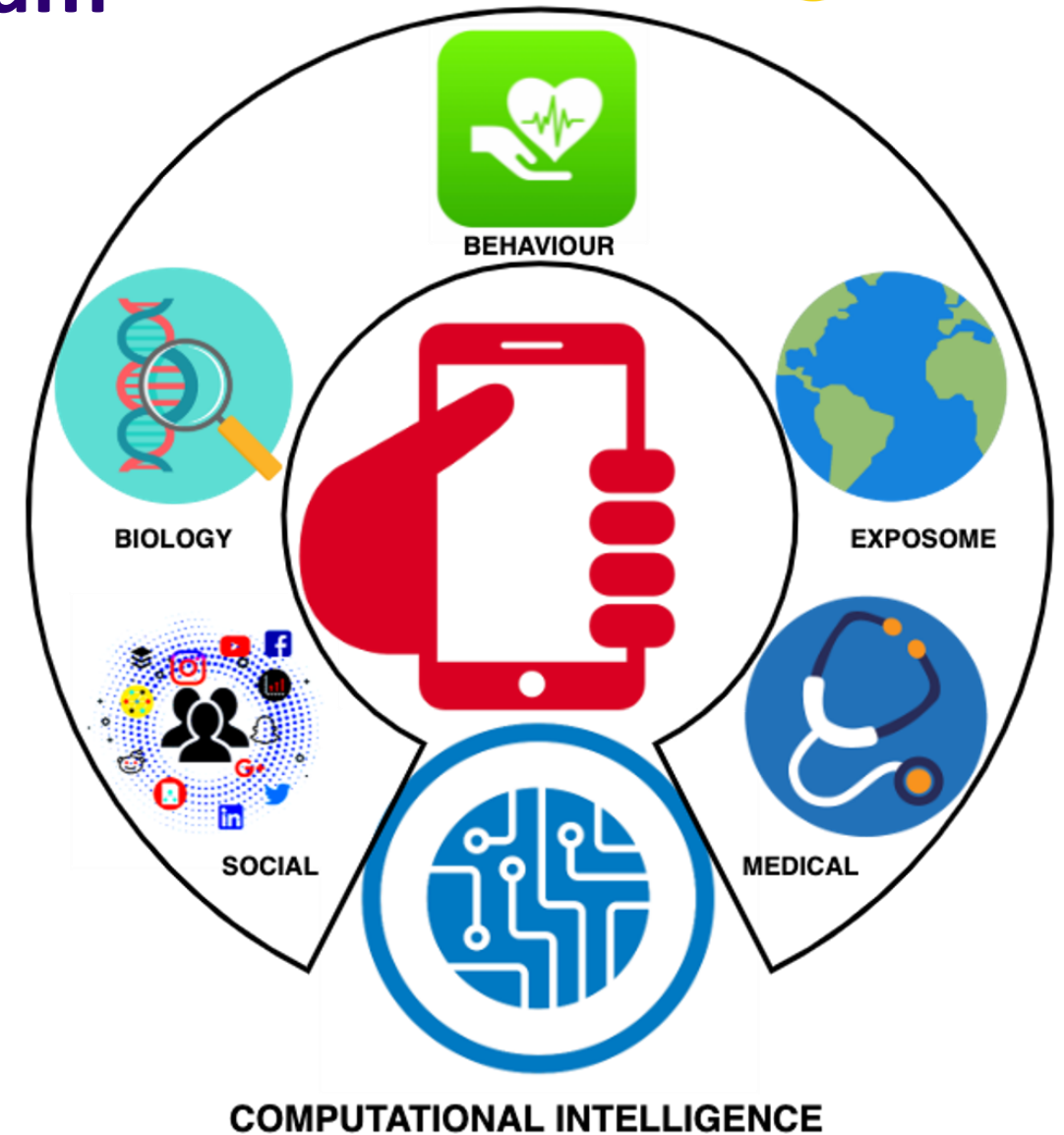
<https://www.biorxiv.org/content/10.1101/2022.11.14.516248v2.abstract>

Personal Health Data Science Team

Leader: Jose Sousa

We focus on:

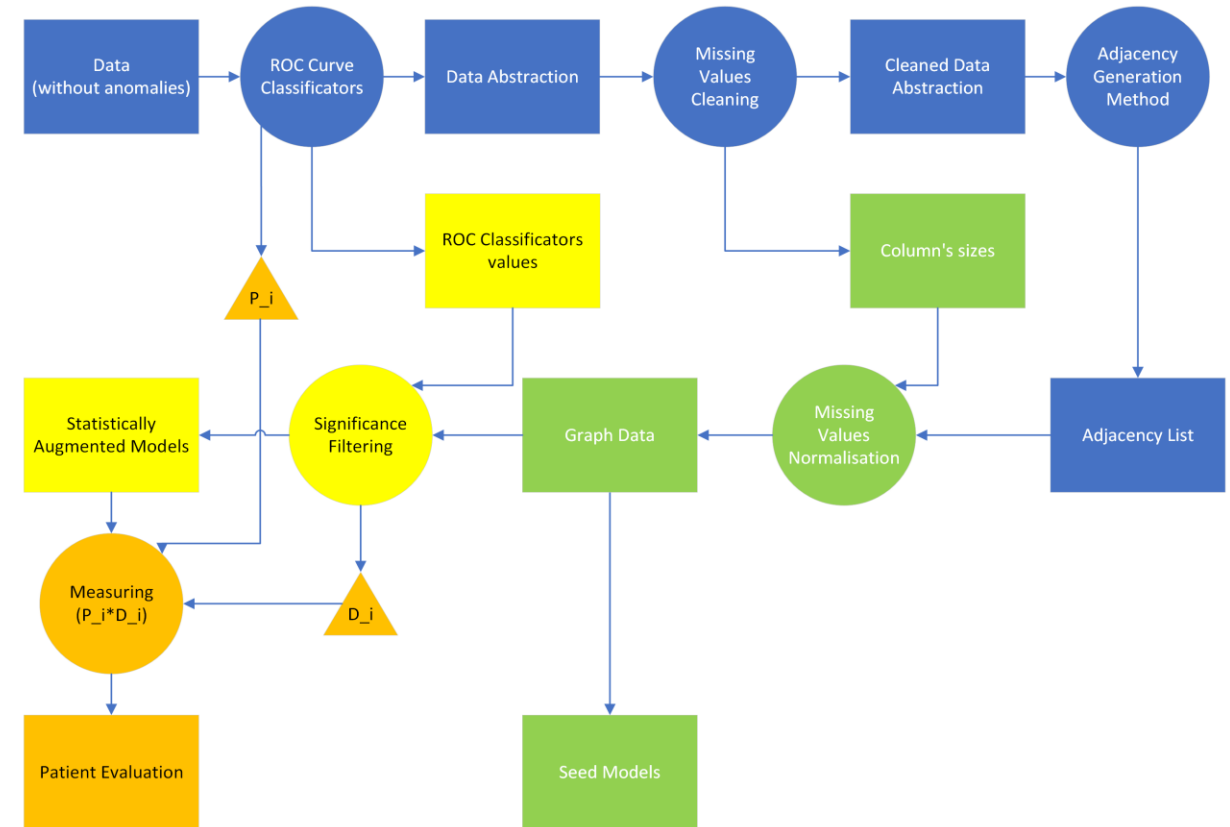
- Fundamental computational approaches aimed at emulating the intricate processes of both "System 1" and "System 2" decision-making in AI systems within a concept of a digital mind
- Disease models' development to support evidence-based health care decision making



SaNDAA small and incomplete dataset analyser

Alfredo Ibias, Varun Ravi Varma, Karol Capala, Luca Gherardini, Jose Sousa

- Personalized health often involves small datasets with missing data.
- Current Machine Learning methods struggle with such datasets due to high data volume requirements.
- SaNDA prioritizes explainability over efficiency due to dataset characteristics and domain importance.
- Evaluation against Random Forest (baseline for explainable methods) and gcForest (state-of-the-art for small datasets).
- SaNDA outperforms Random Forest with more missing data or fewer entries, but less favourable results on larger, well-curated datasets.
- Preferred over gcForest for explainability and privacy protection.
- Valuable for healthcare practitioners given data challenges in the field.

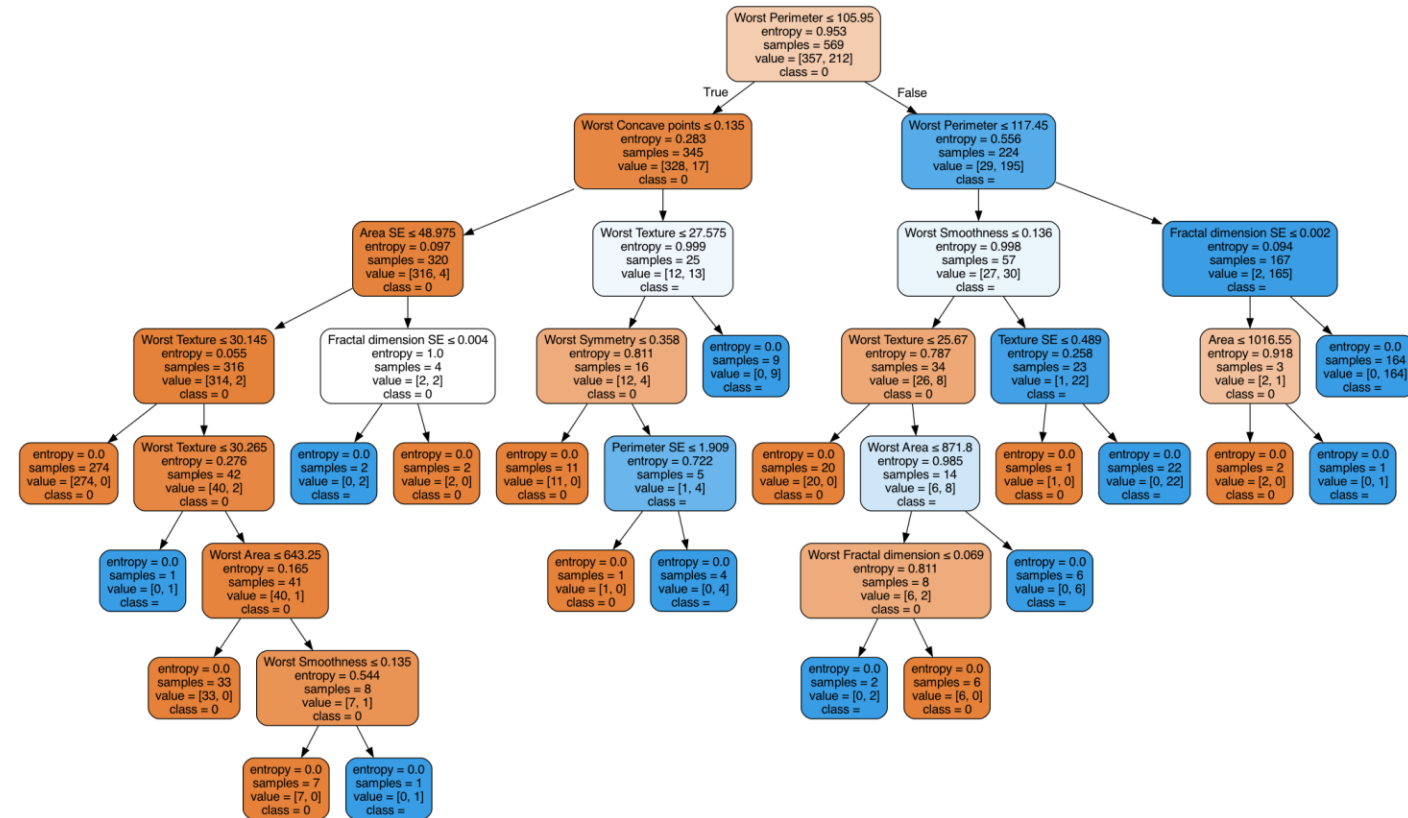


CACTUS: a Comprehensive Abstraction and Classification Tool for Uncovering Structures



Luca Gherardini, Varun Ravi Varma, Karol Capala, Roger Woods, Jose Sousa

- Availability of large data sets drives current AI developments.
- Challenges for small data sets due to practicality, cost, and opacity of deep learning models.
- CACTUS focuses on secure analytics using explainable AI.
- Enhanced support for categorical attributes while preserving their original meaning.
- Optimizes memory usage and speeds up computation through parallelization.
- Displays attribute frequency in each class and ranks them based on discriminative power.
- Assessment conducted on Wisconsin diagnostic breast cancer and Thyroid0387 data sets.



<https://doi.org/10.48550/arXiv.2308.12031>, under review: ACM Transactions on Intelligent Systems and Technology
under review: IEEE Transactions on Pattern Analysis and Machine Intelligence, "Noise Reduction through Abstractions and Its Possibilities for Machine Learning"

Clinical Data Science / Network Medicine & AI



Leader: Ahmed Abdeen Hamed

We focus on:

- Development of innovative network medicine approaches to complex disease research and management
- Development of Generative AI methods to improve clinical decision support for complex disease treatment

Precision Thermostability Prediction: Fueling Drug Discovery for Healthcare's Biological Targets



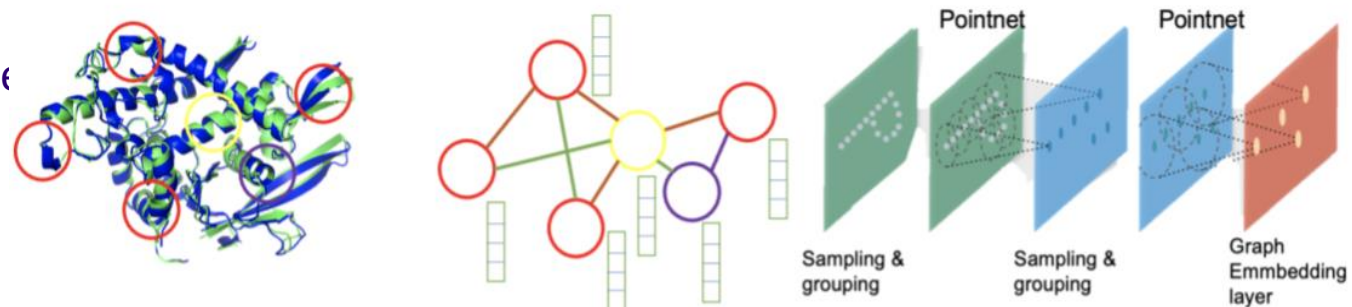
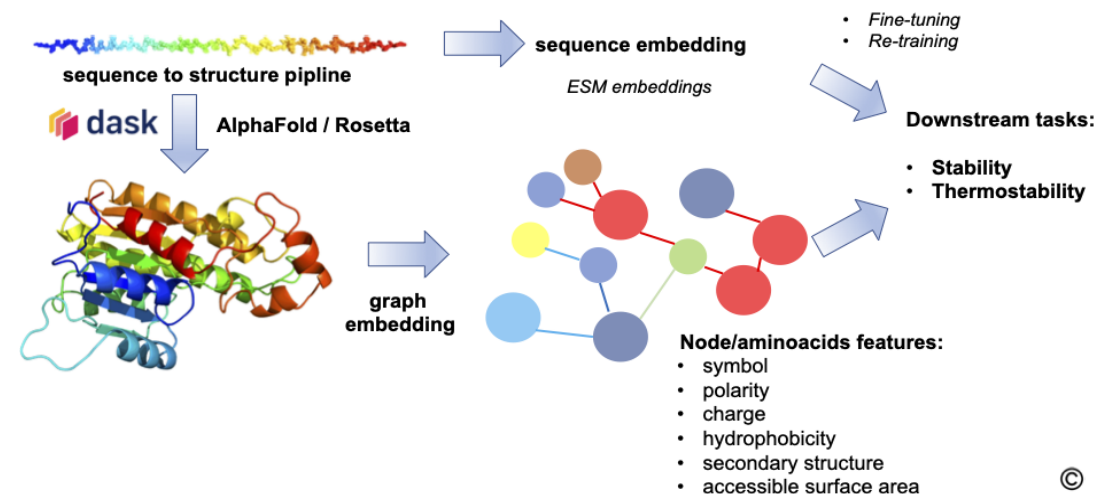
Adam Sulek, Jakub Jonczyk, Ahmed Abdeen Hamed

Question: Can accurate prediction of protein thermostability enhance drug discovery for biological targets in healthcare?

Solution: Precision thermostability prediction: fueling drug discovery for healthcare's biological targets

Features:

- increasing reliance on thermostable enzymes in healthcare.
- challenges in protein design: substantial hurdles in designing thermostable proteins.
- cutting-edge technology use: implementation of advanced graph neural networks (GNN).
- data-driven approach: utilization of publicly available data sources for prediction.
- superior model performance: GNN model outperforms PointNet++ with higher correlation ($R^2 \sim 0.6$).



Establishing Trust in LLM Capabilities

Ground-truth, Fact-Checking, and Computational Verification

Ahmed Abdeen Hamed, Byung Lee

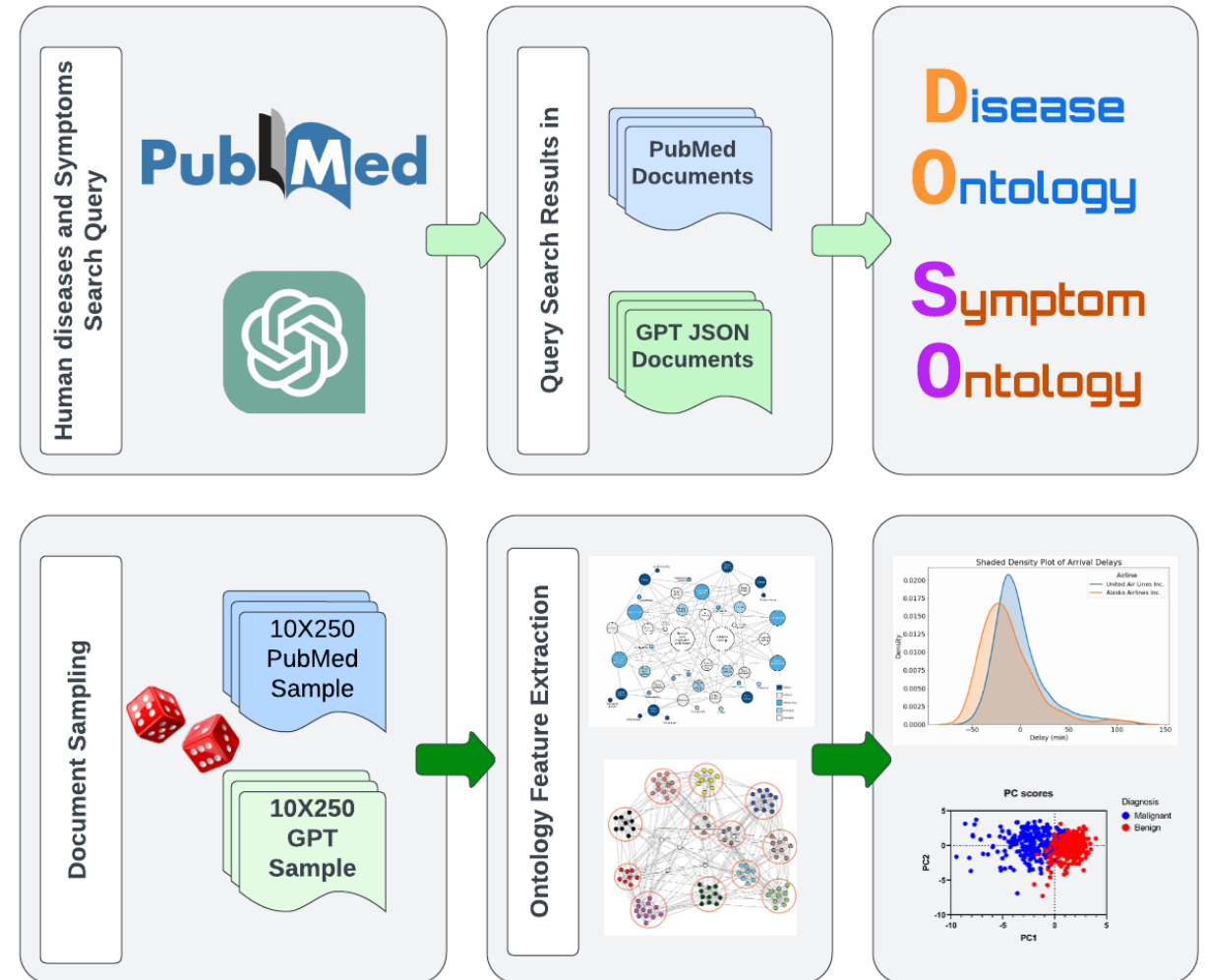


Question: Can trust in LLM capabilities be ensured via ground-truth, fact-checking, and computational verification?

Solution: To strengthen trust in LLM capabilities, incorporating algorithms alongside knowledge graphs and biological ontologies is key. LLMs can deliver more dependable results, especially in healthcare and biotechnology.

Features:

- Ground-Truth for Contextual Accuracy: Providing reliable, domain-specific information based on verified ground truth data.
- Rich BioMedical Knowledge Graphs: Enable context-rich, reproducible responses.
- Biological Ontologies: Ensure precise data interpretation.
- Algorithmic Validation: Rigorous fact-checking.
- Trust Enhancement: Greater reliability in LLM capabilities.



Extreme-Scale Data & Computing Team

Leader: Maciej Malawski

Goal: To apply the experience in large-scale computing systems to medical applications and to push the boundaries of current technologies when required by scale of data and computing demands

Research topics:

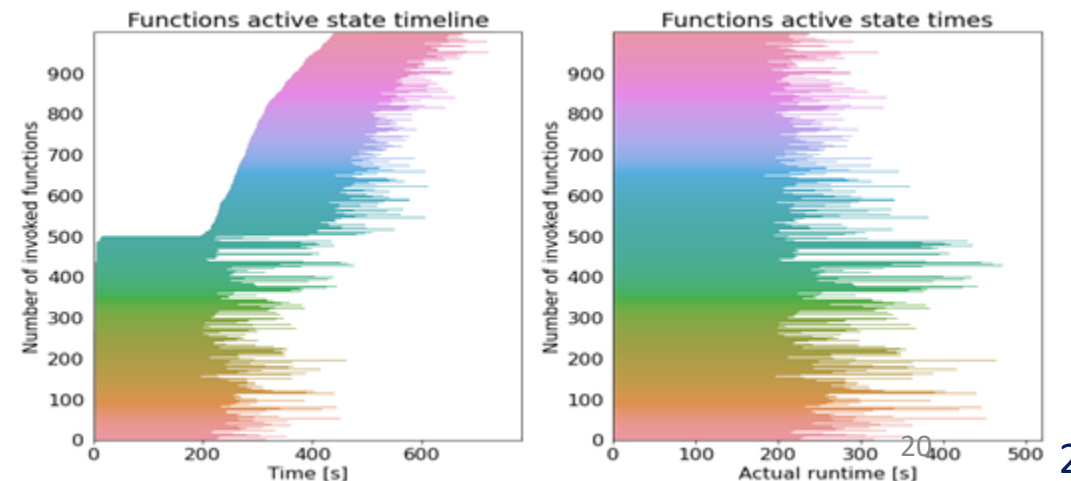
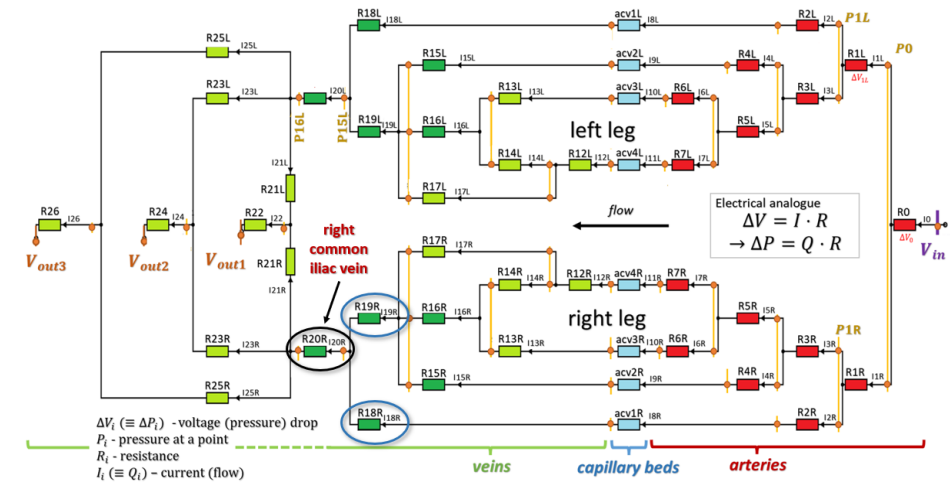
- Efficient processing of large data sets with the use of advanced computing infrastructures (HPC and Cloud)
- Development of tools for large-scale computing and data analysis
- Distributed and federated machine learning
- Performance evaluation and optimization of medical applications on emerging computing infrastructures

Verification, Validation, Uncertainty Quantification of simulation models



Magdalena Otta, Piotr Kica

- 0D, 1D, 3D models of cardiovascular systems
- Example: Venous modelling to inform treatment of patients presenting with lower limb thrombosis (PhD project with University of Sheffield)
- VVUQ, Sensitivity analysis - required in the regulatory process of model certification
- Large parameter studies
 - 50 000 parameters -> tasks
- CloudVVUQ library using serverless computing
- Total time to compute locally \approx **76.1h**
- Total time to compute using Google Cloud \approx **12.5min**



<https://github.com/SanoScience/CloudVVUQ>

Otta, M. et al.. Sensitivity analysis of a model of lower limb haemodynamics (2022) https://doi.org/10.1007/978-3-031-08757-8_7

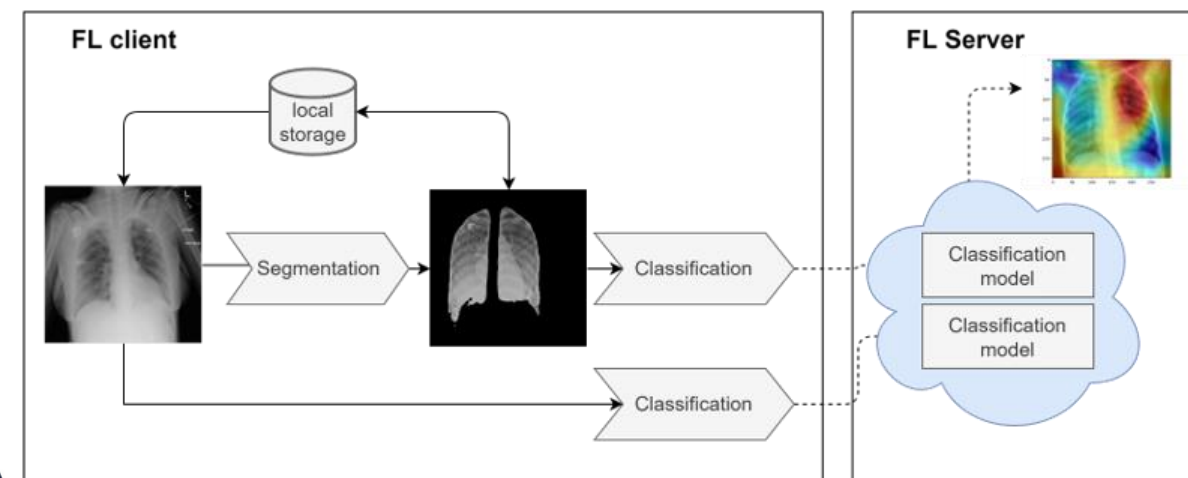
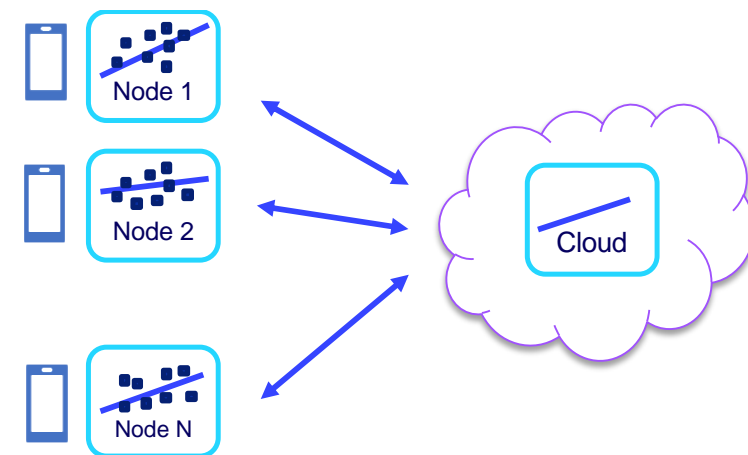
Piotr Kica, Magdalena Otta, Krzysztof Czechowicz, Karol Zajac, Piotr Nowakowski, Andrew J. Narracott, Ian Halliday, Maciej Malawski: Serverless Approach to Sensitivity Analysis of Computational Models, CCGrid 2023

Federated Learning for Medical Applications

Dominika Ciupek, Michał Daniłowski, Jan Przybyszewski



- Training of ML models without transferring data from patients / doctors / hospitals
- Data: Brain imaging, Chest X-Ray, EEG signals, genomic data
- Evaluation in real distributed environments
- Experiments on Prometheus, Ares (GPU) and Google Cloud Platform (GPU Instances)
- Large-scale Chest X-Ray studies (600,000 images) from multiple datasets (classification, segmentation)
- Collaborations: NearData EU Project (U. Tarragona, Barcelona Supercomputing Center, TU Dresden)

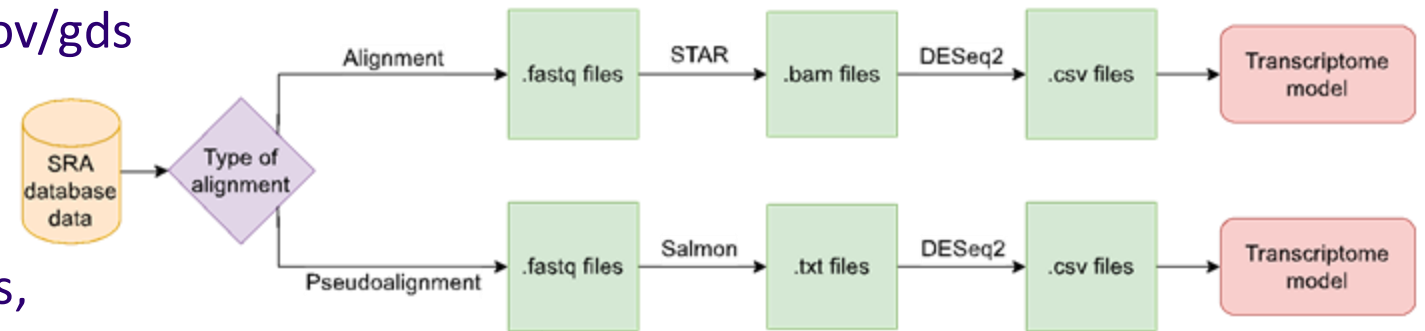


Filip Ślęzyk, Przemysław Jabłocki, Maciej Malawski, Aneta Lisowska and Szymon Plotka: CXR-FL: Deep Learning-based Chest X-ray Image Analysis Using Federated Learning, ICCS, 2022

Transcriptomics Atlas Pipeline

Sabina Lichołai, Piotr Kica, Jan Przybyszewski

- ✦ Big data analysis of transcriptomics databases
- ✦ Based on publicly available data: NCBI GEO <https://www.ncbi.nlm.nih.gov/gds>
- ✦ Differentiated at various levels: organs, tissues, cells versus diseases and conditions.
- ✦ Analysis of 15+ different tissues/organs, 1000+ files of 10s GBs per each tissue, 1h+ on a single core per file
- ✦ Pipeline based on standard bioinformatics packages for preprocessing, alignment and reads counting (STAR, salmon, DESeq2)
- ✦ First results: Comparison of Cloud vs HPC
- ✦ Containerization for portability: Docker and Singularity/Apptainer
- ✦ Collaborations: NearData EU Project (U. Tarragona, Barcelona Supercomputing Center)



P. Kica, S. Lichołai, M. Malawski: Transcriptomics Atlas Pipeline: Cloud vs HPC, submitted to WORKS,23 Workshop at SC23 conference

Scientific Programming Team

Leader: Piotr Nowakowski

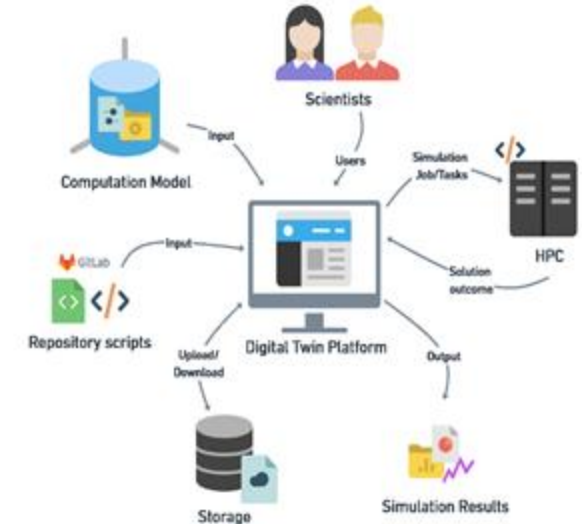
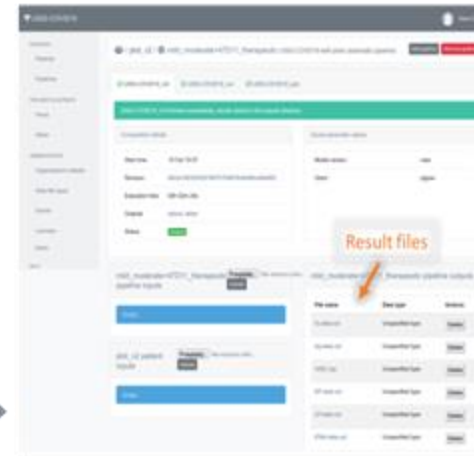
- Each member of SPT is affiliated with one or more research teams, where they participate in development of artifacts, services and customized IT tools
- SPT members also work on research grant proposals, and have been quite successful in this regard (as evidenced e.g. by recent acquisition of the GEMINI research grant)

Platform for simulations of digital twins



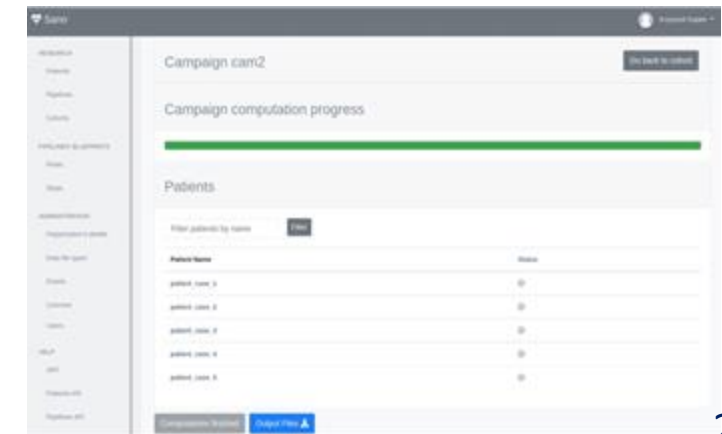
Jan Meizner, Karol Zajac, Adam Nowak, Krzysztof Gadek, Piotr Nowakowski

- Web-based system for management of multi-step computational pipelines on HPC systems
- Patient data-centered workflows:
 - Steps -> Pipelines -> Campaigns
 - Patients -> Cohorts
- Models implemented as software artifacts
 - Versioning: GitHub, GitLab
 - Comparison of results
 - **Reproducibility** -> Robust Science
- File management (HPC storage, Cloud storage)
- Usage in In Silico World EU Project
- Collaboration with Univ. Bologna, Univ. van Amsterdam, KU Leuven



<https://mee.cyfronet.pl/>

Marek Kasztelnik, Piotr Nowakowski, Jan Meizner, Maciej Malawski, Adam Nowak, Krzysztof Gadek, Karol Zajac, Antonino Amedeo La Mattina, Marian Bubak: Digital Twin Simulation Development and Execution on HPC Infrastructures. ICCS (2) 2023: 18-32



HPHOB protein folding calculator

Piotr Nowakowski, Tomasz Gubała, Krzysztof Gądek



- The Hydrophobicity Calculator (HPHOB) tool is used to simulate the folding process of proteins (or protein domains) based on the Fuzzy Oil Drop model, which has been developed at CMUJ
- We have encapsulated the core of the model (which is a standalone executable) with a GUI web application, enabling users to experiment with the model
- The system is also integrated with PDB, providing access to a broad range of protein structures.



PDB ID: [4HW9](#)

Crystal Structure of *Helicobacter pylori* MscS (Closed State)

Classification: MEMBRANE PROTEIN

Organisms: *Helicobacter pylori*

DOI: [10.1002/pro.2222](#)

Image and data retrieved from Protein Data Bank service of RCSB (rcsb.org). Presenting data and image of entry PDB ID 4HW9. Entry citation: Lai, J.Y., Poon, Y.S., Kaiser, J.T., Rees, D.C. Open and shut: crystal structures of the dodecylmaltoside solubilized mechanosensitive channel of small conductance from *Escherichia coli* and *Helicobacter pylori* at 4.4 angstrom and 4.1 angstrom resolutions. *Protein Sci.* (2013) 22 502-509

Process

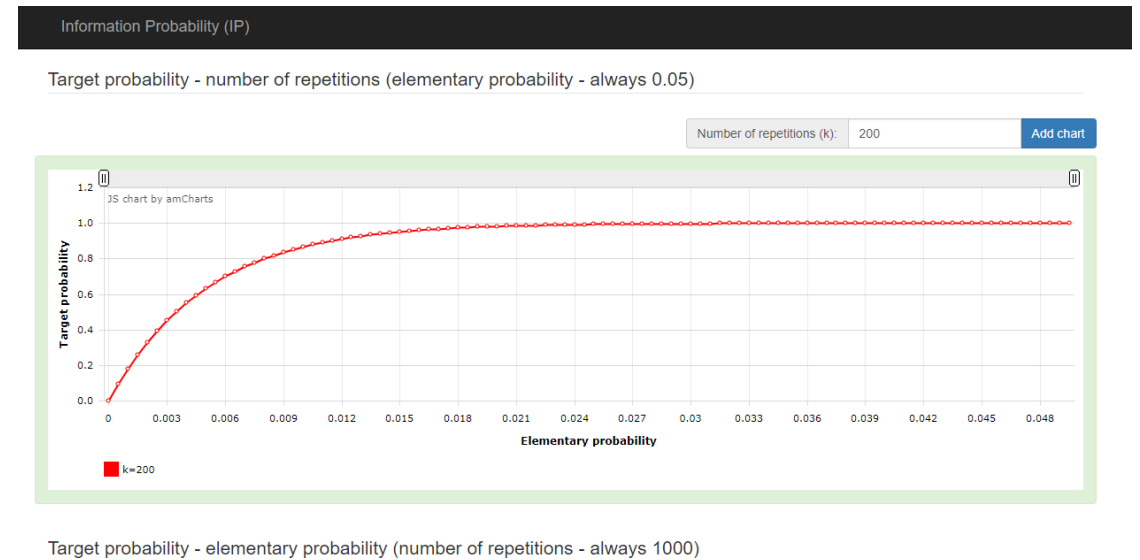
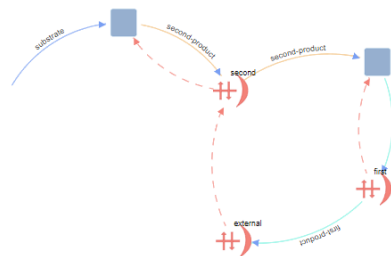
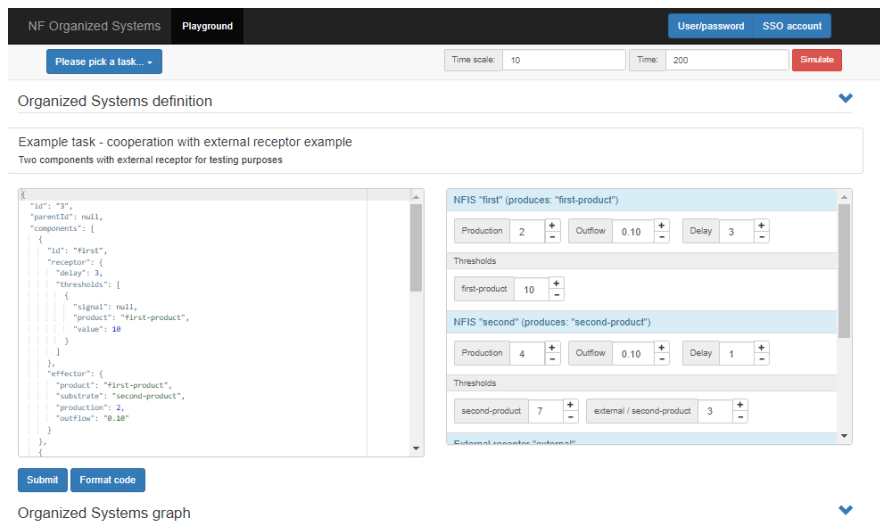
Choose fragment

<https://hphob.sano.science>

Systems Biology simulation tools

Visualizing interactions between feedback loops which comprise biological systems

- We host a selection of tools developed at CMUJ to supplement their research publications and training materials



- The Information Probability (IP) tool and the NF Organized Systems (NFS) calculator have been ported to Sano resources, and are available as training aids
- Both tools are also referenced in books published by our CMUJ collaborators

Sano Computational Medicine Seminars

- A mean to enhance knowledge and build a scientific network
- Started on 22 June 2020; total 103 seminars until 26th June 2023
- The seminars are on Mondays at 14.00 CET/CEST
- Most of seminars are recorded and available at Sano YouTube also as a teaching material
- The list of Sano Seminars is available at Sano website:

<http://sano.science/seminars>



The banner features the Sano logo and the text 'SANO SEMINARS' in large, stylized letters. Below this, the name 'Jarosław Kwiecień' is displayed, followed by his title 'Software developer, data scientist and CTO of CancerCenter'. The seminar title 'A practical view of Artificial Intelligence in histopathology' is prominently shown. At the bottom, the date '9th January 2023' and time '2:00-3:30 PM (CEST)' are provided with corresponding icons. A circular portrait of Jarosław Kwiecień is on the right side of the banner.

 sano

SANO SEMINARS

Jarosław Kwiecień
Software developer, data scientist and CTO of CancerCenter

A practical view of Artificial Intelligence in histopathology

 9th January 2023  2:00-3:30 PM (CEST)

Board, Foundation Council and Scientific Committee



- **Management Board** conducts the daily operations of the Foundation

- **International Scientific Committee** selects research team leaders, evaluates their achievements, and advises the Management Board in scientific matters

Name	Role
Prof. Maciej Malawski	Director, President of the Management Board
Dr Marian Bubak	Director for Scientific Affairs, Member of the Management Board

- **Foundation Council** supervises Sano and appoints Management Board members

Name	Affiliation
Prof. Marco Viceconti, President	Università di Bologna, IT
Prof. Rod Hose	Emeritus Professor in the Department of Infection, Immunity and Cardiovascular Disease, University of Sheffield, UK
Prof. Richard Clayton	University of Sheffield, UK
Prof. Marek Behr	Aachen University, DE
Prof. Holger Gohlke	Heinrich Heine University Düsseldorf, DE
Prof. Dieter Kranzlmüller	Ludwig-Maximilian University Munich, DE
Prof. Zbigniew Nawrat	Silesian Center for Heart Diseases in Zabrze, PL
Prof. Irena Roterman-Konieczna	Jagiellonian University Medical College, PL
Prof. Jerzy Gąsowski	Jagiellonian University Medical College, PL
Prof. Knut Koschätzky	Fraunhofer Institute ISI, DE
Dr Claire Chalopin	University of Leipzig, DE
Dr Stefan Zachow	Zuse Institute Berlin, DE
Prof. Ewa Deelman	University of Southern California, USA
Prof. Peter Sloot	University of Amsterdam, NL
Prof. Jacek Kitowski	AGH University of Science and Technology, PL

Name	Affiliation
Karol Krawentek	ACC Cyfronet AGH, PL
Dr Tanja Bratan	Fraunhofer Institute ISI, DE
Dr Andrew Narracott	University of Sheffield, UK
Dr Olav Zimmermann	Forschungszentrum Jülich, DE
Kazimierz Murzyn President	Klaster LifeScience Kraków, PL

Sano in numbers (as for 31 August 2023)



Oversight and governance:

- Foundation Council – 5 members
- International Scientific Committee – 15 members
- Management Board – 2 members

Research:

- Team leaders – 5
- Senior postdoc – 4
- Postdocs - 12
- PhD students – 19
- Scientific programmers, technical staff – 22
- Visting scientists – 7
- MSc students (FNP scholarships) - 13

Support:

- Operations – 7
- Human talent management – 2
- Finances - 3
- Communication – 3
- Business development – 3
- Legal counsel – 1
- Data protection – 2

Summary - collaboration opportunities



R&D

Developing new methods
computing, algorithms,
models and technologies for
personalized medicine



Education

Contribution to
innovative training
programs & education
that meet the needs of
modern personalized
medicine



sano

Industrial collaboration

Supporting enterprise
development along with
creation of new diagnostic
and therapeutic technologies
(DSS)

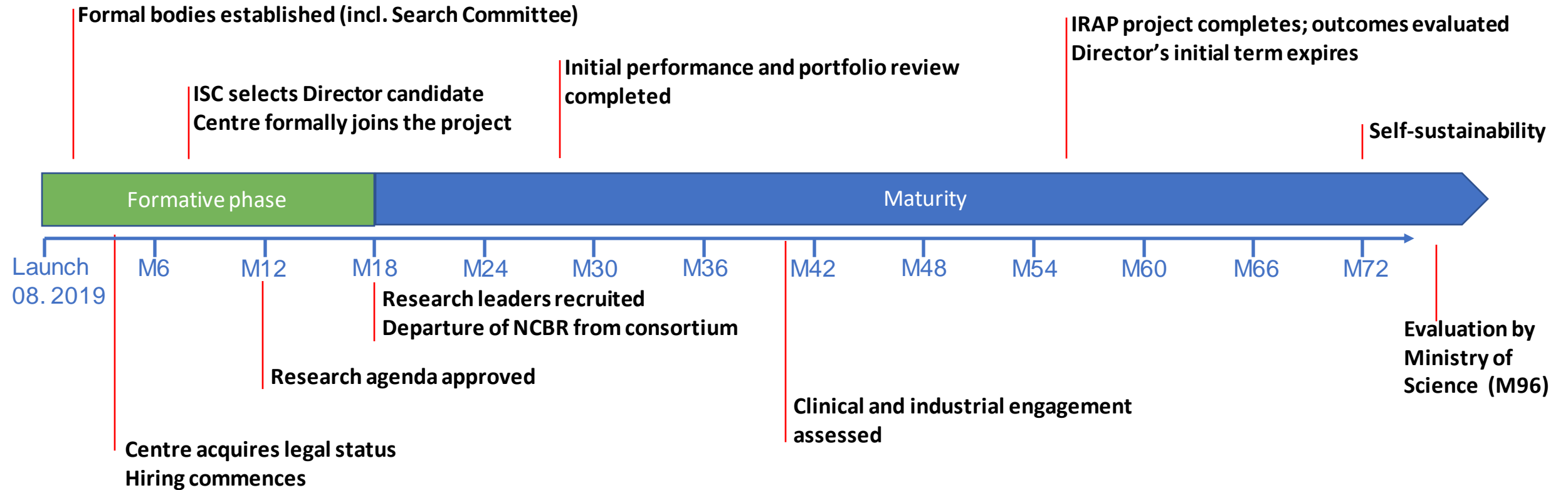


Commercialization

Introducing novel
diagnostic methods and
therapies based on
computer simulations
into clinical practice



Sano timeline



Sano in 2026:
<https://sano.science>

“We prevent and heal disease through the use of computing and data”



Centre for Computational
Medicine

www.sano.science

Marian Bubak
Scientific Affairs Director
m.bubak@sanoscience.org



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857533 and from the International Research Agendas Programme of the Foundation for Polish Science No MAB PLUS/2019/13.



Republic
of Poland



Foundation for
Polish Science

European Union
European Regional
Development Fund

