

Warsaw University of Technology

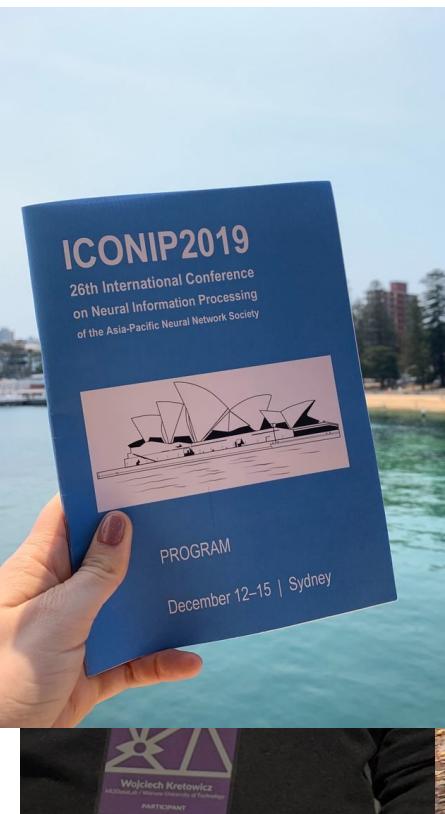
Machine Learning Groups

Przemek Biecek

przemyslaw.biecek@gmail.com

MI2DataLab (Math+Info)^{^2}

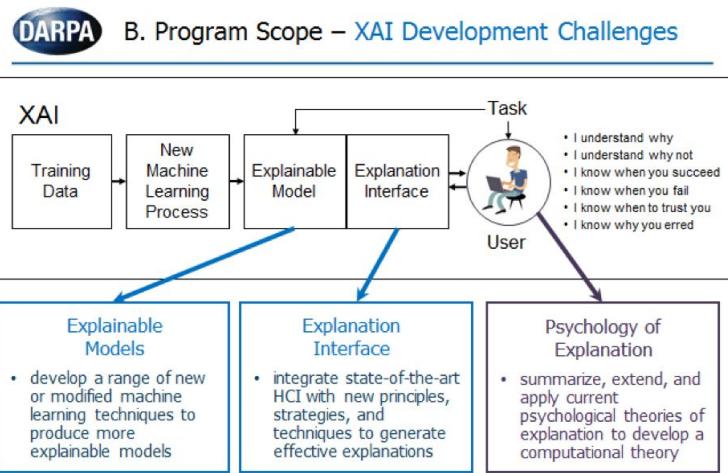
<http://mi2.mini.pw.edu.pl/>



Explainable and Reliable Machine Learning

We are building tools

<https://pbiecek.github.io/explainFIFA20/>



github.com/ModelOriented

The screenshot shows the GitHub repository page for 'Model Oriented'. The repository is owned by 'MI2DataLab @ Warsaw University of Technol...' and has a URL of <https://mi2-warsaw.github.io/>. The page includes navigation links for Repositories (42), Packages, People (20), Teams (2), Projects, and Settings. Below this, a section titled 'Pinned repositories' lists four projects:

- DALEX**: moDel Agnostic Language for Exploration and eXplanation. Developed in Python. 590 stars, 86 forks.
- DrWhy**: DrWhy is the collection of tools for eXplainable AI (XAI). It's based on shared principles and simple grammar for exploration, explanation and visualisation of predictive models. Developed in R. 350 stars, 48 forks.
- randomForestExplainer**: A set of tools to understand what is happening inside a Random Forest. Developed in R. 159 stars, 25 forks.
- modelStudio**: Interactive Studio for Explanatory Model Analysis. Developed in R. 123 stars, 16 forks.

Evidence based Machine Learning

We are debugging myths

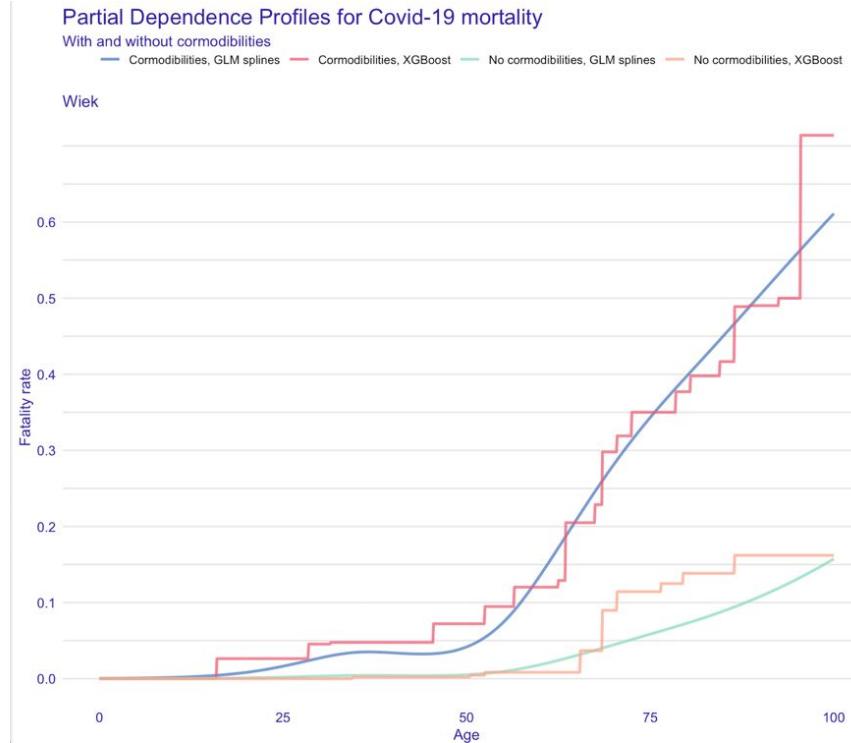
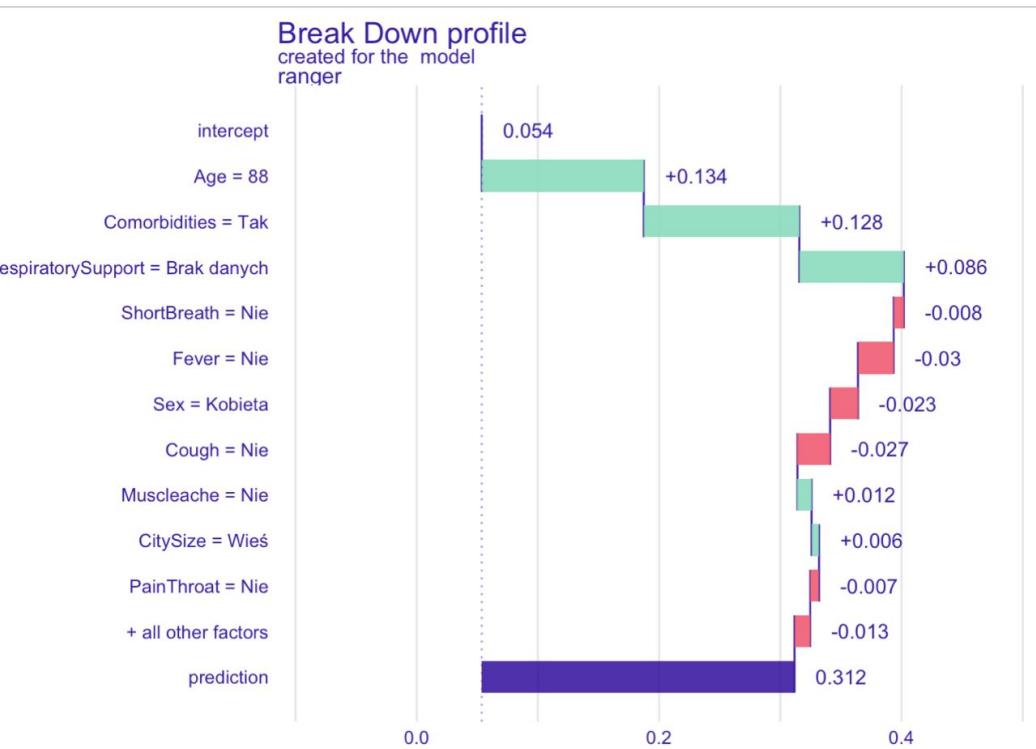
<https://github.com/mini-pw/2020L-WB-Book>

dataset name (dataset ID)	# obs	prc of missings	# numeric	# numeric w. missings
ipums_la_99-small (1018)	8844	7%	15	0
adult (1590)	48842	1%	4	0
eucalyptus (188)	736	3.9%	14	10
dresses-sales (23381)	500	14.7%	1	1
colic (27)	368	16.3%	5	5
credit-approval (29)	690	0.6%	6	2
sick (38)	3772	2.2%	6	6
labor (4)	57	33.6%	8	8
SpeedDating (40536)	8378	1.8%	59	58
hepatitis (55)	155	5.4%	6	5
vote (56)	435	5.3%	0	0
cylinder-bands (6332)	540	5.1%	19	18
echoMonths (944)	130	7.5%	6	6



We are building models

<https://github.com/MOCOS-COVID19/dark-figure>



Looking for PhD students and postdocs
(in short and long projects)



Human Oriented autoMated
machinE leaRning

Computer Vision Lab

Tomasz Trzciński

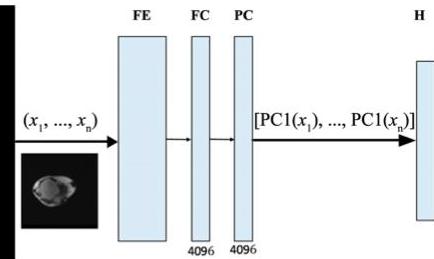
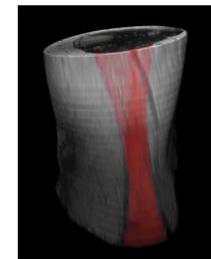
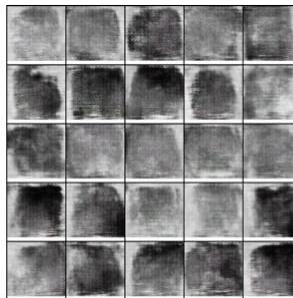
Warsaw University
of Technology



Zakład Grafiki
Komputerowej

Medical imaging analysis

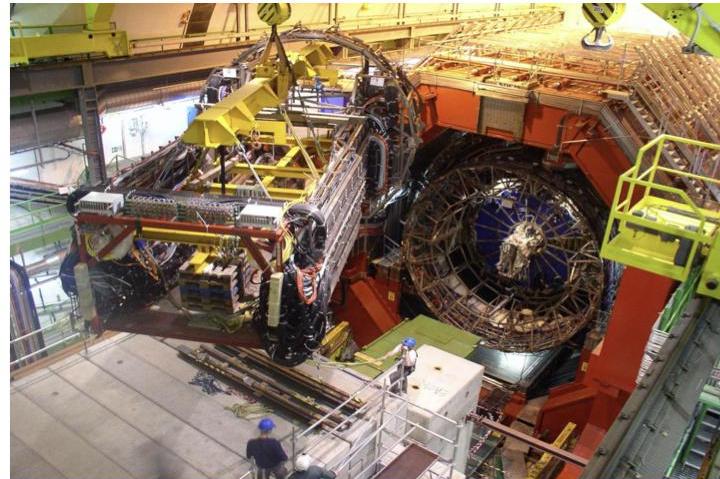
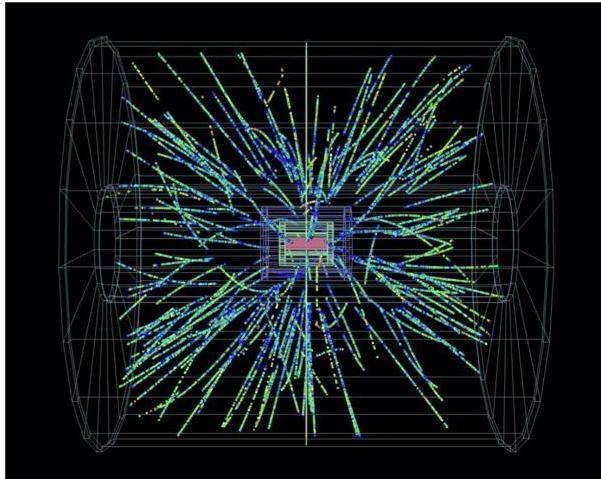
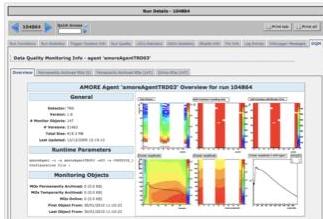
- **Preterm birth prediction** for ultrasound images
- **Generative adversarial** approach for data **privacy** enforcement
- **Achilles tendon healing** monitoring with neural networks



Machine Learning in CERN

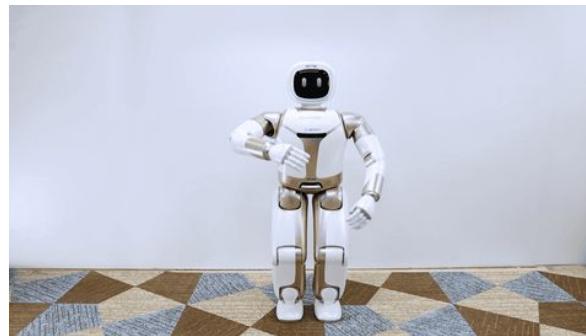
Data Quality Assurance in the ALICE experiment in LHC

- Machine learning for run **quality assessment**
- Collision simulation with **Generative Adversarial Networks**

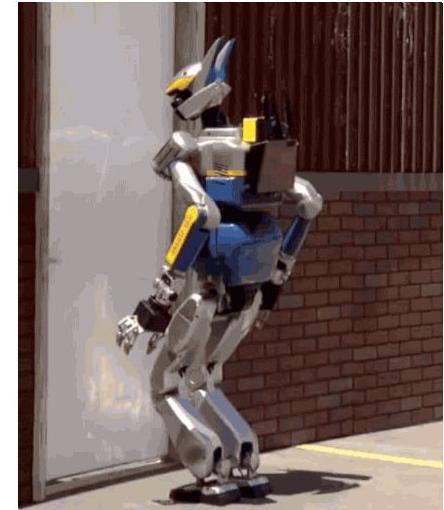


Binary vectors in continual learning

- **Adapting models** to changing conditions
- Data treated as a **stream**, not a constant set
- Changing tasks
- **Catastrophic forgetting**



Model during training



Model during testing

Neural rendering

- **Low Shot Realistic Human Rendering from Partial Information**
- Generating controllable HD video sequences with GANs
- PhD Scholarship programme supported by **Microsoft Research** with up to 3 000 EUR monthly

APPLY till July 31st 2020



Past projects

Visual geolocation (Google ARCore collaboration)



Subjective attribute estimation for social media



Comixify: turning videos into comics



Team

Post docs

Jacek Komorowski
Łukasz Dąbała
Karol Piczak

PhD students

Kamil Deja
Witold Oleszkiewicz
Tomasz Włodarczyk

Grad students

Szymon Płotka
Dominik Trusiński
Daniel Iwanicki

Daniel Marczak

Karolina Pawluk

Professor

Przemysław Rokita



Funding:



Narodowe Centrum
Badań i Rozwoju

Thank you

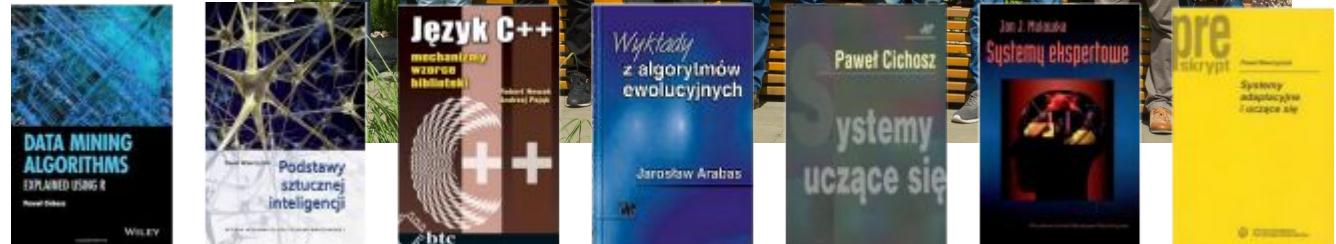
tomasz.trzcinski@pw.edu.pl

ii.pw.edu.pl/~ttrzcins



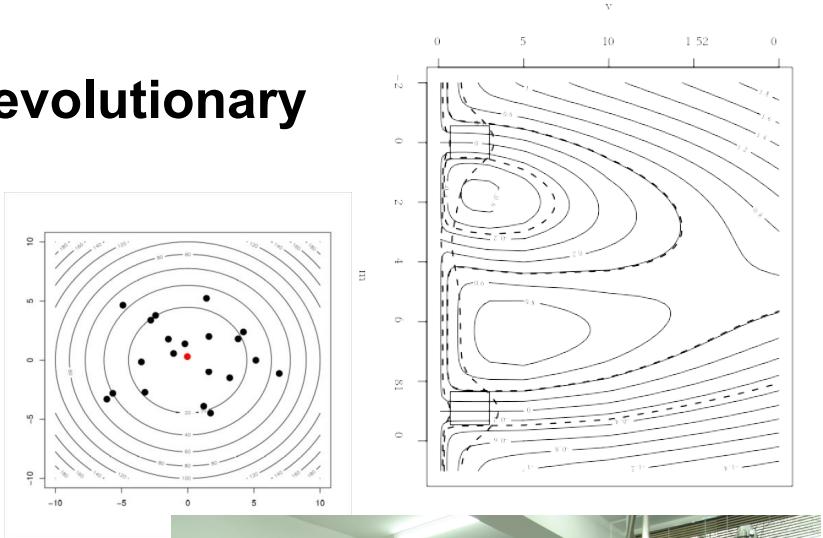
Artificial Intelligence Division (13 staff, 10 PhD students, 40 diplomants)

- Jan Mulawka, Prof, PhD, DSc
- Jarosław Arabas, PhD, DSc
- Robert Nowak, PhD, DSc (head)
- Paweł Wawryński, PhD, DSc
- Rafał Biedrzycki, PhD
- Paweł Cichosz, PhD
- Stanisław Kozdrowski, PhD
- Karol Piczak, PhD
- Bartłomiej Twardowski, PhD
- Paweł Zawistowski, PhD
- Wiktor Kuśmirek
- Łukasz Neumann
- Witold Oleszkiewicz



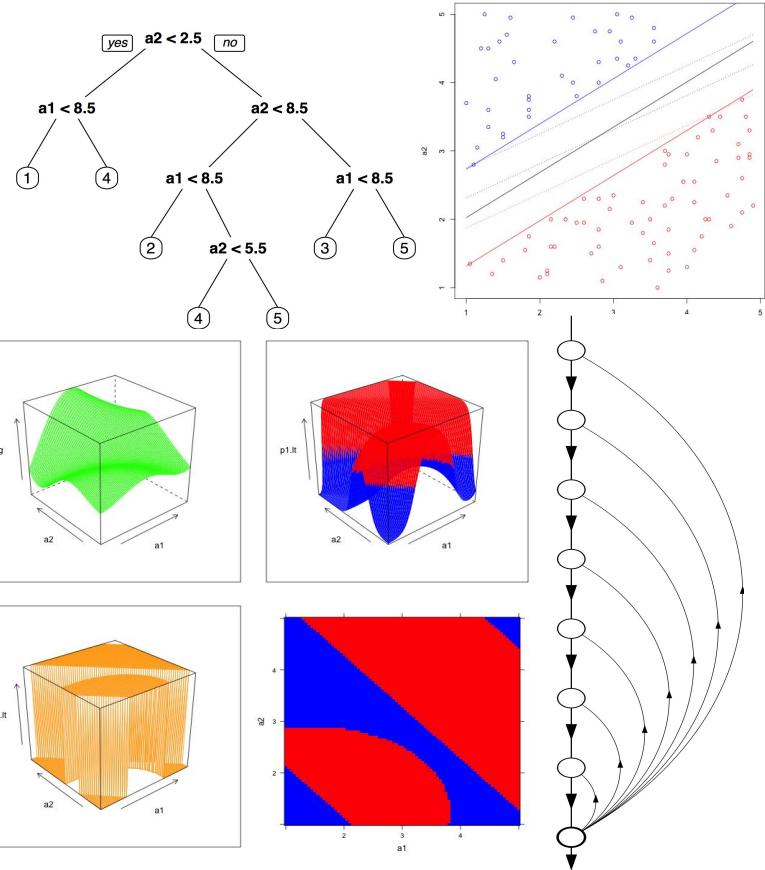
New optimization methods based on evolutionary algorithms

- Modeling of evolutionary algorithm properties
- New version of optimization algorithms
 - high ranking in international competitions:
- Optimisation methods used in practice:
 - Semiconductor mirror
 - Calibration standards
 - Optimization Jiles-Atherton model
 - Optimization of motif searching in DNA

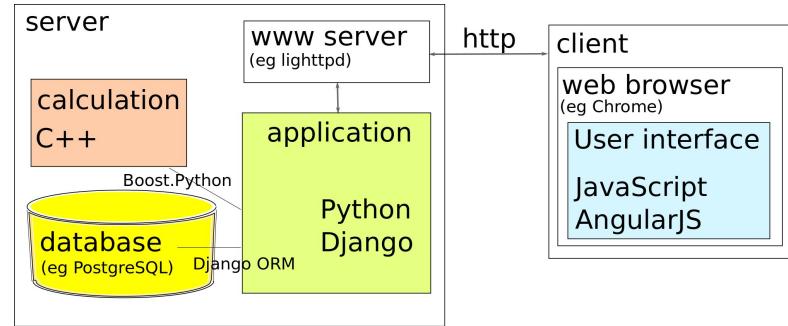
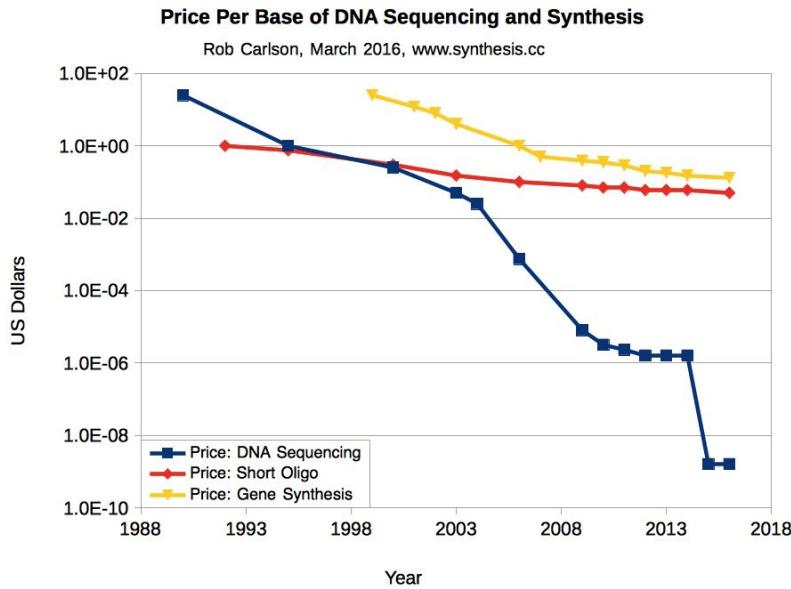


Machine Learning Algorithms

- **Predictive modeling:** learning classification, regression, and clustering models
- **Neural Networks:** deep learning, convolutional, recursive, LSTM and others
- **Anomaly detection:** learning patterns of normal and anomalous behavior or events
- **Text classification and clustering:** assigning or discovering document classes
- **Collaborative filtering:** learning user/customer interest patterns from ratings
- **Reinforcement learning:** learning decision/control policies from delayed rewards



Genomics and synthetic biology



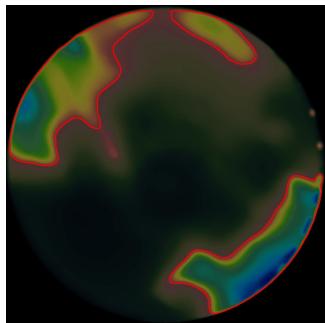
Areas of interest: Assemblers NGS, Genomic Variant Data Warehouse, Variant Prioritization, Quality Control Software for NGS data, Copy Number Variation Detection, Haplotype Analysis, Artificial Gene Synthesis

Technologies: Bioweb framework, Hadoop, Apache, Kafka, Kylin, R, C++, CUDA

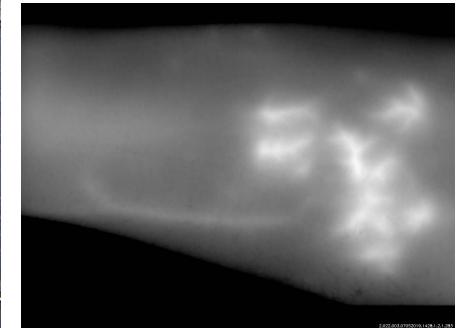
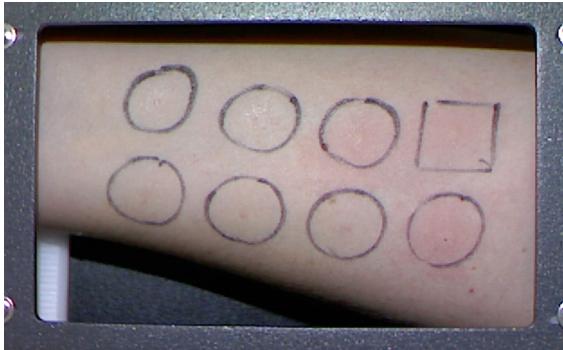
Collaboration: Warsaw Medical University, Warsaw University of Life Sciences, Institute of Mother and Child in Warsaw, ETH Zurich, Baylor College of Medicine (Houston)

Medical image analysis

Breast cancer detection using thermal images, 1 M PLN, Braster SA



Skin allergy response detection using visual & thermal images, Milton-Essex SA



- image processing, feature extraction, feature significance
- machine learning (classifiers)
- efficient implementation in C++, concurrency
- used in production as SaaS, cooperation with external systems.

Financial data analysis

Data cleaning: missing attribute values calculation, misspells (dictionary based, ontologies),

Feature generation, eg. public transport availability, population of the city, weather, weather forecast, public holidays

Model building: bayesian, random forests, neural networks, SVM and others

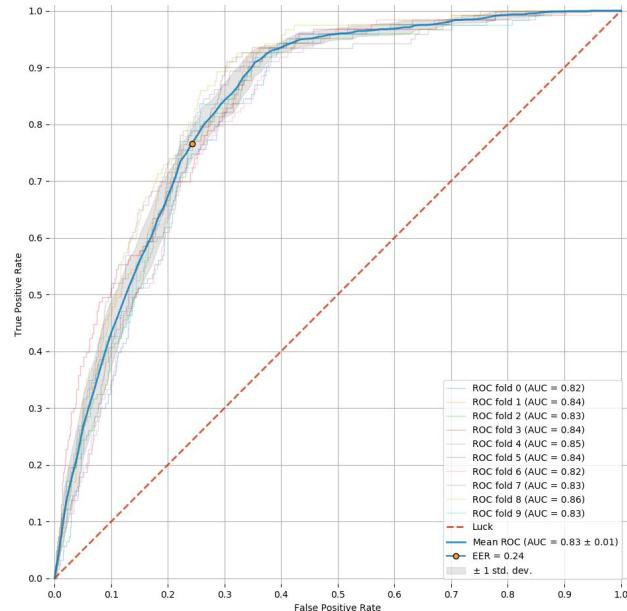
Attribute significance

Cost matrix

Business process change calculation

ROI estimation

Example: the need for replacement vehicle for car insurance companies





Institute of Computer Science
Warsaw University of Technology

Artificial Intelligence Division

<http://ai.ii.pw.edu.pl>

Thank you

robert.nowak@pw.edu.pl