HELMHOLTZ AI KICK-OFF MEETING
5 March 2020 - Lenbach Palais, Munich

POSTER ABSTRACTS
HELMHOLTZ AI KICK-OFF MEETING
5 March 2020 - Lenbach Palais, Munich

POSTER ABSTRACTS
### Content

<table>
<thead>
<tr>
<th>PS-01</th>
<th>Workshop on machine learning in Earth System Sciences</th>
</tr>
</thead>
<tbody>
<tr>
<td>PS-02</td>
<td>A connectivity based Psychometric Prediction Framework for Brain-Behavior Relationship Studies</td>
</tr>
<tr>
<td>PS-03</td>
<td>Machine Learning at GEOMAR- from sparse measurements</td>
</tr>
<tr>
<td>PS-04</td>
<td>Deep learning speeds up gapless cytoarchitectonic mapping in the human brain</td>
</tr>
<tr>
<td>PS-05</td>
<td>Helmholtz AI at Forschungszentrum Jülich for research field key technologies/Information</td>
</tr>
<tr>
<td>PS-06</td>
<td>HeAT- Helmholtz Analytics Toolkit</td>
</tr>
<tr>
<td>PS-07</td>
<td>Not one model fits all: unfairness in RFSC-based prediction of behavioral data in African American</td>
</tr>
<tr>
<td>PS-08</td>
<td>ML applications in High Energy Physics</td>
</tr>
<tr>
<td>PS-09</td>
<td>Helmholtz AI For Matter at Helmholtz-Zentrum Dresden-Rossendorf</td>
</tr>
<tr>
<td>PS-10</td>
<td>Helmholtz AI Consultants at Helmholtz Zentrum Dresden-Rossendorf</td>
</tr>
<tr>
<td>PS-11</td>
<td>Helmholtz. ai YIG on Digital Twinning for High Energy Density Science</td>
</tr>
<tr>
<td>PS-12</td>
<td>AI in medical image computing</td>
</tr>
<tr>
<td>PS-13</td>
<td>Estimating model uncertainty of Neural Networks in Sparse Information Form</td>
</tr>
<tr>
<td>PS-14</td>
<td>Deep Learning in Particle Physics for the CMS experiment at the LHC</td>
</tr>
<tr>
<td>PS-15</td>
<td>Learning for Discovery- Significance as Loss</td>
</tr>
<tr>
<td>PS-16</td>
<td>Discovery of Magnesium Degradation Modulating Additives using Machine Learning Techniques</td>
</tr>
<tr>
<td>PS-17</td>
<td>Helmholtz AI Consultants for Medical Imaging Analysis</td>
</tr>
<tr>
<td>PS-18</td>
<td>Automatic analysis of cortical areas in Whole Brain Histological Sections using Convolutional Neural Networks</td>
</tr>
<tr>
<td>PS-19</td>
<td>Repository mining; from Code to README Generation</td>
</tr>
<tr>
<td>PS-20</td>
<td>Artificial Intelligence to predict individual characteristics for personalised medicine</td>
</tr>
<tr>
<td>PS-21</td>
<td>Datasets and AI projects of the German Mouse Clinic</td>
</tr>
<tr>
<td>PS-22</td>
<td>The Munich School of Data Science @ Helmholtz, TUM and LMU</td>
</tr>
<tr>
<td>PS-23</td>
<td>Artificial Intelligence and Integration of Multi-omics datasets for Individualised Infection Medicine</td>
</tr>
<tr>
<td>PS-24</td>
<td>Helmholtz Imaging Platform (HIP)</td>
</tr>
<tr>
<td>PS-25</td>
<td>Coping with Concept Drifts in Load Forecasting using Machine Learning</td>
</tr>
<tr>
<td>PS-26</td>
<td>Machine Learning in Soil Research</td>
</tr>
<tr>
<td>PS-27</td>
<td>Adopting Conversational Interfaces for Exploring OSGi- based Software Architecture in Augmented Reality</td>
</tr>
<tr>
<td>PS-28</td>
<td>Machine Learning in Translational Genomics</td>
</tr>
<tr>
<td>PS-29</td>
<td>HIFIS: Helmholtz Infrastructure for Federated ICT services</td>
</tr>
<tr>
<td>PS-30</td>
<td>The Helmholtz Metadata Collaboration (HMC) - addressing the metadata problem</td>
</tr>
<tr>
<td>PS-31</td>
<td>Climate change and Health Impacts: An applied use case for Machine Learning</td>
</tr>
<tr>
<td>PS-32</td>
<td>FD-Net: Fourier Descriptors for Microstructural Object Detection</td>
</tr>
<tr>
<td>PS-33</td>
<td>Helmholtz Information and Data Science Academy (HIDA)</td>
</tr>
<tr>
<td>PS-34</td>
<td>Deep learning enables in silico chemical effect prediction</td>
</tr>
<tr>
<td>PS-35</td>
<td>Applying Al to persistent knowledge graphs to learn on complex multi-omics data</td>
</tr>
<tr>
<td>PS-36</td>
<td>Machine Learning Group @ DLR institute of data science</td>
</tr>
<tr>
<td>PS-37</td>
<td>Helmholtz AI Voucher system</td>
</tr>
<tr>
<td>PS-38</td>
<td>Spectro-image representation of DNA sequencing for Deep Learning</td>
</tr>
<tr>
<td>PS-39</td>
<td>Helmholtz AI Central Unit and Local Unit Health</td>
</tr>
<tr>
<td>PS-40</td>
<td>Helmholtz AI Cooperation Unit</td>
</tr>
<tr>
<td>PS-41</td>
<td>Meno: An AI- powered, peer-to-peer publishing platform to revolutionize scientific publishing</td>
</tr>
<tr>
<td>PS-42</td>
<td>Uncertainty quantification of soil moisture predictions</td>
</tr>
<tr>
<td>PS-43</td>
<td>BlenderProc</td>
</tr>
<tr>
<td>PS-44</td>
<td>Deep learning for Climate and Weather</td>
</tr>
<tr>
<td>PS-45</td>
<td>HAIACU-MASTr: Munich@Aeronautics, Space and Transport</td>
</tr>
<tr>
<td>PS-46</td>
<td>Deep learning for cell-type annotation taks does not outperform classical machine learning</td>
</tr>
</tbody>
</table>

Author Index

Keyword Index
PS-01

Workshop on Machine Learning in Earth System sciences

Dr. Tobias Weigel¹, Dr. Laurens Bouwer², Dr. David Greenberg³, Dr. Eduardo Zorita³, Prof. Corinna Schrum³, Prof. Thomas Ludwig¹

¹ DKRZ, Hamburg, Germany; ² GERICS, Hamburg, Germany; ³ HZG, Geesthacht, Germany

The Helmholtz AI local unit AIM will bring Artificial Intelligence (AI) and Machine Learning (ML) innovations to applications within Earth System modelling and analytics. The AIM research team will be located at the Helmholtz-Zentrum Geesthacht (HZG) and the consultant team will be located at the German Climate Computing Centre (DKRZ). In order to foster early collaboration with researchers and potential users, DKRZ and HZG (Institute of Coastal Research and GERICS) co-organized a workshop on machine learning in Earth system sciences, held on February 3 and 4 at DKRZ in Hamburg. More than 50 researchers from Earth system science and computer science came together to share their experiences with various methods, discuss applications in various fields of earth sciences, and consider future challenges in the application of AI and ML to Earth system modelling and analytics. The audience represented a wide range of interests and institutions, including several Max Planck institutes and Helmholtz centers, the European Centre for Medium-range Weather Forecasts (ECMWF), and national and international universities.

The workshop featured presentation and break-out discussion sessions. Presentations focused on a range of current activities and future directions, including the substitution of existing parameterization schemes in Earth system models with trained Artificial Neural Networks, the challenges of handling statistical uncertainty, causality, prediction biases and explainability, and tools and services that lower technical barriers to implementation. It was widely felt that the community needs to move beyond the exploratory phase of applying the new techniques to specific problems, and proceed towards reaping practically relevant, measurable benefits from integrating the new methods in model execution and analytics workflows in the coming years.

The workshop provided valuable input for the AIM consultant team and the research group at HZG in terms of initial scope and collaboration opportunities. The tasks of the consultant team include building knowledge about the methods useful to researchers in Earth system sciences, providing suitable work environments and helping researchers bring their applications to these environments, which includes integration of computing and data resources. Future activities as a result of the workshop can be expected, including the possibility for a follow-up workshop with a wider audience and a more extensive programme.

Keywords: Outreach, Workshop, Earth and Environment, AIM, Earth System Science
PS-02

A Connectivity-based Psychometric Prediction Framework for Brain-behavior Relationship Studies

PhD/MD student Jianxiao Wu¹, Prof. Simon Eickhoff¹, Dr. Felix Hoffstaedter¹, Dr. Kaustubh R. Patil¹, Prof. Holger Schwender², Dr. Sarah Genon¹

¹ Forschungszentrum Julich, Institute of Neuroscience and Medicine, Jülich, Germany; ² Heinrich-Heine University, Mathematical Institute, Düsseldorf, Germany

The recent availability of population-based studies with standard neuroimaging measurements and extensive psychometric characterization opens promising perspectives to investigate the relationships between interindividual variability in brain regions’ connectivity and behavioral phenotypes. However, the multivariate nature of the prediction model based on connectivity within a network of brain regions severely limits the interpretation of the brain-behavior patterns from a cognitive neuroscience perspective. To address this issue, we here propose a connectivity-based psychometric prediction (CBPP) framework based on individual region’s connectivity profile. Preliminary to the development of this region-wise machine learning approach, we performed an extensive assessment of the general CBPP framework based on whole-brain connectivity information. Because a systematic evaluation of different parameters was lacking from previous literature, we evaluated several approaches pertaining to the different steps of a CBPP study. We hence tested 72 different approach combinations in a cohort of over 900 healthy adults across 98 psychometric variables. Overall, our extensive evaluation combined to an innovative region-wise machine learning approach, offers a framework that optimizes both, prediction performance and neurobiological validity (and hence interpretability) to study brain-behavior relationships.

Keywords: Brain, MRI, Behavior, Prediction, Connectivity
PS-03

Machine Learning at GEOMAR - from sparse measurements to global maps

Dr. Timm Schoening, On behalf of all machine learners at GEOMAR.

GEOMAR, DeepSea Monitoring, Kiel, Germany

This poster presents selected machine learning use cases developed at GEOMAR. Included are unsupervised pattern recognition examples and supervised methods for various biological, geological and anthropogenic missions. The poster looks at monitoring hazards from munition dumped into the oceans, extracting information on deep sea fauna from camera footage, providing global maps of geochemical parameters, assessing deep sea high-tech metal abundance and enabling high-performance computing at sea. AI enables ocean scientists at GEOMAR to link heterogeneous datasets of diverging resolution and coverage regarding time and/or space. It provides a connection between the sparse sampling methods that can access the deep oceans and the regional and global maps that are needed to communicate insights with context. The use cases presented here are just a glimpse into the ongoing AI activities at GEOMAR. Further applications are machine learning for model tuning and speed-up, high-throughput plankton identification combining AI and crowdsourcing and developing the next generation of marine robots that will bring AI into the deep ocean for truly deep learning.

Keywords: Machine Learning, Prediction, Detection, Monitoring, Sparse Measurements, Global Maps
Deep Learning speeds up gapless cytoarchitectonic mapping in the human brain

M.Sc./M.A. Christian Schiffer¹, M.Sc./M.A. Hannah Spitzer¹, M.Sc./M.A. Kai Kiwitz², Prof. Katrin Amunts¹,², Dr. Timo Dickscheid¹

¹ Forschungszentrum Jülich, Institute of Neuroscience and Medicine (INM-1), Jülich, Germany; ² University Hospital Düsseldorf, Heinrich Heine University Düsseldorf, Cécile & Oskar Vogt Institute of Brain Research, Medical Faculty, Düsseldorf, Germany

Multilevel human brain atlases are fundamental tools in neuroscience, helping to analyse and better understand the structure and functional organization of the brain. A crucial part of such atlases are delineations of cytoarchitectonic areas, which are considered a gold standard in structural parcellations. They are characterized by variations of distribution, size, density and shape of neurons in the cerebral cortex, which are analyzed in microscopic scans of brain tissue sections, with a resolution of 1 micrometer. Currently employed well-established semi-automatic and observer-independent methods [1] identify boundaries between adjacent areas based on statistical differences along the cortical ribbon. This procedure is precise and reliable, but cannot keep up with the acquisition speed of modern high-throughput microscopes. We develop Deep Learning methods for automated cytoarchitectonic mapping with human level precision, and have already demonstrated that automated cytoarchitectonic mapping is possible [2]. Yet, results are not accurate enough to fully automate the mapping process.

We propose a method to train specialized neural networks for precisely segmenting individual areas in local parts of the brain with minimal training data. We train specialized models on pairs of spatially close, manually annotated sections. Each model specializes on the local properties of cytoarchitecture and tissue morphology in the respective local region, and focuses on a single selected cytoarchitectonic area. Trained models can then be used to automatically segment all sections between the training sections - the “annotation gap” - with high accuracy.

We successfully applied our method for nine areas in the BigBrain dataset [3], and achieved a gapless labeling of these areas. Based on the resulting highly detailed stack of section-wise segmentations, we computed high-resolution 3D maps at 20 micron resolution by exploiting the existing 3D reconstruction of the BigBrain dataset.


Keywords: Deep Learning, Cytoarchitecture, Segmentation, Human Brain, Postmortem, Neuroanatomy, BigBrain, CNN, Computer Vision, Histology
Helmholtz AI at Forschungszentrum Jülich for research field Key Technologies / Information

Dr. Susanne Wenzel\textsuperscript{1}, Dr. Timo Dickscheid\textsuperscript{1}, Dr. Jenia Jitsev\textsuperscript{2}, Prof. Morris Riedel\textsuperscript{2}

\textsuperscript{1}Forschungszentrum Jülich, Institute of Neuroscience and Medicine, Structural and functional organisation of the brain (INM-1), Jülich, Germany; \textsuperscript{2}Forschungszentrum Jülich, Jülich Supercomputing Centre, Jülich, Germany

Helmholtz AI at FZJ addresses the current transformation of science regarding different aspects of digitization especially at the overlap of AI with high-performance computing (HPC), and neuroscience. The unit is built on the long and intense interdisciplinary partnership between Juelich Supercomputing Centre (JSC) and Institute of Neuroscience and Medicine, Structural and functional organisation of the brain (INM-1) within the Helmholtz Programmes ‘Decoding the Human Brain’ and ‘Supercomputing and Big Data’, which has already become the main driver of the European research flagship “Human Brain Project”. The focus of Jülich’s Helmholtz AI unit will be on robust deep learning methods for high-resolution scientific image analysis, as well as on large-scale, self-organized continual learning transferable across different tasks and domains. Driven by high-throughput data acquisition and the ambition to quickly transfer learned knowledge across tasks and scientific domains, the implementation of AI methods on large-scale high performance computing systems is a key aspect of this work.

Helmholtz AI at FZJ will be implemented with a research group at INM-1 and a tandem of groups at JSC that cover both research and research support. The research group at INM-1 will focus on Biomedical Computer Vision, especially deep learning methods for analyzing large and complex scientific image data with limited availability of training examples. Driven by continuously increasing image resolutions and data volumes in high-throughput settings, methods for distributed operation on HPC systems will be developed. To ensure a close exchange of research and support, the team at JSC consists of the research-oriented Cross Sectional Team Deep Learning (CST-DL) and the software development and research support-oriented High Level Support Team (HLST). The research focus and long-term agenda driven by JSC in the field of AI will be on enabling large-scale self-organized continual learning in multi-task scenarios. This line of research will create methods capable of growing generic models from incoming streams of data, extracting knowledge and skills quickly transferable across different tasks and domains - a still grand, open scientific question. The activity will be strongly dedicated to open science and open source software, making all the results transparent and all the tools available to scientific communities and public.

Keywords: Neuroscience, HPC, Biomedical Computer Vision, Robust Deep Learning, Continual Learning, Multi-task Learning, Active Learning, Transfer Learning, Large-scale Distributed Learning
PS-06

**HeAT - Helmholtz Analytics Toolkit**

**Dr. Markus Götz**¹, Dr. Martin Siggel², Dr. Kai Krajsek³, Prof. Achim Streit¹, M.Sc./M.A. Daniel Coque-лин¹³, Dr. Charlotte Debus², M.Sc./M.A. Björn Hagemeier³, Dr. Claudia Comito³, Dr. Philipp Knechtges², M.Sc./M.A. Michael Tarnawa³, Simon Hanselmann¹, On behalf of the Helmholtz Analytics Framework (HAF)

¹ Karlsruhe Institute of Technology (KIT), Steinbuch Centre for Computing (SCC), Eggenstein-Leopoldhafen, Germany; ² German Aerospace Centre (DLR), Simulation and Software Technology (SC), Cologne, Germany; ³ Forschungszentrum Jülich (FZJ), Juelich Supercomputing Centre (JSC), Jülich, Germany

We present HeAT - a programming framework for large-scale data analysis and machine learning on high-performance computing (HPC) systems. Building on top of PyTorch it provides common features like CPU and GPU support, highly-optimized linear algebra routines and imperative eager execution. On top of that, HeAT provides additional functionality like CUDA-aware MPI support, distributed automatic differentiation and automatic data communication to enable high-performance machine learning modelling. The NumPy-like interface allows effortless porting of existing code to any HPC near you. We demonstrate HeAT's superior processing capabilities based on the clustering analysis of rocket combustion data and deep image classification.

Keywords: AI, Machine Learning, HPC, GPU, Distributed Deep Learning, Clustering, HeAT, Rocket Science, Open Source
PS-07

Not one model fits all: unfairness in RSFC-based prediction of behavioral data in African American

Dr. Jingwei Li¹,³, Dr. Danilo Bzdok², Dr. Avram Holmes⁴, Dr. Thomas Yeo³, Dr. Sarah Genon¹

¹ Forschungszentrum Jülich, Institute of Neuroscience and Medicine, Jülich, Germany; ² McGill University, Department of Biomedical Imaging, Montreal, Canada; ³ National University of Singapore, ECE, CSC, CIRC, N.1 & MNP, Singapore, Singapore; ⁴ Yale University, New Haven, USA

While predictive models are expected to play a major role in personalized medicine approaches in the future, biases towards specific population groups have been evidenced, hence raising concerns about the risks of unfairness of machine learning algorithms. As great hopes and intense work have been invested recently in the prediction of behavioral phenotypes based on brain resting-state functional connectivity (RSFC), we here examined potential differences in RSFC-based predictive models of behavioral data between African American (AA) and White American (WA) samples matched for the main demographic, anthropometric, behavioral and in-scanner motion variables.

We used resting-fMRI data with 58 behavioral measures of 953 subjects comprising 130 African American (AA) and 724 White American (WA). For each subject, a 419 x 419 matrix summarizing connectivity of 419 brain regions was computed.

Matching between AA and WA was performed at the subject level by creating 102 pairs of AA and WA subjects, matched for 6 types of variables (age, sex, intracranial volume, education, in-scanner motion and behavioral scores). We performed 10-fold nested cross-validation by randomly splitting the 102 pairs across 10 sets. The remaining 749 subjects were also divided across the 10 sets. A predictive model was built for each behavioral variable by using kernel ridge regression.

All analyses focused on the 102 matched AA and WA groups. After FDR correction (q < 0.05), no significant difference was found between the matched AA and WA groups for the matching variables.

Out of 58 behavioral variables, 38 showed significantly above chance prediction accuracies (based on permutation test, FDR corrected). Overall, average prediction performance for these variables was higher in the WA group than in the AA group. Furthermore, significant differences in prediction performance between the two groups were found in 35 behavioral variables (FDR corrected; q < 0.05).

Our results suggest that RSFC-based prediction models of behavioral phenotype trained on the entire HCP population show different prediction performance in different subsets of the population. This suggest that one model might not fit all that, in some cases, RSFC-based predictive models might have poorer prediction accuracies for African Americans compared to matched White Americans. Future work should evaluate the factors contributing to these discrepancies and the potential consequences, as well as possible recommendations.

Keywords: Brain, MRI, Behavior, Prediction, Fairness
PS-08
ML applications in High Energy Physics

Dr. Kirill Grevtsov, Priv.-Doz. Judith Katzy

DESY, Hamburg, Germany

One of the main objectives of particle physics is to exploit the full physics potential of the Large Hadron Collider (LHC) and its upgrade, the high luminosity LHC (HL-LHC). The physics reach of the experiments strongly depends on the performance of algorithms and computational resources used to explore the detector capabilities. Machine learning (ML) algorithms are crucial in this effort: it is widely used in pattern recognition, event and object reconstruction, calibration and as an analysis technique to obtain high-level physics results. ML is essential in searches for rare physics processes of interest (signal), which are hidden behind overwhelming amounts of other physics processes with similar signature (background) - techniques like Boosted Decision Tree or Neural Network are employed to discriminate these events.

The LHC experiments profit from very detailed simulated Monte Carlo (MC) data sets - large statistics labeled data, orders of magnitude higher than available unlabeled experimental data. These simulated samples are used for developing, testing and optimization of the ML algorithms.

In this material, we present an overview of the ML use-cases in the ATLAS experiment and discuss the prospects for future applications.

Keywords: Machine Learning, High Energy Physics, LHC, ATLAS, Supervised Learning, Higgs Boson, Simulated Data, Neural Network, Boosted Decision Tree, Adversarial Network
PS-09

Helmholtz.AI For Matter at Helmholtz-Zentrum Dresden-Rossendorf

Peter Steinbach¹, Nico Hoffmann²

¹ Helmholtz-Zentrum Dresden-Rossendorf, Scientific Computing Group, Dept. of Information Services & Computing, Dresden, Germany; ² Helmholtz-Zentrum Dresden-Rossendorf, Institute for Radiation Physics, Dresden, Germany

Our Poster aspires to present the Helmholtz.AI Local Group at Helmholtz-Zentrum Dresden-Rossendorf. We would like to present a brief introduction to our center and its mission. In addition, the young investigator group and the consultant team is presented. With this, we aim to provide a basis for discussion and present ourselves to the community.

Keywords: HZDR, YIG, Consultants, Matter, AI
PS-10

Helmholtz.AI Consultants at Helmholtz-Zentrum Dresden-Rossendorf

Peter Steinbach

Helmholtz-Zentrum Dresden-Rossendorf, 2 Scientific Computing Group, Dept. of Information Services & Computing, Dresden, Germany

We provide an introduction to our ongoing work as Helmholtz.AI Consultants within the research Field Matter. The poster gives insights in our methods and typical datasets. This emphasizes challenges and opportunities when working with broad diversity of collaborators. We also highlight methods used and areas of expertise we like to specialize in.

Keywords: HZDR, Matter, Consultants, Denoising, Segmentation, Inference
PS-11

Helmholtz AI YIG on Digital Twinning for High Energy Density Science

Dr. Nico Hoffmann, Dr. Peter Steinbach

Helmholtz-Zentrum Dresden-Rossendorf, Dresden, Germany

Matter and Technologies (MT) denotes a major program of Helmholtz association’s research field Matter. MT primarily focuses on researching novel accelerator- as well as detector technologies that eventually result in compact plasma-based accelerators for science, medicine and industry. That research is typically carried out at large research facilities such as LHC (Switzerland), European-XFEL (Hamburg, Germany), or BESSY II (Berlin, Germany). The development of advanced accelerator technologies requires novel detectors that provide insights into the processes inside the accelerator. Detectors produce large amounts of imaging data at high frame rates that needs to be processed in order to actually understand the physics inside the accelerator. Prior large-scale plasma simulations already provide some knowledge about physical phenomena inside these accelerators, though experimental validation and improvement of these pre-simulations promote the understanding of physics under extreme conditions even further. Experimental validation is a challenging task since the reconstruction of detector images requires ultra-fast and reliable algorithms that solve ill-posed mathematical problems (phase retrieval). The young investigators group is going to research data-driven approaches for digital twinning that yield high quality surrogate models and reconstruction of ambiguous data at compelling speed. The joint usage of prior knowledge about the dynamics of physical systems as well as empirical knowledge learnt from large amounts of data for designing and training of neural networks is going to provide new means for fast and reliable comprehension physics under extreme conditions in theory and experiment.

Keywords: HZDR, YIG, Digital Twin, Machine Learning
PS-12

AI in Medical Image Computing

David Kügler¹, Leonie Henschel¹, Santiago Estrada¹, Kersten Diers¹, Martin Reuter¹,²

¹ German Center for Neurodegenerative Diseases, Image Analysis Group, Bonn, Germany; ² Harvard Medical School, Department of Radiology, Boston, USA

The transition of healthcare to a predictive paradigm hinges on reliable, trustworthy, explainable, and easy-to-use AI methods for rich data sources such as medical images. Our research focuses on the development of novel AI and computational methods to automatically evaluate multi-modal images for medical and clinical applications. Empowered by the increasing availability of large cohort studies, we aim to assess treatment response, to recognize early disease effects, and to identify risk and preserving factors of neuro-degenerative diseases. The following four examples illustrate our research agenda: 1) MRI segmentation, 2) MRI reconstruction, 3) Geometric Learning, and 4) Shape modeling.

Segmentation: Traditional neuroimaging pipelines involve computationally intensive, time-consuming optimization steps, and thus, do not scale well to large cohort studies. Through the introduction of spatial context, localized weights, and competition we developed a deep-learning based pipeline FastSurfer to automatically process structural human MRI brain scans in minutes and FatSegNet for adipose tissue segmentation in the Rhineland Study. Extensions focus on including an attention mechanism to improve cortical details in high-resolutional MRI.

Reconstruction: By undersampling during MRI data acquisition (compressed sensing) significant speed-ups are achieved requiring more involved MR reconstruction schemes. Our solution outperforms the state-of-the art by integrating complex operations and Fourier transformations into the neural network architecture.

Geometric Learning: Cortical neuroimaging data (thickness, functional activations, folding patterns, curvature, etc.) are typically analysed on the cortical surface using spherical coordinate systems on triangle meshes. We perform surface-based segmentation of the human cortex, where our 2D parameter space approach with view-aggregation (p³CNN) outperforms spherical CNNs with respect to accuracy.

Shape Models: Anatomical structures are inherently geometric objects. We show that descriptive local and global shape features are better suited at detecting early structural changes, such as lateral asymmetry, compared to traditional volume estimates. For example, quantifying local thickness in hippocampal substructures supports research into etiology of several neurodegenerative diseases.

Our interdisciplinary research provides validated and reliable methods with a direct impact on the applied medical field.

Keywords: Medical Imaging, Neuro-degenerative Diseases, Segmentation, Surface-based Deep Learning, Shape Modeling, MRI Reconstruction
Estimating Model Uncertainty of Neural Networks in Sparse Information Form

Jongseok Lee, Matthias Humt, Jiangxiang Feng, Rudolph Triebel

German Aerospace Center, Institute for Robotics and Mechatronics, Wessling, Germany

Whenever machine learning methods are used for safety-critical applications such as mobile robotics, autonomous driving, medical image analysis or space exploration, it is crucial to provide a precise estimation of the failure probability of the learned predictor. Therefore, most of the current learning approaches return distributions rather than single, most-likely predictions. However, in case of Deep Neural Networks (DNNs), this true failure probability tends to be severely underestimated, leading to overconfident predictions. The main reason for this is that DNNs are typically trained with a principle of maximum likelihood, neglecting their epistemic or model uncertainty with the point estimates of parameters.

Unfortunately, estimating model uncertainty in DNNs is computationally a challenging task due to a large number of parameters and size of data-sets. Consequently, approximate inference on DNNs posterior often rely on mean field approximations or simplify the covariance matrix into Kronecker products of two smaller matrices regardless of the inference principles such as variational inference or Laplace approximation, even though there exists ample evidence that these simplifications induce crude approximation to the true parameter posterior.

To this end, we alternatively present a sparse representation of model uncertainty for DNNs where the parameter posterior is approximated with an inverse formulation of the Multivariate Normal Distribution (MND), also known as the information form. The key insight of our work is that the information matrix, i.e. the inverse of the covariance matrix, is sparse in its spectrum. Therefore, dimensionality reduction techniques such as low rank approximations can be effectively exploited. To achieve this, we develop a novel sparsification algorithm and derive a cost-effective analytical sampler. As a result, we show that the information form of MND can be scalably applied to represent model uncertainty in DNNs. In our experiments, we showcase the state-of-the-art performance in approximating information content of the parameters, competitiveness in predictive uncertainty estimation, scalability to large network structures and data-sets, and superiority in memory consumption amongst structured estimators to the true posterior.

Keywords: Bayesian Deep Learning, Robotics, Approximate Bayesian Inference
Deep Learning in Particle Physics for the CMS experiment at the LHC

David Brunner, Leonid Didukh, Dr. Dirk Krücker, Ashraf Mohamed, Dr. Isabell Melzer-Pellmann

Deutsches Elektronen-Synchrotron DESY, Hamburg, Germany

Deep Learning (DL) has become of essential importance for the CMS experiment at the Large Hadron Collider (LHC) at CERN. After a quick introduction of the experiment, we present two of the flagship applications of DL within the CMS collaboration and the newly started work on Graph networks within the DESY CMS group. Neural nets for the identification of b-quarks are used within the CMS experiment since 2015. In the last 2 years, complex networks build from convolutional and recurrent units had been constructed to provide the best available identification performance. Another application of DL in CMS is tagging highly boosted, hadronically decaying particles, e.g. top-quark pairs. Graph networks have become popular in recent years to describe non-image data. Geometric DL on point clouds is also a promising approach for the identification of tau-lepton jets.

Keywords: Particle Physics, Jet Identification, Deep Learning, Point Clouds, CMS, LHC
PS-15

Learning for Discovery - Significance as Loss

Dr. Adam Elwood, Dr. Dirk Krücker

Deutsches Elektronen-Synchrotron DESY, Hamburg, Germany

We introduce two new loss functions designed to directly optimize the statistical significance of the expected number of signal events when training neural networks to classify events as signal or background in the scenario of a search for new physics at a particle collider. The loss functions are designed to directly maximize commonly used estimates of the statistical significance, s/sqrt(s+b), and the Asimov estimate, ZA. We consider their use in a toy SUSY search with 30/fb of 14 TeV data collected at the LHC. In the case that the search for the SUSY model is dominated by systematic uncertainties, it is found that the loss function based on ZA can outperform the binary cross entropy in defining an optimal search region.

Keywords: Particle Physics, Deep Learning, Asimov Loss
Discovery of Magnesium Degradation Modulating Additives using Machine Learning Techniques

Tim Würger\textsuperscript{1,2}, Dr. Christian Feiler\textsuperscript{1}, Dr. Sviatlana Lamaka\textsuperscript{1}, Prof. Robert H. Meißner\textsuperscript{1,2}, Prof. Mikhail Zheludkevich\textsuperscript{1,3}

\textsuperscript{1} Helmholtz-Zentrum Geesthacht, Institute of Materials Research, MagIC - Magnesium Innovation Centre, Geesthacht, Germany; \textsuperscript{2} Hamburg University of Technology, Institute of Polymers and Composites, Hamburg, Germany; \textsuperscript{3} University of Kiel, Institute for Materials Science, Faculty of Engineering, Kiel, Germany

Magnesium (Mg) is a material with high potential for a wide range of applications in areas such as transport, energy, and medicine. However, a prerequisite to unlock the full potential of Mg-based materials is to gain control over the degradation properties as each target implementation demands degradation properties specific to the field of application. A promising approach to gain control over the degradation rate of Mg are so-called dissolution modulators.\textsuperscript{1} However, the vast number of small molecules with potentially useful properties (inhibitors or accelerators) renders current experimental discovery methods time- and resource-consuming. Fortunately, emerging computer-assisted methods can explore large areas of chemical space with less effort. Here we present two applications of data-driven machine learning techniques that facilitate the performance prediction of untested dissolution modulating additives based on data from an existing experimental database.\textsuperscript{1} Firstly, we show how density functional theory calculations and machine learning methods can work synergistically to generate robust and predictive models that recapitulate experimentally-derived corrosion inhibition efficiencies of small organic compounds for pure magnesium. We further validate our methods by predicting \textit{a priori} the corrosion modulation properties of seven hitherto untested small molecules and confirm the prediction in subsequent experiments.\textsuperscript{2} Secondly, we developed a concept based on high throughput calculations, combining dimensionality reduction algorithms and corrosion experiments. Physical properties of the molecules, such as the corrosion inhibition efficiency, are presented in a two-dimensional sketch map based on their molecular similarity, which allows to identify structure-property relationships. By using out-of-sample embedding, this approach can facilitate the search for new molecules with inhibiting or promoting degradation modulating properties and could thus contribute significantly to a better understanding of corrosion inhibition mechanisms.\textsuperscript{3}

Acknowledgement:
Funding by HZG MMDi IDEA project (Modulators of magnesium dissolution for biomedical applications) is gratefully acknowledged.

\textsuperscript{1}S. V. Lamaka et al., Corrosion Science 128 (2017) 224–240  
\textsuperscript{2}C. Feiler et al., Corrosion Science 163 (2020) 108245  
\textsuperscript{3}T. Würger et al., Frontiers in Materials 6 (2019) 53

Keywords: Molecular Dynamics, Structure-Property-Relationship, Density Functional Theory, Dimensionality Reduction, Artificial Neural Networks, Out-of-sample Embedding, Corrosion, Magnesium, Corrosion Inhibition
PS-17

Helmholtz AI Consultants for Medical Imaging Analysis

Christina Bukas

Helmholtz Muenchen, Helmholtz AI, Neuhrberg, Germany

The Helmholtz AI Central Unit in Munich works together with Helmholtz colleagues and associates to connect healthcare and computer science via modern machine and deep learning methods. Currently, the Helmholtz AI Consultant Team is in the recruitment phase and the first scientists have joined. Some pilot vouchers are already under way and new exciting AI projects have been defined as the current AI Consultants are expanding their network through collaborations with local Helmholtz Munich researchers and partners. AI Consultant Ario Sadafi recently published his work on detection of red blood cells using multiclass deep active learning. In addition, newly joined AI Consultant Christina Bukas has started working on extracting the pulmonary artery flow parameters from Magnetic Resonance Images to help explain the cause of Bronchopulmonary Dysplasia in newborns. A collaboration has been established with the German Mouse Clinic through which AI Consultants will be working on big data interpretation. Here, we plan on automatizing feature extraction from medical images of mice to enable unbiased objective measurements of biological features and detection of abnormalities. Through a close collaboration with the Institute of Computational Biology we have established a deep learning pipeline which will serve in automating and accelerating standard DL practices. Moreover, the outcomes of last month’s Hacking for Health Hackathon have been reviewed and discussed and AI Consultants will take part in the continuation of several health challenges. All the above demonstrates a dynamic start of the Helmholtz AI Consultant Team in Munich leading to wider application of AI methods on healthcare-relevant data sets.

Keywords: Segmentation, Reconstruction, Detection, Classification
Automatic Analysis of Cortical Areas in Whole Brain Histological Sections using Convolutional Neural Networks

Hannah Spitzer\textsuperscript{1,2}, Prof. Katrin Amunts\textsuperscript{1,3}, Prof. Stefan Harmeling\textsuperscript{4}, Dr. Timo Dickscheid\textsuperscript{1}

\textsuperscript{1} Forschungszentrum Juelich, Institute of Neuroscience and Medicine (INM-1), Juelich, Germany; \textsuperscript{2} Helmholtzzentrum Muenchen, Institute of Computational Biology (ICB), Munich, Germany; \textsuperscript{3} Heinrich-Heine University Duesseldorf, C. and O. Vogt Institute of Brain Research, Duesseldorf, Germany; \textsuperscript{4} Heinrich-Heine University Duesseldorf, Institute of Computer Science, Duesseldorf, Germany

Cytoarchitectonic parcellations of the human brain serve as anatomical references in multimodal atlas frameworks [1]. They are based on analysis of cell-body stained histological sections and the identification of borders between brain areas. The current standard to determine borders is based upon an observer-independent method that uses image analysis and multivariate statistical tools to capture changes in the distribution of cell bodies [2]. Nowadays, new technologies for high-throughput microscopy allow rapid digitization of histological sections, which increases the need for a fully automatic brain area segmentation method. This task is extremely challenging due to the high inter-individual variability in cortical folding, sectioning artifacts, limited labeled training data, and the need for large input sizes for automatic methods.

We have developed a Convolutional Neural Network model for semantic segmentation of cortical areas [3]. This model combines texture input given by high-resolution extracts of the histological sections with prior knowledge given by an existing probabilistic brain area atlas. This atlas prior helps the model to localize the texture input in the brain. To overcome the limited amount of brain area annotations, the model is pre-trained on a self-supervised auxiliary task based on predicting the 3D distance between image extracts from sections of the same brain [4]. This significantly increases the segmentation performance.

Using the self-supervised task alone, the model learns a compact internal feature representation of the input using the inherent structure of the brain, without having explicit access to the concept of brain areas. We show that these features allow to distinguish between several cortical areas – indicating that the self-supervised task is an appropriate auxiliary task for cytoarchitectonic mapping.


Keywords: Convolutional Neural Network, Deep Learning, Human Brain, Cytoarchitectonic Areas, BigBrain, Semantic Segmentation
PS-19

Repository Mining: from Code to README Generation

M.Sc./M.A. Roxanne El Baff¹, Dr. Tobias Hecking²

¹ German Aerospace Center (DLR), Intelligent Software Systems Group / Institute for Software Technology, Weßling, Oberpfaffenhofen, Germany; ² German Aerospace Center (DLR), Intelligent Software Systems Group / Institute for Software Technology, Köln, Germany

With more than 10 millions repositories holding software artifacts online, it is hard for developers to search semantically through piles of code using natural language. It is, however, easier to discover relevant repositories if the project is well documented. For example, if the source code in a repository contains documentation for each function and README/Wiki files explaining the features of the software, traditional information retrieval techniques such as vector space models can be used to match a user query to the text in the README or function documentation since both are written in natural language.

Problems for searching software repositories arise, however, if they are not or incompletely documented and the source code is the main available information. In order to support retrieval of such poorly documented software projects, a first challenge is to map the user’s query in natural language (e.g. English) to the source code in a project (e.g. Python), which requires creating a common semantic space between them. The issue of mapping natural language to source code has already been approached by using Deep Learning techniques (Husein, 2018), where programme functions and short documentation strings (docString) are transformed into each other. Since such docstrings only describe a small piece of code but do not explain the overall purpose and features of the software project. Thus a follow-up question is to what extent README files in natural language can be generated from given or generated docstrings?

We conduct experiments with language transformation models to translate software code into a natural language describing the main features of a software project. Our solution will (1) assist software developers in documenting their projects by automatically generating initial features descriptions from the source code itself and (2) improve indexing and retrieval of software repositories that lack sufficient documentation.

Our solution is divided into three consecutive phases: (1) The first phase deals with automating the process of extracting sentences describing software features from README files. (2) The second phase tackles the generation of features descriptions for a repository from code only. And (3) the third phase involves building an information retrieval system where we index all the generated features’ descriptions and retrieve repositories by mapping the user query to the indexed data.

Keywords: Natural Language Generation, Repository Mining
PS-20

Artificial intelligence to predict individual characteristics for personalized medicine

Dr. Kaustubh R. Patil1,2, Dr. Juergen Dukart1,2, Dr. Susanne Weis1,2, Dr. Robert Langner1,2, Prof. Simon Eickhoff1,2

1 Forschungszentrum Jülich, Institute of Neuroscience and Medicine (INM-7: Brain and Behaviour), Jülich, Germany; 2 Heinrich Heine University Düsseldorf, Institute of Systems Neuroscience, Düsseldorf, Germany

As cognitive, health and medical research move towards a more personalized viewpoint, artificial intelligence / machine learning (AI) methods are taking a center stage. Many promising research outcomes using AI methods are available and showcase their potential to make individualized predictions. Such methods can help in understanding inter-individual differences in cognition and behavior as well as in clinical conditions, and thus have a great potential for personalizing interventions at different levels, ranging from cognitive to pharmacological.

Several challenges remain in application of AI methods in these fields. Notably, limited generalizability of the models (i.e. low prediction accuracy on new data) remains a major hurdle arising mainly from scarcity of data and/or lack of proper measurements. We have developed methods to address these issues in systematic and biologically informative ways. Specifically, our methods integrate biologically-motivated feature reduction that improves the generalization of the models and makes the results more intuitively interpretable. To this end, we demonstrated successful use of parcel-wise prediction using resting-state functional connectivity (RSFC) neuroimaging data in predicting sex and schizophrenia sub-types. In both cases we performed functional decoding of the highly predictive brain regions to understand where the differences lie. We also highlight the use of literature-derived meta-analytic networks as priors for RSFC in predicting several clinical disorders and personality traits. Towards the aim of improving the scope of measurements, we show application of data derived identification of Parkinson’s disease from sensor-based smartphone assessments.

Taken together, we show several applications of AI methods to predict individual characteristics and clinical status. Such methods can shed light on inter-individual differences in health and improve monitoring and care for individual patients.

Keywords: Artificial Intelligence, Machine Learning, Neuroimaging, Neurological Disorders, Individual Predictions, Personalized Medicine, Sensors
Datasets and AI projects of the German Mouse Clinic

Dr. Holger Maier¹, Elida Schneizer¹, Isabella Galter¹, Christine Schütt¹, Dr. Helmut Fuchs¹, Dr. Valerie Gailus-Durner¹, Prof. Martin Hrabé de Angelis¹,²

¹ Helmholtz Zentrum München, German Mouse Clinic, Institute of Experimental Genetics, Neuherberg, Germany; ² Technische Universität München, Chair of Experimental Genetics, Freising, Germany; ³ German Center for Diabetes Research, Neuherberg, Germany

The German Mouse Clinic (GM) offers a large-scale standardised phenotypic platform for mouse models of human diseases. Mice are phenotyped in a clinical “check-up” with more than 600 parameters, representing the areas of behaviour, bone and cartilage development, neurology, clinical chemistry, eye development, immunology, allergy, energy metabolism, lung function, vision and pain perception, molecular phenotyping, cardiovascular analyses and pathology. The GMC collaborates with clinicians and scientists worldwide and is a founding member of the International Mouse Phenotyping Consortium (IMPC).

The unique data resource of the GMC hosts data of >10,000 mice, representing hundreds of mutant and wildtype strains. The complete phenotyping data set is captured from the same individual mice along their lifetime and available as raw data. Also, the complete demographic mouse data and metadata is available and can be linked to every single data point. Data is captured and managed in an ISO 9001:2015 – certified environment, according to standard operation procedures (SOPs), which have been developed within the IMPC.

AI offers new possibilities for analysis of raw data using the complete available dataset. In an ongoing project, bones of mice in X-ray images are automatically segmented and their features (length, shape, ...) are extracted. These features provide a rich, so far unavailable set of phenotyping data that can be used to further improve characterization of mutant mouse lines. In another ongoing project, also used as a challenge in a Helmholtz AI hackathon, digital histopathology slides are used for automated segmentation of subcutaneous white adipose tissue cells in order to extract their features. Again, these features will provide a rich, and so far unavailable source of phenotypic data. Examples of other raw data include cardiology (ECG and Echocardiography), auditory brainstem response (ABR), immunology (FACS), eye imaging (OCT) and energy metabolism (indirect calorimetry), where AI methods will be used to improve data analysis and scientific insight.

Keywords: Phenotyping, Mouse
PS-22

The Munich School for Data Science @ Helmholtz, TUM&LMU

Julia S. Schlehe¹, Fabian Theis¹,²

¹ Helmholtz Zentrum München für Gesundheit und Umwelt GmbH, Institute of Computational Biology, Neuherberg, Germany; ² Technische Universität München, Department of Mathematics, Garching bei München, Germany

The Munich School for Data Science (MUDS) trains the next generation of data scientists at the interface of data science and four different application domain sciences: biomedicine, plasma physics, earth observation, and robotics. Our research school strengthens the domain-driven research within the Helmholtz Association by teaching methodological data science skills in an interdisciplinary and application-oriented fashion. MUDS offers joint research projects designed by two partners, a domain-specific application partner and a methodological partner, both supervising the PhD student and therefore ensuring methodological as well as application specific education. At the metropolitan region of Munich, the universities TUM and LMU, and three regional Helmholtz centers (HMGU, IPP, DLR) joined forces for an internationally visible and highly attractive consortium at a prime location for computational sciences in Germany. Additionally, MUDS is cooperating with Roche Penzberg to promote application-oriented PhD projects in biomedicine, and is a member of the Helmholtz Information & Data Science Academy (HIDA).

Keywords: Research School, PhD, Data Science, Biomedicine, Plasma Physics, Earth Observation, Robotics Research School, Robotics
PS-23

Artificial Intelligence and Integration of Multi-omics Datasets for Individualised Infection Medicine

PhD/MD student Yunus Kuijpers, PhD/MD student Cancan Qi, PhD/MD student Xiaojing Chu, PhD/MD Bowen Zhang, PhD/MD Michael Beckstette, PhD/MD student Martin Graßhoff, PhD/MD Cheng-Jian Xu, Prof. Yang Li

Helmholtz-Centre for Infection Research (HZI) and Medizinische Hochschule Hannover (MHH), Centre for Individualised Infection Medicine (CiiM), Hannover, Germany

The variety between different individuals regarding different immune phenotypes such as autoimmune diseases or infectious disease risk can be attributed to both genetic and environmental factors (Li et al. 2016) (ter Horst et al. 2016). Data obtained from various different origins such as the microbiome, transcriptome, epigenome, metabolome, proteome, and other immune phenotypes can all be used in addition to the genotype data in order to predict various immune phenotypes.

These multi-omics datasets need to be integrated in a sensible fashion in order to extract the optimal amount of information to make a predictive model. The best way to do this is to incorporate them an artificial intelligence based model that best fits the purpose of the task at hand. Various examples of how AI and multi-omics data work well together are the prediction of stimulated cytokine response, allergic disease status prediction, and type 1 diabetes status prediction.

We found that multi-omics driven AI models were better at predicting stimulated cytokine expression, allergic disease status, and type 1 diabetes status. Integrating data from multiple categories besides just genotype data consistently outperformed models using exclusively genotype data. Furthermore by testing different models using various composition of input data we identified which datasources were most beneficial in predicting allergic disease status.

AI for individualized infection medicine will also be used to study and predict vaccination response.

Keywords: Elastic Net Regularization, Machine Learning, Multi-omics, Cytokines, Autoimmune Diseases, Supervised Learning, Classification
PS-24

Helmholtz Imaging Platform (HIP)

Prof. Christian G. Schroer\textsuperscript{1,2}, Prof. Volker Gülzow\textsuperscript{4}, Prof. Lena Maier-Hein\textsuperscript{3}, Priv.-Doz. Klaus Maier-Hein\textsuperscript{3}, Prof. Thoralf Niendorf\textsuperscript{5}, Dr. Stephan Preibisch\textsuperscript{5}

\textsuperscript{1} DESY, FS-PETRA, Hamburg, Germany; \textsuperscript{2} Universität Hamburg, Institut für Nanostruktur- und Festkörperphysik, Hamburg, Germany; \textsuperscript{3} DKFZ, Heidelberg, Germany; \textsuperscript{4} DESY, IT, Hamburg, Germany; \textsuperscript{5} MDC, Berlin, Germany

The Helmholtz Imaging Platform (HIP) brings scientists and engineers in the Helmholtz Association together to promote and develop imaging science and foster synergies across imaging modalities and applications within the Helmholtz Association. Any scientist or engineer within the Helmholtz Association can be a member of the HIP Network and can contribute to and benefit from the services provided by the platform. The central funding instruments are HIP Projects. They are granted to cross-disciplinary research teams to promote innovative imaging approaches. A first call for proposals will be issued in 2020. The results and imaging tools developed within HIP are made available to the HIP Network via the HIP Solutions framework. The Helmholtz Imaging Platform is currently being implemented by the HIP Core Team located at DESY, MDC, and DKFZ. The core team will run the platform and give scientific support along the full imaging pipeline, from data acquisition and image reconstruction (inverse problems) over data stitching, fusion and visualisation to manual and semi-automatic labelling of image data, automatic image analysis, validation and benchmarking.

Keywords: Helmholtz Imaging Platform, HIP Network, HIP Projects, HIP Solutions, Image Reconstruction, Data Fusion, Visualisation, Image Analysis, Validation
Coping with Concept Drifts in Load Forecasting using Machine Learning

M.Sc./M.A. Benedikt Heidrich, M.Sc./M.A. Marian Turowski, M.Sc./M.A. Nicole Ludwig, Prof. Ralf Mikut, Prof. Veit Hagenmeyer

Karlsruhe Institute of Technology, Institute for Automation and Applied Informatics, Eggenstein-Leopoldshafen, Germany

Network operators have to ensure the safety and stability of the electrical grid, which includes balancing the demand and generation of electricity. To achieve this balancing, the network operators need accurate forecasts for load time series.

In general, a variety of forecasting models for load time series exist, ranging from statistical approaches such as ARIMA to deep neural networks. However, the specific properties of load time series make many models unsuitable. For example, the models often assume that the underlying stochastic process of electricity demand is stationary since their parameters are time-invariant. However, often the stochastic process varies, which results in concept drifts. Load time series may contain gradual, sudden, and recurrent concept drifts. Gradual concept drifts can occur if the users’ behaviour changes over time, while sudden concept drifts can take place due to events such as the renovation of buildings. Finally, recurring concept drifts can appear in factories that operate in campaigns. Each of these drifts poses an unmanageable challenge for forecasting models with time-invariant model parameters.

The present poster emphasizes the importance of considering concept drifts during the design process of (deep learning) forecasting models. It also presents current state-of-the-art strategies for dealing with concept drifts. Additionally, we present a new approach to better cope with concept drifts that combines statistical information and deep convolutional neural networks. Lastly, the poster provides an outlook for further research directions towards more robust energy load forecasters.

Keywords: Machine Learning, Concept Drift, Load Forecasting, Neural Network
PS-26

Machine learning in soil research

Dr. Mareike Ließ

Helmholtz Centre for Environmental Research - UFZ, Soil System Science, Halle, Germany

Soils are at the centre of terrestrial ecosystems. They control water flows, nutrient storage and release. The multivariate soil parameter space forms the basis for the comprehensive understanding of terrestrial systems and matter flows by coupling it to models of plant growth and hydrology. Hence, landscape-scale high-resolution soil parameter fields generated by pedometric modelling approaches are required to answer a wide variety of research questions of high topicality and societal relevance. Overall, large data sets in the environmental and geosciences, and the complex relationships they contain, pose ever greater challenges that need to be met with adequate data science methods.

Pedometrics is an interdisciplinary science integrating soil science, applied mathematics/statistics and geoinformatics. The object of investigation is the spatial-temporal soil distribution at multiple scales. Modelling approaches are used along with multiple aspects of soil sensing and geodata analysis. Currently, the use of algorithms and optimization methods from the field of spatial data science is becoming increasingly important. The complexity of the task ranges from (A) single variable values at geographic point locations that are projected into the continuous 2-dimensional univariate parameter space up to (B) multivariate auto-correlated transect data (soil profile database) that need to be projected into the continuous 3-dimensional multivariate space. Targeting variables of high temporal dynamic requires the time dimension to be present in the point/transect data as well as the data cube of explanatory variables.

In pedometrics, machine learning is used in a wide variety of applications such as landscape-scale soil modelling, soil monitoring and sensing, or the development of pedotransfer functions. An example is presented concerning model development to predict topsoil texture of soils under agricultural use (Germany).

Keywords: Multi-target Machine Learning, Spatial Data Science, Optimization, Remote and Proximal Sensing, Geodata Analysis
PS-27

Adopting Conversational Interfaces for Exploring OSGi-based Software Architectures in Augmented Reality

M.Sc./M.A. Sivasurya Santhanam

German Aerospace Center (DLR), SC- Intelligent and Distributed Systems, Cologne, Germany

We propose Conversational user interfaces as a convenient and intuitive way for software developers to explore OSGi-based software architectures in immersive Augmented Reality (AR). By providing natural user interfaces, we aim to remedy some peculiarities of AR devices, but also enhancing the exploration task at hand. We present an architecture to integrate the conversational interfaces to the existing software visualization system. The proposed model has been implemented and tested in the Microsoft HoloLens device. We exemplify a use case and sketch how different user utterances can be used to retrieve information about the to-be-explored OSGi-based software architecture. We identify crucial components such as natural language generation and intent recognition which are required to implement the user story and we outline the status of our implementation.

Keywords: Conversational Interfaces, Chatbots, Augmented Reality, User Interfaces, Natural Language Processing, Software Visualization
PS-28

Machine Learning in Translational Genomics

Dr. Nigel W. Rayner, Prof. Eleftheria Zeggini

Helmholtz Muenchen, Institute of Translational Genomics, Muenchen, Germany

Biology and medicine are becoming ever more data-intensive with the prediction that genomics will equal or surpass other fields in data generation and usage within a decade.

To cope with this dramatic increase in data at the institute of translational genomics we are utilising a wide variety of machine learning techniques to further the translation of basic science to medically relevant outcomes.

We are using a wide variety of different techniques, such as Gaussian mixture models, two dimensional convolutional neural networks (2D CNN), shrunken centroid methods, elastic nets and random forest.

These are applied via existing tools, such as variant detection in high throughput sequencing which has for a long time been using machine learning techniques, for example our large scale sequencing utilises the Genome Analysis ToolKit (GATK) which has traditionally used a Gaussian mixture model to classify variants based on the clustering of their annotation values given a training set of high-confidence variants. The latest technique for variant quality recalibration in GATK is now based on a 2D CNN which outperforms its predecessor giving greater precision with no loss of sensitivity. Other variant calling methods such as the annotation of the Variant Call Format (VCF) data files are also using scores from a convolutional neural net.

We are also utilising a variety of machine learning methodologies to predict disease or disease progression in osteoarthritis and cardiometabolic diseases, such as:

• A nearest shrunken centroid method to construct a gene classifier that distinguishes between disease subtypes in osteoarthritis.

• A random forest based classifier to characterise methylation loci that have high predictive power to distinguish between intact and degraded cartilage in osteoarthritis.

• The use an elastic net to choose the best combination of variables in a Polygenic Risk Score (PRS) to predict disease occurrence or trait variance in cholesterol levels.

• The development of predictors of SNP function using a random forest method.

In future we are looking to expand the use of machine learning to all aspects of the work, especially that around the use of electronic health records and multi-omics data.

Keywords: Nearest Shrunken Centroid, Random Forest, Elastic Net, Gaussian Mixture Model, Convolutional Neural Network
PS-29

HIFIS: Helmholtz Infrastructure for Federated ICT Services

Dr. Uwe Jandt, on behalf of HIFIS

DESY, IT, Hamburg, Germany

The newly launched platform “Helmholtz Infrastructure for Federated ICT Services” (HIFIS) aims to build an outstanding federated IT infrastructure of the Helmholtz Association. Its ultimate goal is to combine the capabilities of all Helmholtz centres in order to promote science at a broad spectrum, ranging from large-scaled and computationally demanding projects to small-scaled and focused research works. The aim of this platform is to expand to all Helmholtz centres and prospectively to the national and international scientific community, including the National Research Data Infrastructure (NFDI) and the European Open Science Cloud (EOSC). To reach this goal, a federated approach was chosen from the beginning, pioneered by centres that already have considerable expertise in their respective fields. Key elements of this platform are the establishment of a trusted Virtual Private Network (VPN) between the centres and the installation of basic services such as trust relationships and a common authentication and authorisation infrastructure (AAI). Despite still in the early stages, the most important backbone infrastructures have already been technically established and basic HIFIS services are being implemented.

This Helmholtz backbone is precondition for higher-level services such as a Helmholtz Cloud, which is currently being established, with first implementation tests ongoing in several HIFIS centres. This includes collaborative services as well as data storage and retrieval, furthermore high performance computation clusters and documentation services. The specific portfolio of services is being determined and prioritized based on an extensive study involving all 19 Helmholtz centres. Its purpose is to integrate scientists as well as service providers from the very beginning and address their requirements and capabilities continuously.

On top of this, a platform is set up to support modern and sustainable software engineering methods and infrastructures. Developed software solutions are made accessible, visible and usable at a defined quality level via dedicated software repositories.

A first set of infrastructure services and training offers is available for the Helmholtz community, facilitating new scientific projects and generating additional value by making use of the synergies of federated and integrated IT services. This service platform to ensure a sustainable research software development is unique in the scientific landscape in Germany.

Keywords: Community Services, Information and Communication Technologies, Digitalization, Artificial Intelligence, Seamless Infrastructure, Sustainable Software
PS-30

The Helmholtz Metadata Collaboration (HMC) – addressing the metadata problem

Sören Lorenz¹, Michael Denker², Ants Finke³, Christian Langenbach⁴, Rainer Stotzka⁵, Frank Ückert⁶, Heike Görzig³, Tanja Höpker⁶, Wolfgang Suess⁵, Thomas Jejkal⁵, Emanuel Soeding¹

¹ GEOMAR Helmholtz Centre for Ocean Research Kiel, Kiel, Germany; ² Forschungszentrum Jülich, Jülich, Germany; ³ Helmholtz Zentrum Berlin, Berlin, Germany; ⁴ German Aerospace Center, Köln, Germany; ⁵ Karlsruhe Institute of Technology, Karlsruhe, Germany; ⁶ German Cancer Research Center, Heidelberg, Germany

The Helmholtz Association of German Research Centres and its international partners produce and exchange outstanding research data. Metadata are essential information for finding, understanding and reusing research data as outlined by the FAIR Principles. Researchers are, however, often challenged to collect, store and document their metadata in accordance with the common standards. To improve the metadata handling within its organization the Helmholtz Association recently launched a major initiative to establish the Helmholtz Metadata Collaboration (HMC) Platform. The HMC will provide comprehensive services, consulting, information and tools for an efficient handling of metadata. Community-Expertise to handle the metadata within the six Helmholtz research areas is jointly developed, shared and consolidated in the platform. Within the Helmholtz Association, the HMC will develop and establish key competences and services for these purposes.

The HMC platform is based on a distributed structure of discipline specific Metadata Hubs for each of the Helmholtz research fields. These hubs activate competences, nurture ideas and collect demands of their domains to develop solutions to current metadata challenges. The Metadata Hubs are accompanied by central service units, tasked with the technical developments, support and project logistics, to implement recommended solutions and activities derived from the hubs as services or tools. The technical units will make use of existing infrastructures and concepts where appropriate, e.g. FAIR data objects or existing standards, ontologies or vocabularies, and will develop new services when appropriate. Annual open project calls addressing metadata implementation challenges allow for an active participation of all Helmholtz researchers and their partners in the HMC. Generically usable processes, technical solutions as well as training, education and data consulting services are set up and made available to the Metadata Hubs for specific adaptation and utilization. Helmholtz Association’s ultimate goal is, to create a platform not only for its internal use, but to establish a public, open, long-term available community service handling metadata. This activity tackles one of the oldest and most pressing challenges in Data Science. We invite interested parties to share thoughts and ideas with us, in setting up such a platform.

Keywords: Metadata, FAIR, Digital Objects, Research Data
PS-31

Climate Change and Health Impacts: An Applied Use Case for Machine Learning

Lennart Marien¹, Mahyar Valizadeh², Dr. Wolfgang zu Castell², Dr. Alexandra Schneider², Dr. Kathrin Wolf², Dr. Diana Rechid¹, Dr. Laurens Bouwer¹

¹ Climate Service Center Germany (GERICS), Helmholtz-Zentrum Geesthacht (HZG), Hamburg, Germany; ² Helmholtz Zentrum München - Deutsches Forschungszentrum für Gesundheit und Umwelt (HMGU), München, Germany

Myocardial infarctions (MI) are a major cause of death worldwide. In addition to well-known individual risk factors, studies have suggested that temperature extremes, such as heat waves, may adversely affect MI (Chen et al., 2019). The frequency and intensity of heat waves is increasing due to climate change, and will likely increase further, even at levels limited to 1.5°- or 2°-degree global warming (Sieck et al., 2020). The relationship between health impacts and climate is very complex, depending on a multitude of climatic, environmental, sociodemographic and behavioral factors. Machine Learning (ML) is a powerful tool in investigating complex and unknown relationships between environmental conditions and their adverse impacts (Wagenaar et al., 2017). However, ML has seldomly been applied in a large multi-variable setting for studying health effects of heat waves. Combining heterogeneous health, climatic, environmental and socio-economic datasets this study is a first step towards modelling MI risk due to heat waves with ML.

We will develop ML algorithms based on data from the KORA cohort study (Holle et al., 2005) in the Augsburg region of Bavaria, Germany, comprising detailed information on MI and underlying health conditions. Additionally, weather and climate data (e.g., climate projections from the EURO-CORDEX initiative); air pollution data (e.g., PM₁₀, PM₂.₅, nitrogen oxides, ozone); building characteristics (e.g., type and age); and socio-economic data (e.g., household income, education) will be used for this study.

A small ensemble of ML models for multi-variable relationships between environmental conditions and health impacts will be developed. Careful validation will allow to assess model performance and to estimate the magnitude of generalization errors. Ultimately, we have the goal to project risks for future time periods making use of a detailed regional climate change projections as well as demographic and socio-economic projections.

Keywords: Myocardial Infarction, Heatwaves, Climate Projections, Machine Learning, Climate and Health, Climate Impacts, Digital Earth
We propose a network that predicts Fourier Descriptors to accurately regress object shapes in a fix sized vector form. We show that segmentation masks that can be derived from the descriptors produce adequate results in comparison to pixel based segmentation methods. We hypothesize that in many cases it is inherently more efficient to learn the distribution of possible shapes and derive pixel classes from there, than it is to learn the distribution of individual object pixels before inferring the shape. Hence, our approach is especially suitable for data that shows many instances with consistent shape characteristics. The learned object embedding aims to be invariant against changes in texture and translation and equivariant against transformations like scaling. We hypothesize that contour shape based instance segmentation suffers less under otherwise prevalent morphological inconsistencies produced by pixel based methods. Also, in contrast to pixel based methods, our approach is able to directly model overlapping instances. This property is crucial for our task of segmenting overlapping instances of neuronal cell bodies in histological sections of human brains.

Keywords: Deep Learning, Machine Learning, Object Detection, Instance Segmentation, Segmentation, Computer Vision, Histology
Helmholtz Information & Data Science Academy (HIDA)

M.Sc./M.A. Susan Trinitz

Helmholtz Information & Data Science Academy, Berlin, Germany

The Helmholtz Information & Data Science Academy (HIDA) is Germany’s largest post-graduate training network in the information and data sciences. We prepare the next generation of scientists for a data-heavy future of research. HIDA connects and serves as the roof to six data science research schools, which are linked by a network of 15 national labs and 17 top-tier universities across Germany. During the next five years, these data science research schools will train over 250 fully funded doctoral researchers. They will deepen their knowledge in data science methods and learn to combine knowledge from the six Helmholtz research areas – energy, earth and environment, health, aeronautics, space and transport, matter, and information – with data science methods. For these purposes, all doctoral researchers receive dual supervision in data science and their scientific domain. In addition, HIDA offers doctoral researchers and scientists attractive opportunities to obtain training and continuing education in a wide range of methods and to become part of an international data science network.

Keywords: Education, Training, Network, Information and Data Science Academy, HIDA
Deep learning enables *in silico* chemical-effect prediction

Dr. Jana Schor, Dr. Jörg Hackermüller

*Helmholtz Centre for Environmental Research, Molecular Systems Biology, Leipzig, Germany*

All living species are exposed to a plethora of chemical substances. In addition to food and endogenous chemicals there are drugs and pollutants. Many chemicals are associated to the risk of developing severe diseases due to their interaction with bio-molecules, like proteins or nucleic acids. Hundreds of thousands of chemicals are listed in public databases worldwide, and there are similarly many bio-molecules encoded in the genomes of species. The advances in high-throughput sequencing technologies in genomics and high-throughput robotic testing in toxicology provide a great source of complex data (for a fraction of chemicals) that must be integrated on a large scale. There is an increasing need for systematic prediction, for example to prioritize chemicals for risk assessment, to enable a smart selection of chemicals for monitoring or for sustainable design of future chemicals. We present our deepFPlearn approach that uses deep learning to associate the molecular structure of chemicals with target genes involved in endocrine disruption - an interference with the production, metabolism or action of hormones in the body which is associated to the development of many severe diseases and disorders in humans. A deep autoencoder has been trained to reduce the number of features that describe the molecular structure of chemicals. The compressed representation is then provided to a further deep neural network to predict the interaction with certain genes/molecular pathways. Trained on 7,248 thousand chemicals for which an interaction with 6 target genes of interest has been measured, the program reached 82% prediction accuracy. Its application to the 777,887 toxCast chemicals identified a plethora of additional candidates that might be involved in endocrine disruption in less than 2 minutes of computation time. With deepFPlearn we demonstrate that transforming the enormous quantity of data in genomics and toxicology into value using deep learning will pave the way towards predictive toxicology.

Keywords: deepFPlearn, Deep Learning, Predictive Toxicology, Autoencoder, Chemical Structure, Endocrine Disruption, Chemical Exposure
Applying AI to persistent knowledge graphs to learn on complex multi-omics data

**Dr. Daniel Lang**, M.Sc./M.A. Leon van Ess, M.Sc./M.A. Michael Seidel, Dr. Manuel Spannagl, Prof. Klaus F. Mayer

Helmhotz Center Munich, Plant Genome and Systems Biology (PGSB), Neuherberg, Germany

Many of the problems and tasks we encounter in our research in the fields of genomics, phylogenomics, pangenomics, population genomics and transcriptomics deal with the fundamental issue of identifying and understanding individual nodes or connected components of biological networks. The strategy to tackle these issues often involves graph-based algorithms and data structures. Whether it is obvious use cases like the study of genetic, interaction or gene regulatory networks or less obvious cases like phylogenetic inference, text mining, ontology annotation, the prediction of gene functions or the integration of data over multiple scales - graphs almost always offer powerful solutions. So far, most of these graphs are largely ephemeral and specific i.e. they are created for a specific problem, analyzed within a specific software framework, stored in specific formats/environments and utilized in a specific context. Subsequent query, analysis and integrating of multiple of these graphs to quickly address novel questions that go beyond the original intention are often impractical, laborious and thus mostly not pursued.

Persistent knowledge graphs as they already are applied in other IT domains would provide a perfect solution to this problem. Preliminary trials to apply graph databases and frameworks like neo4j and Apache Spark Graphx to our phylogenomics data sets have been promising, but did not yet scale to our data sets.

The combination of knowledge graphs and graph-based learning would provide an ideal toolkit to tackle the issues described above in one framework. The integration of results from multiple iterations, algorithms and evidence sources into single knowledge graph database, followed by the subsequent in-depth analysis and classification based on existing training sets would allow us to utilize this data-rich resource to employ graph-based learning approaches to extract and learn graph patterns and attributes for classification of orthologous and other gene family relationships on less feature-rich graphs of the same genes. Subsequently these trained models could be used to classify and iteratively integrate novel taxa more reliably. If successful, this approach could also be extended to solve other classification/clustering questions on biological networks e.g. the inference of GRNs or the reconstruction of metabolic pathways or prediction of gene functions etc.

Keywords: Knowledge Graphs, Graph Databases, Multi-omics, neo4j, Interaction Networks, Connected Components
PS-36

Machine Learning Group @ DLR Institute of Data Science

Dr. Anna Kruspe

DLR, Institute of Data Science, Jena, Germany

The DLR Institute of Data Science (Institut für Datenwissenschaften) in Jena was founded in 2017. It conducts research on a large variety of data science topics relevant within DLR and the broader research community. Its research directions include data management technologies, climate informatics, visual analysis, digital production platforms, secure software engineering, citizen science, and machine learning.

The machine learning group is situated at the junction of fundamental machine learning research and practical applications within the German Aerospace Center, such as computer vision for earth observation data, natural language processing for social media, and many more. It aims to be at the state of the art in deep learning, to further develop such methods, and determine how to put them into practice for DLR problems. Consequently, the group considers machine learning not just for a specific set of applications or data sets, but from a holistic perspective. As the group’s research is grounded in the requirements of DLR, they are frequently concerned with real-world data, not just academic corpora. Ongoing research topics include machine learning methods for high-resolution geodata, architectures for data fusion, approaches for small and noisy data sets, image and text mining from social networks, sparse and low-rank methods for machine learning models, and the integration of (physical) knowledge into machine learning models. Because the developed models are deployed in potentially critical real-world scenarios, the group develops these approaches in the context of quality assurance, i.e. making models more stable, certain, or understandable for their operators. Aspects of this include validation, stability, robustness, uncertainty estimation, interpretability, and explainability. The group is being established in cooperation with the DLR Remote Sensing Institute (Prof. Xiaoxiang Zhu and Prof. Richard Bamler).

Keywords: Deep Learning, Anomaly Detection, Uncertainty Modeling, Social Media, NLP, Computer Vision, Data Fusion, Knowledge Integration, Stability and Robustness, Interpretability and Explainability
PS-37

Helmholtz AI Voucher System

Dr. Mohammad Mirkazemi, Dr. Angela Jurik-Zeiller

HMGU, Munich, Germany

The Helmholtz AI voucher system offers scientists from all Helmholtz centres of the Helmholtz Association the opportunity to issue vouchers in order to enable AI cooperation with unused data sets and to spread know-how. The annual budget for the authorization of expenses for all vouchers amounts to 700,000 € and is provided by the Impulse and Networking Fund (INF) of the HGF President. Vouchers are issued in a competitive procedure in which they are accepted or rejected according to clearly defined guidelines and processed within the various Helmholtz AI units. AI Consultants will provide high-level support to users in a flexible, dynamic and transparent way. High-level support in general is defined as creative work that requires selection, adaptation and application of methods, and suggesting solutions that were not worked out already before by the partner.

Keywords: Vouchers, AI Consulting
PS-38

Spectro-image representation of DNA sequence for Deep Learning

Sabrina Martini¹, Dr. Carlos Garcia-Perez¹, Dr. Keiichi Ito¹, Prof. Wolfgang zu Castell¹, PhD/MD student Javier Betel¹

¹ helmholtz-muenchen, ICT, muenchen, Germany; ² helmholtz-muenchen, ICT, muenchen, Germany; ³ helmholtz-muenchen, ICT, muenchen, Germany; ⁴ helmholtz-muenchen, ICT, muenchen, Germany; ⁵ helmholtz-muenchen, ICT, muenchen, Germany

In many classification problems, input data representation is an important consideration that affects the performance of learning. Classification of bacterial genomes requires large amounts of data in training of the model which requires long training time on CPUs and GPUs. We propose a new representation of DNA sequence that takes advantages of recent image classification method called Convolutional Neural Network. In this poster we present our first results with short sequence of 16s rRNA as a proof of concept for the full sequence genome classification. The proposed method classifies this multiclass classification problem with very high accuracy.

Keywords: Deep Learning, Machine Learning, Genome
PS-39

Helmholtz AI Central Unit & Local Unit Health

**Dr. Christoph Feest**¹, Meike Kalb¹, Dr. Angela Jurik-Zeiller¹, Gemma Fornons¹, Dr. Mohammad Mirka-zemi¹, M.Sc./M.A. Christina Bukas¹, Ario Sadafi¹,², Andrea Mikulandra¹, Prof. Fabian Theis¹,²

¹ HMGU, Helmholtz AI Central Unit, Neuherberg, Germany; ² HMGU, Institute of Computational Biology, Neuherberg, Germany

Helmholtz Munich (HMGU) is home to the Central Unit of the Helmholtz Artificial Intelligence Cooperation Unit, which covers research field health. With its strong and expanding array of computationally-focused institutes (ICB), a dynamically growing digital health environment (ITG), and the bioengineering and imaging-focused HPC, HMGU is an ideal and well-connected host for the central unit.

The Helmholtz AI Science Management team at the Central Unit is integral to implementing and running Helmholtz AI, with its science and outreach managers coordinating platform activities and projects across all Helmholtz AI Units and within the Helmholtz Association that ensure wide dissemination of methods and that increase international visibility. The team is responsible for outreach and events, the Voucher System and Helmholtz AI Projects, as well as facilitating the meetings of the Helmholtz AI Steering Board and the Scientific Advisory Committee. Through liaison with the Association’s head office, its various bodies and other platforms of the Incubator Information and Data Science the Science Management team ensures a seamless flow of information, alignment with overall strategy and concise reporting. Helmholtz AI Central Unit also includes a team of health-focused AI Consultants, that will provide technical support and assessment to researchers across the Helmholtz Association through the Voucher System.

The website [www.helmholtz.ai](http://www.helmholtz.ai) serves as the central entry portal providing information on events, projects and funding calls, as well as the latest news (also on Twitter: @helmholtz_ai) and job offers at all units.

Keywords: Coordination, Outreach, Projects, Vouchers, Strategy
Helmholtz AI Cooperation Unit

Dr. Christoph Feest¹, Prof. Fabian Theis¹,²

¹ HMGU, Helmholtz AI, Neuherberg, Germany; ² HMGU, Institute of Computational Biology, Neuherberg, Germany

Our ambition at the Helmholtz AI Cooperation Unit is to reach an internationally visible leadership position in applied Artificial Intelligence (AI)/Machine Learning (ML) by combining unique research questions, data sets and expertise with newly developed AI/ML-based tools and democratized access to them in an open and dynamic community.

We are a research-driven hub for applied AI that i) fosters cross-field creativity, ii) identifies and leverages similarities between applications, iii) integrates field-specific excellence and AI/ML prowess, iv) improves the quality, scalability and timely availability of emerging methods and tools and v) empowers and trains the current and next generation of scientists.

We are structured as a hub-and-spoke model with six units across the Helmholtz Association; these units fund research groups and employ Helmholtz AI Consultant teams, ensuring a strong anchoring of our activities in all research fields. The central unit coordinates platform strategy and activities, particularly outreach, the voucher system and Helmholtz AI Projects, a funding line for collaborative projects.

Our goal is to enable the efficient and agile development and implementation of AI/ML assets across the whole Helmholtz Association.

Keywords: AI, Cooperation, Research, Projects, Outreach
The world's 8 million scientists must publish scientific articles in journals to advance their careers and share their work with their colleagues. With scientific careers dependent on corporate journals, publishers have managed to create an absurd, and absurdly expensive, system: 1) scientists donate their labor to publishers in the form of peer reviews that determine what does or does not get published; 2) authors donate their articles to publishers, who then turn around and sell access to these papers to institutions around the world; or 3) authors pay Open Access fees to publishers to get their articles published in journals, which are then made available for free to readers. Scientists provide the content, filter that content for correctness and quality, and then pay publishers for the privilege of having done all the work. The result is that scientists perform 14 million peer reviews annually, for free, which costs academia an estimated $2 billion dollars every year. And from this free labor, publishers generate $10 billion in annual revenues. Since the system serves publishers so well, they have little incentive to change. Current publishing workflows have hardly changed since the 1950s, despite technological advances which could greatly ease, and accelerate, scholarly communication. Authors suffer acutely from these inefficiencies. It takes an average of three months for an author to receive peer review feedback from a journal. And because of the outdated workflows, authors are forced to approach one journal at a time. Rejection at the first journal of choice leads authors to approach a second journal, or even a third and fourth, before they can get their work published, or even reviewed. These delays prevent scientists from getting validation of their ideas quickly, slow the pivot to new research questions, and unnecessarily inhibit the virtuous feedback circle that leads to scientific advance. Besides publishers, nobody is happy with scientific publishing, but there isn't currently a viable alternative that can effectively replace this expensive, extractive, and inefficient system. Meno is building that alternative: an AI-driven, peer-to-peer platform that radically drives down costs, accelerates and modernizes the entire publishing process, returns control of the publishing to scientists, and generates value for Meno, for scientists and their institutions, and for society.
Uncertainty quantification of soil moisture predictions
Dr. Miroslav Bačák, Dr. Martin Schrön, Dr. Stephan Thober, Dr. Hendrik Paasche

Helmholtz Centre for Environmental Research - UFZ, Leipzig, Germany

The agricultural drought, which represents one of the most impactful consequences of the climate change, is typically characterized by soil moisture values. However, direct measurements of soil moisture are very costly. Indeed, we have only very sparse soil moisture measurements obtained by Cosmic-Ray Neutron Sensing within the MOSES project at UFZ at our disposal.

While classic regionalization methods (e.g. kriging and its variants) provide reasonable predictions for interpolation problems when the available data is uniformly distributed across the area under consideration, they are of little practical use in the present project since we deal with extremely sparse and non-uniformly distributed data. That is why we use more sophisticated methods from machine learning and predict soil moisture from other quantities (for instance aspect, bulk density, elevation, precipitation, soil type), which are easier to measure. We employ several methods for regression including random forest and deep neural networks, however, there is at the moment no obvious choice among the existing machine learning approaches, which would capture both the spatial and temporal dimensions of our data. Finding the right method therefore represents a major challenge in the present project. Other questions are related to the input datasets. Which features are important for soil moisture predictions and which are not? What is the optimal sampling strategy for soil moisture measurements?

One of the key aspects of our project is that we want to quantify the prediction uncertainty. This is crucial for measuring the reliability and robustness of our models. To this end we need to know the uncertainties of the input data and investigate how they propagate through the model and get mixed with the model uncertainty.

Having a machine learning model for soil moisture predictions is of utmost importance. In addition, we can then compare our results with process-based soil moisture models, for instance with the Mesoscale Hydrologic Model (mHM), and use possibly our findings for the parameter calibration of these process-based models. This way we would like to contribute to the improvement of The German Drought Monitor.

Keywords: Uncertainty Quantification, Deep Learning, Random Forest, Regression
PS-43

BlenderProc

Maximilian Denninger, Martin Sundermeyer, Dominik Winkelbauer, Youssef Zidan, Dmitry Olefir, Ahsan Lodhi, Harinandan T. Katam, Mohamad Elbadrawy

German Aerospace Center (DLR), Institute of Robotics and Mechatronics, Weßling, Germany

BlenderProc is a modular procedural pipeline, which helps in generating real looking images for the training of convolutional neural networks. These can be used in a variety of use cases including segmentation, depth, normal and pose estimation and many others. A key feature of our extension of blender is the simple to use modular pipeline, which was designed to be easily extendable. By offering standard modules, which cover a variety of scenarios, we provide a starting point on which new modules can be created.

Keywords: Sim2Real, Machine Learning, Deep Learning, Robotics, Computer Vision, Pattern Recognition
PS-44

Deep Learning for Climate and Weather

Stan Posey, Dr. David Hall, Collaborations include NOAA, USA

NVIDIA, ESS Domain, Santa Clara, USA

Deep learning (DL) provides researchers with a new set of tools, complimentary to traditional physics-based models and statistical approaches. In several areas, deep learning has enabled the automated design of algorithms that far exceed the capabilities of those designed by experts. This is particularly true for computer vision tasks, but it has also been demonstrated in time-series analysis, language translation, speech synthesis, anomaly detection, strategy, autonomous navigation, and robotics. Such tools are new enough that many researchers remain unfamiliar with the potential scope of its capabilities. In this discussion, we will survey existing applications of deep learning in remote sensing and weather forecasting and discuss our own proof-of-concept projects designed to explore its potential. Example applications include extreme event detection, automated dataset translation, data assimilation, satellite simulation, image colorization, super-resolution, slow-motion, interpolation, physical parameterization emulation, de-noising, and many others.

Keywords: Deep Learning, Weather Model, Climate Model, Detection, Emulation, Down Scaling
PS-45

Helmholtz AI - MASTr: Munich @ Aeronautics, Space and Transport

Prof. Xiaoxiang Zhu¹, Dr. Rudolph Triebel², Dr. Martin Siggel³

¹ DLR, Remote Sensing Technology Institute (IMF), Wessling, Germany; ² DLR, Institute of Robotics and Mechatronics, Wessling, Germany; ³ DLR, Simulation and Software Technology, Köln, Germany

The Helmholtz AI Local Unit “MASTr: HAICU Munich @ Aeronautics, Space and Transport” consists of a Young Investigator Group (YIG) in the application domain Earth observation and an AI Consultant Team providing the expertise from Earth observation, robotics, computer vision and an HPC/HPDA support unit. It involves three DLR institutes:
- Remote Sensing Technology Institute (IMF), Oberpfaffenhofen – coordinator,
- Institute of Robotics and Mechatronics (RM), Oberpfaffenhofen,
- Facility Simulation and Software Technology (SC), Cologne.

The participating DLR institutes have been active in the field of machine learning and artificial intelligence in the last decade.

Keywords: Local Unit, AI Consulting, YIG, Earth Observation, AI4EO, Robotics, Data Mining, HPC/HPDA, Unsupervised Learning, Helmholtz AI
Deep learning for cell-type annotation tasks does not outperform classical machine learning

Maren Büttner, Niklas Köhler, Niry Andriamanga, Fabian Theis

Helmholtz Zentrum München, ICB, Neuherberg, Germany

Deep learning has revolutionized image analysis and natural language processing with remarkable accuracies in prediction tasks, such as image labeling or word identification. The origin of this revolution was arguably the deep learning approach by the Hinton lab in 2012, which halved the error rate of existing classifiers in the then 2-year-old ImageNet database. In hindsight, the combination of algorithmic and hardware advances with the appearance of large and well-labeled datasets has led up to this seminal contribution.

The emergence of large amounts of data from single-cell RNA-seq and the recent global effort to chart all cell types in the Human Cell Atlas has attracted an interest in deep-learning applications. However, all current approaches are unsupervised, i.e., learning of latent spaces without using any cell labels, even though supervised learning approaches are often more powerful in feature learning. Therefore, we investigate whether the increasingly large datasets and cell-type labels can be a playground for supervised deep learning in order to learn cell-type identity in new single-cell datasets. Notably, deep learning does not outperform classical machine-learning methods in the task. Cell-type prediction based on gene-signature derived cell-type labels is potentially a simplistic task for complex non-linear methods, and better labels of functional single-cell readouts are ideally required. We, therefore, are still waiting for the “ImageNet moment” in single-cell genomics.

Keywords: Deep Learning, Machine Learning, Automation, Single Cell Genomics
Author Index

A
Amunts, Katrin .................................. PS-32, PS-18, PS-04
Andriamanga, Niry ............................. PS-46

B
Bačák, Miroslav .................................. PS-42
Beckstette, Michael ............................. PS-23
Betel, Javier ....................................... PS-38
Bouwer, Laurens .................................. PS-31, PS-01
Brunner, David .................................... PS-14
Bukas, Christina .................................. PS-39, PS-17
Büttner, Maren .................................... PS-46
Bzdok, Danilo ...................................... PS-07

C
Chu, Xiaojing ................................... PS-23
Collaborations include NOAA, USA ........ PS-44
Comito, Claudia .................................. PS-06
Coquelin, Daniel .................................. PS-06

D
Debus, Charlotte .................................. PS-06
Denker, Michael .................................. PS-30
Denninger, Maximilian ......................... PS-43
Dickscheid, Timo .................................. PS-32, PS-18, PS-05, PS-04
Didukh, Leonid .................................. PS-14
Diers, Kersten .................................... PS-12
Dukart, Juergen .................................. PS-20

E
Eichhoff, Simon .................................. PS-20, PS-02
El Baff, Roxanne ................................. PS-19
Elbadrawy, Mohamad ............................ PS-43
Elwood, Adam .................................... PS-15
Estrada, Santiago .................. PS-40, PS-39, PS-16
Feiler, Christian ................................. PS-16
Feng, Jiangxiang ................................. PS-13
Finke, Ants ........................................ PS-30
Fomons, Gemma .................................. PS-39
Fuchs, Helmut .................................... PS-21

G
Gailus-Durner, Valerie ......................... PS-21
Galter, Isabella .................................. PS-21
Garcia-Perez, Carlos .......................... PS-38
Genon, Sarah ..................................... PS-07, PS-02
Görzig, Heike ..................................... PS-30
Götz, Markus ..................................... PS-06
Grafšhoff, Martin ............................... PS-23
Greenberg, David ............................... PS-01
Grevtsov, Kirill .................................. PS-08
Gülzow, Volker .................................. PS-24

H
Hackermüller, Jörg ............................. PS-34
Hagemeier, Björn ............................... PS-06
Hagemeyer, Veit .................................. PS-25
Hall, David ........................................ PS-44
Hanselman, Simon .............................. PS-06
Harmeling, Stefan .............................. PS-32, PS-18
Hecking, Tobias ................................. PS-19
Heidrich, Benedikt .................. PS-32, PS-25
Henschef, Leonie ............................... PS-12
Hoffmann, Nico .................................. PS-11, PS-09
Hoffstaedter, Felix ............................. PS-02
Holmes, Avram ................................. PS-07
Höpker, Tanja .................................... PS-30
Hrabě de Angelis, Martin .................... PS-21
Humt, Matthias ................................. PS-13

I
Ito, Keiichi ........................................ PS-38

J
Jandt, Uwe ........................................ PS-29
Jejakl, Thomas ................................. PS-30
Jitsev, Jenia ............................ PS-39, PS-37
Jurik-Zeiller, Angela ......................... PS-39, PS-37

K
Kalb, Meike ..................................... PS-39
Katam, Harinandan T. ......................... PS-43
Katzy, Judith ................................. PS-08
Kiwit, Kai ......................................... PS-04
Knechtges, Philipp .................. PS-32, PS-05
Köhler, Niklas ................................. PS-46
Krajsek, Kai ..................................... PS-06
Krücker, Dirk ............................. PS-15, PS-14
Kruspe, Anna ................................... PS-36
Kügler, David .................................. PS-12
Kuijpers, Yunus ................................. PS-23
POSTER ABSTRACTS

HELMHOLTZ AI KICK-OFF MEETING
5 MARCH 2020 - LENBACH PALAIS, MUNICH

L
Lamaka, Sviatlana ........................................PS-16
Lang, Daniel ................................................PS-35
Langenbach, Christian ................................PS-30
Langner, Robert ...........................................PS-20
Lee, Jongseok ..............................................PS-13
Li, Jingwei ................................................PS-07
Li, Yang ....................................................PS-23
Ließ, Mareike .........................................PS-26
Lodhi, Ahsan ..............................................PS-43
Lorenz, Sören ..........................................PS-30
Ludwig, Nicole .........................................PS-25
Ludwig, Thomas ..........................................PS-01

M
Maier, Holger .............................................PS-21
Maier-Hein, Klaus ......................................PS-24
Maier-Hein, Lena ........................................PS-24
Marien, Lennart .......................................PS-31
Martini, Sabrina .......................................PS-38
Mayer, Klaus F. .........................................PS-35
Meißner, Robert H. ....................................PS-16
Melzer-Pellmann, Isabella ............................PS-14
Mikulandra, Andrea ................................PS-39
Mikut, Ralf ................................................PS-25
Mirkazemi, Mohammad ..........................PS-39, PS-37
Mohamed, Ashraf ......................................PS-14

N
Niendorf, Thoralf ........................................PS-24

O
Olefir, Dmitry ..............................................PS-43
On behalf of all machine learners
at GEOMAR ...............................................PS-03
on behalf of HIFIS ......................................PS-29
On behalf of the Helmholtz Analytics
Framework (HAF) ......................................PS-06

P
Paasche, Hendrik ......................................PS-42
Patil, Kaustubh R ..................................PS-20, PS-02
Posey, Stan .................................................PS-44
Preibisch, Stephan .....................................PS-24

Q
Qi, Cancan ................................................PS-23

R
Rayner, Nigel W. ......................................PS-28
Rechid, Diana .............................................PS-31
Reuter, Martin ..........................................PS-12
Riedel, Morris .........................................PS-05

S
Sadafi, Ario ...............................................PS-39
Santhanam, Sivasurya ................................PS-27
Schiffer, Christian ......................................PS-04
Schlehe, Julia S .........................................PS-22
Schneider, Alexandra ................................PS-31
Schmelzer, Elida .........................................PS-21
Schoening, Timm .......................................PS-03
Schor, Jana ...............................................PS-34
Schoer, Christian G ....................................PS-24
Schrän, Martin ..........................................PS-42
Schrum, Corinna .......................................PS-01
Schütt, Christine ......................................PS-21
Schwarz-Romond, Thomas .......................PS-41
Schwender, Holger ....................................PS-02
Seidel, Michael ........................................PS-35
Siggel, Martin .......................................PS-45, PS-06
Soeding, Emanuel ....................................PS-30
Spannagl, Manuel ......................................PS-35
Spitzer, Hannah ........................................PS-18, PS-04
Steinbach, Peter ........................................PS-11, PS-10, PS-09
Stotzka, Rainer .........................................PS-30
Streit, Achim .............................................PS-06
Suess, Wolfgang .......................................PS-30
Sundermeyer, Martin ...............................PS-43

T
Tarnawa, Michael ......................................PS-06
Theis, Fabian ..................................PS-46, PS-40, PS-39, PS-22
Thofer, Stephan ........................................PS-42
Triebel, Rudolph ......................................PS-45, PS-13
Trinitz, Susan ...........................................PS-33
Turowski, Marian .....................................PS-25

U
Ückert, Frank .............................................PS-30
Upschulte, Eric ........................................PS-32

V
Valizadeh, Mahyar ..................................PS-31
van Ess, Leon ........................................PS-35
van Lanen, Edward ................................PS-41

W
Weigel, Tobias ...........................................PS-01
Wei, Susanne ...........................................PS-20
Wenzel, Susanne ......................................PS-05
Winkelbauer, Dominik ................................PS-43
Wolf, Kathrin ..............................................PS-31
Wu, Jianxiao ..............................................PS-02
Würger, Tim ..............................................PS-16
X
Xu, Cheng-Jian ...........................................PS-23
Y
Yeo, Thomas ..............................................PS-07
Z
Zeggini, Eleftheria .......................................PS-28
Zhang, Bowen ............................................PS-23
Zheludkevich, Mikhail ...................................PS-16
Zhu, Xiaoxiang ..........................................PS-45
Zidan, Youssef ...........................................PS-43
Zorita, Eduardo ..........................................PS-01
zu Castell, Wolfgang .....................................PS-38, PS-31
Keyword Index

A
Active Learning .................................. PS-05
Adversarial Network ........................ PS-08
AI ................................... PS-40, PS-09, PS-06
AI Consulting .................................. PS-45, PS-37
AI4EO .................................. PS-45
AIM .................................. PS-01
Anomaly Detection ................................ PS-36
Approximate Bayesian Inference ........ PS-13
Artificial Intelligence ...................... PS-29, PS-20
Artificial Neural Networks ........ PS-16
Asimov Loss .................................. PS-15
ATLAS .................................. PS-08
Augmented Reality .......................... PS-27
Autoencoder .................................. PS-34
Autoimmune Diseases ................ PS-23
Automation .................................. PS-46

B
Bayesian Deep Learning .................. PS-13
Behavior .................................. PS-07, PS-02
BigBrain .................................. PS-18, PS-04
Biomedical Computer Vision .......... PS-05
Biomedicine .................................. PS-22
Boosted Decision Tree ........ PS-08
Brain .................................. PS-07, PS-02

C
Chatbots .................................. PS-27
Chemical Exposure ........................ PS-34
Chemical Structure ................ PS-34
Classification ................................ PS-23, PS-17
Climate and Health ................ PS-31
Climate Impacts ................ PS-31
Climate Model ........................ PS-44
Climate Projections ................ PS-31
Clustering .................................. PS-06
CMS .................................. PS-14
CNN .................................. PS-04
Community Services ................ PS-29
Computer Vision ................ PS-43, PS-36, PS-32, PS-04
Concept Drift ........................ PS-25
Connected Components ................ PS-35
Connectivity ................................ PS-02
Consultants ................................ PS-10, PS-09
Continual Learning ................ PS-05
Conversational Interfaces ........ PS-27
Convolutional Neural Network ........ PS-28, PS-18

D
Data Fusion .................................. PS-36, PS-24
Data Mining .................................. PS-45
Data Science .................................. PS-22
Deep Learning .................. PS-46, PS-44, PS-43,
                                PS-42, PS-38, PS-36, PS-34,
                                PS-32, PS-18, PS-15, PS-14, PS-04
depFPlearn ................................ PS-34
Denoising .................................. PS-10
Density Functional Theory ........ PS-16
Detection ................................ PS-44, PS-17, PS-03
Digital Earth ........................ PS-31
Digital Objects ........................ PS-30
Digital Twin ........................ PS-11
Digitalization ................................ PS-29
Dimensionality Reduction ........ PS-16
Distributed Deep Learning .......... PS-06
Down Scaling ................................ PS-44

E
Earth and Environment .................. PS-01
Earth Observation ........................ PS-45, PS-22
Earth System Science ........ PS-01
Education .................................. PS-33
Elastic Net ........................ PS-28
Elastic Net Regularization .......... PS-23
Emulation ................................ PS-44
Endocrine Disruption ................ PS-34

F
FAIR .................................. PS-30
Fairness .................................. PS-07

G
Gaussian Mixture Model ........ PS-28
Genome .................................. PS-38
Geodata Analysis ................ PS-26
Global Maps ................................ PS-03
GPU .................................. PS-06
Graph Databases ................ PS-35
<table>
<thead>
<tr>
<th>H</th>
<th>PS numbering</th>
</tr>
</thead>
<tbody>
<tr>
<td>HeAT</td>
<td>PS-06</td>
</tr>
<tr>
<td>Heatwaves</td>
<td>PS-31</td>
</tr>
<tr>
<td>Helmholtz AI</td>
<td>PS-45</td>
</tr>
<tr>
<td>Helmholtz Imaging Platform</td>
<td>PS-24</td>
</tr>
<tr>
<td>HIDA</td>
<td>PS-33</td>
</tr>
<tr>
<td>Higgs Boson</td>
<td>PS-08</td>
</tr>
<tr>
<td>High Energy Physics</td>
<td>PS-08</td>
</tr>
<tr>
<td>HIP Network</td>
<td>PS-24</td>
</tr>
<tr>
<td>HIP Projects</td>
<td>PS-24</td>
</tr>
<tr>
<td>HIP Solutions</td>
<td>PS-24</td>
</tr>
<tr>
<td>Histology</td>
<td>PS-32, PS-04</td>
</tr>
<tr>
<td>HPC</td>
<td>PS-06, PS-05</td>
</tr>
<tr>
<td>HPC/HDPA</td>
<td>PS-45</td>
</tr>
<tr>
<td>Human Brain</td>
<td>PS-18, PS-04</td>
</tr>
<tr>
<td>HZDR</td>
<td>PS-11, PS-10, PS-09</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>I</th>
<th>PS numbering</th>
</tr>
</thead>
<tbody>
<tr>
<td>Image Analysis</td>
<td>PS-24</td>
</tr>
<tr>
<td>Image Reconstruction</td>
<td>PS-24</td>
</tr>
<tr>
<td>Individual Predictions</td>
<td>PS-20</td>
</tr>
<tr>
<td>Inference</td>
<td>PS-10</td>
</tr>
<tr>
<td>Information and Communication Technologies</td>
<td>PS-29</td>
</tr>
<tr>
<td>Information and Data Science Academy</td>
<td>PS-33</td>
</tr>
<tr>
<td>Instance Segmentation</td>
<td>PS-32</td>
</tr>
<tr>
<td>Interaction Networks</td>
<td>PS-35</td>
</tr>
<tr>
<td>Interpretability and Explainability</td>
<td>PS-36</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>J</th>
<th>PS numbering</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jet Identification</td>
<td>PS-14</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>K</th>
<th>PS numbering</th>
</tr>
</thead>
<tbody>
<tr>
<td>Knowledge Graphs</td>
<td>PS-35</td>
</tr>
<tr>
<td>Knowledge Integration</td>
<td>PS-36</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>L</th>
<th>PS numbering</th>
</tr>
</thead>
<tbody>
<tr>
<td>Large-scale Distributed Learning</td>
<td>PS-05</td>
</tr>
<tr>
<td>LHC</td>
<td>PS-14, PS-08</td>
</tr>
<tr>
<td>Load Forecasting</td>
<td>PS-25</td>
</tr>
<tr>
<td>Local Unit</td>
<td>PS-45</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>M</th>
<th>PS numbering</th>
</tr>
</thead>
<tbody>
<tr>
<td>Magnesium</td>
<td>PS-16</td>
</tr>
<tr>
<td>Matter</td>
<td>PS-10, PS-09</td>
</tr>
<tr>
<td>Medical Imaging</td>
<td>PS-12</td>
</tr>
<tr>
<td>Metadata</td>
<td>PS-30</td>
</tr>
<tr>
<td>Molecular Dynamics</td>
<td>PS-16</td>
</tr>
<tr>
<td>Monitoring</td>
<td>PS-03</td>
</tr>
<tr>
<td>Mouse</td>
<td>PS-21</td>
</tr>
<tr>
<td>MRI</td>
<td>PS-07, PS-02</td>
</tr>
<tr>
<td>MRI Reconstruction</td>
<td>PS-12</td>
</tr>
<tr>
<td>Multi-omics</td>
<td>PS-35, PS-23</td>
</tr>
<tr>
<td>Multi-target Machine Learning</td>
<td>PS-26</td>
</tr>
<tr>
<td>Multi-task Learning</td>
<td>PS-05</td>
</tr>
<tr>
<td>Myocardial Infarction</td>
<td>PS-31</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>N</th>
<th>PS numbering</th>
</tr>
</thead>
<tbody>
<tr>
<td>Natural Language Generation</td>
<td>PS-19</td>
</tr>
<tr>
<td>Natural Language Processing</td>
<td>PS-41, PS-27</td>
</tr>
<tr>
<td>Nearest Shrunken Centroid</td>
<td>PS-28</td>
</tr>
<tr>
<td>Network</td>
<td>PS-33</td>
</tr>
<tr>
<td>Neural Network</td>
<td>PS-25, PS-08</td>
</tr>
<tr>
<td>Neuroanatomy</td>
<td>PS-04</td>
</tr>
<tr>
<td>Neuro-degenerative Diseases</td>
<td>PS-12</td>
</tr>
<tr>
<td>Neuroimaging</td>
<td>PS-20</td>
</tr>
<tr>
<td>Neurological Disorders</td>
<td>PS-20</td>
</tr>
<tr>
<td>Neuroscience</td>
<td>PS-05</td>
</tr>
<tr>
<td>NLP</td>
<td>PS-36</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>O</th>
<th>PS numbering</th>
</tr>
</thead>
<tbody>
<tr>
<td>Object Detection</td>
<td>PS-32</td>
</tr>
<tr>
<td>Open Source</td>
<td>PS-06</td>
</tr>
<tr>
<td>Optimization</td>
<td>PS-26</td>
</tr>
<tr>
<td>Out-of-sample Embedding</td>
<td>PS-16</td>
</tr>
<tr>
<td>Outreach</td>
<td>PS-40, PS-39, PS-01</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>P</th>
<th>PS numbering</th>
</tr>
</thead>
<tbody>
<tr>
<td>Particle Physics</td>
<td>PS-15, PS-14</td>
</tr>
<tr>
<td>Pattern Recognition</td>
<td>PS-43</td>
</tr>
<tr>
<td>Peer Review</td>
<td>PS-41</td>
</tr>
<tr>
<td>Personalized Medicine</td>
<td>PS-20</td>
</tr>
<tr>
<td>PhD</td>
<td>PS-22</td>
</tr>
<tr>
<td>Phenotyping</td>
<td>PS-21</td>
</tr>
<tr>
<td>Plasma Physics</td>
<td>PS-22</td>
</tr>
<tr>
<td>Point Clouds</td>
<td>PS-14</td>
</tr>
<tr>
<td>Postmortem</td>
<td>PS-04</td>
</tr>
<tr>
<td>Prediction</td>
<td>PS-07, PS-03, PS-02</td>
</tr>
<tr>
<td>Predictive Toxicology</td>
<td>PS-34</td>
</tr>
<tr>
<td>Preprints</td>
<td>PS-41</td>
</tr>
<tr>
<td>Projects</td>
<td>PS-40, PS-39</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>R</th>
<th>PS numbering</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random Forest</td>
<td>PS-42, PS-28</td>
</tr>
<tr>
<td>Reconstruction</td>
<td>PS-17</td>
</tr>
<tr>
<td>Regression</td>
<td>PS-42</td>
</tr>
<tr>
<td>Remote and Proximal Sensing</td>
<td>PS-26</td>
</tr>
<tr>
<td>Repository Mining</td>
<td>PS-19</td>
</tr>
</tbody>
</table>
Research ................................................... PS-40
Research Data ......................................... PS-30
Research School ...................................... PS-22
Robotics .................................................. PS-45, PS-43, PS-22, PS-13
Robotics Research School ............................ PS-22
Robust Deep Learning ................................ PS-05
Robotics Research School ............................ PS-22
Robust Deep Learning ................................ PS-05
Scientific Publishing .................................. PS-41
Seamless Infrastructure ............................... PS-29
Segmentation .......................................... PS-32, PS-17, PS-12, PS-10, PS-04
Semantic Segmentation ................................ PS-18
Sensors ................................................... PS-20
Shape Modeling ....................................... PS-12
Sim2Real .................................................. PS-43
Simulated Data ........................................ PS-08
Single Cell Genomics ................................ PS-46
Social Media .......................................... PS-36
Software Visualization ............................... PS-27
Sparse Measurements ................................ PS-03
Spatial Data Science .................................. PS-26
Stability and Robustness ............................ PS-36
Strategy .................................................. PS-39
Structure-Property-Relationship .................. PS-16
Supervised Learning .................................. PS-23, PS-08
Surface-based Deep Learning ..................... PS-12
Sustainable Software ............................... PS-29

T
Training .................................................. PS-33
Transfer Learning ..................................... PS-05

U
Uncertainty Modeling .............................. PS-36
Uncertainty Quantification ......................... PS-42
Unsupervised Learning ............................ PS-45
User Interfaces ....................................... PS-27

V
Validation ........................................ PS-24
Visualisation ....................................... PS-24
Vouchers ........................................ PS-39, PS-37

W
Weather Model ....................................... PS-44
Workshop .......................................... PS-01

Y
YIG ................................................ PS-45, PS-11, PS-09