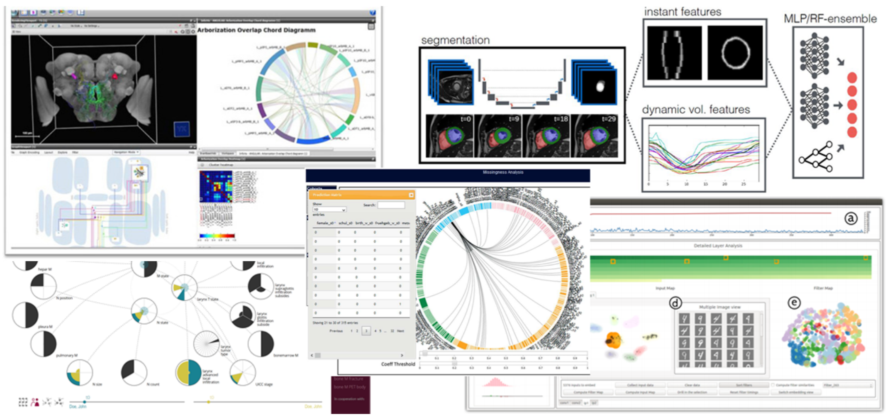
Blending Visualization with Data Mining and Machine Learning for Biomedical Data Analysis

Steffen Oeltze-Jafra, Anna Vilanova, Katja Bühler, Sandy Engelhardt, Nicola Pezzotti, and Bernhard Preim



1 Organizing Team

Steffen Oeltze-Jafra
University Leipzig, Innovation Center Computer Assisted Surgery (ICCAS), Germany & University of Magdeburg, Faculty of Computer Science, Germany
Steffen.Oeltze-Jafra@medizin.uni-leipzig.de
www.iccas.de/forschung/dpm

Steffen Oeltze-Jafra is Scientific Director of Digital Patient and Process Model (DPM) and DPM Group Leader at the Innovation Center Computer Assisted Surgery (ICCAS), University of Leipzig, Germany. He is also an associate professor at the Faculty of Computer Science at the University of Magdeburg, Germany. His research interests are in digital patient modeling, clinical decision support, and visual analytics of clinical, biological and epidemiological data. In 2016, Steffen received a habilitation (venia legendi) in computer science from the University of Magdeburg (UoM). In 2010, he received his Ph.D. in computer science and in 2004, his diploma (M.Sc.) degree in computational visualistics from the the same university. In 2017, Steffen won the Dirk Bartz Prize 2nd place for Visual Computing in Medicine from Eurographics. Between 2006 and 2015, Steffen has organized and participated in 13 tutorials and workshops at prestigious conferences in the fields of visualization, medical informatics, and medical image processing.

Anna Vilanova
Delft University of Technology, Computer Graphics & Visualization (CGV), The Netherlands
A.Vilanova@tudelft.nl
https://graphics.tudelft.nl/

Anna Vilanova is associate professor, since August 2013, at the Faculty of Electrical Engineering, Mathematics and Computer Science, Department of Intelligent Systems, Computer Graphics & Visualization Group at the University of Deft, the Netherlands. Before she was Assistant Professor at the Biomedical Image Analysis group of the Biomedical Engineering Department at the Eindhoven University of Technology. She is leading a research group in the subject of visual analytics and multivalued image analysis and visualization, focusing on high-dimensional biomedical data. Her research interests include visual analytics, medical visualization, volume visualization, multivalued visualization, and medical image analysis. In 2005, she was awarded a NWO-Veni personal grant with title ”Visualization of global tensor information for diffusion tensor imaging”. In 2014, she got a NWO-Aspasia. She has been chair and editor of relevant conferences and journals (e.g., EuroVis 2008, Computer & Graphics). She is member of the steering committee of IEEE VGTC EuroVis since 2014 and the EUROGRAPHICS executive committee since 2015.

Katja Bühler
Virtual Reality and Visualisation Research Centre GmbH (VRVis), Vienna, Austria
buehler@vrvis.at
www.vrvis.at/research/biomedical-image-informatics-group/

Katja Büler graduated in Mathematics at University of Karlsruhe (KIT) in 1996. She joined the Institute for Computer Graphics and Algorithms at TU Vienna from 1998-2002 as assistant professor and received a Ph.D. in Computer Science in 2001. In 2002, she continued her career as senior researcher at VRVis and as external lecturer at TU Vienna and became head of the Biomedical Image Informatics Group at VRVis in 2003. Since 2010, she is also managing the VRVis Research Area Complex Systems. Katja Bühler received the Austrian science2business award in 2012 for the management of the successful long-term research collaboration with AGFA Healthcare. Katja's research is focused on holistic approaches to provide intuitive and fast access to image information ranging from classical medical image analysis and deep learning to biomedical visualization and image data mining with a strong focus on computer supported radiology and neuroscientific applications. She is currently part of the management board of the Austrian Correlated Multi Modal Imaging Node and the steering committees of ”GI Fachgruppe Visual Computing in der Medizin” and the EG VCBM Workshop.

Sandy Engelhardt
University of Magdeburg, Department of Simulation and Graphics, Computer Assisted Surgery Group, Germany
engelhsa@isg.cs.uni-magdeburg.de
isgwww.cs.uni-magdeburg.de/cas/member.php?name=Engelhardt

Sandy Engelhardt studied Computational Visualistics at University of Koblenz-Landau and received the M.Sc. degree in 2012. From 2012 to 2016, she did her Ph.D. at the ”German Cancer Research Center” (DKFZ) in Heidelberg in the division of ”Medical and

Biological Informatics". During her Ph.D., she was part of the Collaborative Research Center SFB/TRR 125 Cognition-guided surgery and contributed to an assistance system for reconstructive mitral valve surgery. The system was awarded with the MICCAI AE-CAI Best Paper Award in 2014. Recently, her dissertation "Computer-assisted Quantitative Mitral Valve Surgery" granted the BVM-Award 2017, 1st place. At MICCAI 2017, she won the "Automated Cardiac Diagnosis Challenge"(ACDC) at the STACOM Workshop using a Deep Learning Approach. Her work mainly addresses topics of cardiology and computer-assisted heart surgery, thereby combining methods and technologies from the field of image segmentation, tracking systems and augmented reality. Sandy Engelhardt is currently Visiting Lecturer and deputy of the Computer-assisted Surgery Group at Magdeburg University.

Nicola Pezzotti
Delft University of Technology, Computer Graphics & Visualization (CGV), The Netherlands
N.Pezzotti@tudelft.nl
https://graphics.tudelft.nl/nicola-pezzotti/
Nicola Pezzotti received the Laurea Magistrale in Ingegneria Informatica (MSc) from the University of Brescia, Italy, in 2011. Previously he worked as a research fellow at the University of Brescia and as a research and development engineer in Open Technologies Srl. Since 2014, he is a PhD student at the Delft University of Technology in the Computer Graphics and Visualization group. His main research interest is in the development of efficient manifold learning algorithms for the exploratory analysis of large high-dimensional data. His algorithms are published in the most important data-visualization venues and are used for the analysis of Deep Learning architectures and in several medical research contexts. Nicola Pezzotti joined Google Research as an intern in 2018 for a research project focused on the visualization and improvement of Deep Neural Networks.

Bernhard Preim
University of Magdeburg, Department of Simulation and Graphics, Visualization Group, Germany
bernhard@isg.cs.uni-magdeburg.de
www.vismd.de
Bernhard Preim received the diploma in computer science in 1994 (minor in mathematics) and a Ph.D. in 1998 from the Otto-von-Guericke University of Magdeburg. In 1999, he joined MeVis Research, Bremen. In close collaboration with radiologists and surgeons he directed the work on "computer-aided planning in liver surgery". In June 2002, he received the Habilitation degree (venia legendi) from the University of Bremen. Since March 2003, he is full professor for "Visualization" at the Otto-von-Guericke-University of Magdeburg, heading a research group which is focused on medical visualization and visual analytics in healthcare. He authored the textbooks: "Visualization in Medicine" (Co-author Dirk Bartz, 2007) and "Visual Computing in Medicine" (Co-author: C. Botha, 2013). Bernhard Preim founded the working group Medical Visualization in the German Society for Computer Science and was president of the German Society for Computer- and Robot-Assisted Surgery from 2013 to 2015. He was Co-Chair and Co-Organizer of the first and second VCBM workshop and is now member of the steering committee of that workshop.

2 Tutorial Description

Groundbreaking results are achieved by neural networks in medical image processing. This success is reflected by a high and presumably increasing number of Deep Learning related submissions to the MICCAI conference. Despite the enormous progress in this field crucial challenges remain such as understanding the learned features, which often reside in high-dimensional space, as well as the tailor-made design and improvement of neural networks. In this tutorial, we will show how visualization can help in tackling these challenges. In traditional machine learning, features are designed instead of being learned, which is often a tedious process requiring skilled experts. We will also show how the feature design can be supported by blending visualization and data mining techniques. Causal networks supporting prognostic reasoning and discovering functional interactions are frequently being learned from complex biomedical and epidemiological data. We will show how visualization can assist an exploration and verification of the learned networks in order to, e.g., remove spurious dependencies.

Data mining systematically applies statistical and mathematical models to complex data in order to reveal patterns and trends and eventually, to infer knowledge from data. Many biomedical problems are being addressed using data mining techniques. Important challenges thereby are providing guidance to data mining laymen, e.g., physicians, in adjusting the parameters of a data mining algorithm and in interpreting its results, e.g., clusters in high-dimensional space or data sub-spaces. Again, visualization has been demonstrated to be beneficial in tackling theses challenges. It assists in understanding the parameter space and cluster structure.

In the tutorial, we address the blending of visualization with data mining and machine learning from a research and an application-oriented perspective. The latter focuses on cardiac surgery planning, understand gene-structure behavior in neurosciences, tumor tissue characterization, risk factor identification in epidemiology, and clinical decision support.

2.1 Part I - Introduction

We will start the tutorial with a brief introduction of the speakers, a presentation of the tutorial structure, and a motivation of the topic.

2.2 Part II - Image Analysis Meets Deep Learning and Visualization in Cardiology and Cardiac Surgery

Cardiac imaging improves on diagnosis of cardiovascular diseases by providing images at high spatiotemporal resolution. Manual evaluation of these time-series, however, is expensive and prone to biased and non-reproducible outcomes. At the same time, Convolutional neural networks (CNN) have recently shown outstanding performance in medical image processing. We present a method that addresses named limitations by integrating segmentation and disease classification into a fully automatic processing pipeline. We will reflect on our winning contribution for MICCAI's 2017 Automated Cardiac Diagnosis Challenge (STACOM) [8]. The fully automatic processing pipeline constitutes an attractive software for clinical decision support due to the visualization of segmentation maps, the comprehensive quantification of cardiologic assessment and the rapid processing speed of less than 40sec. Furthermore, we will show how the robustness of Deep Learning methods can aid for further image analysis and visualization tasks. In this regard, we will present novel visualizations for patient-specific shape modeling to predict response to pulmonary valve replacement in patients with congenital heart disease [6]. We will further highlight the great potential of using VR headsets for patient education and surgery planning considering the highly diverse pathomorphologies of congenital heart disease.

2.3 Part III - Visual Analytics from Feature Design to Deep Neural Networks Understanding

Traditional machine learning algorithms rely on carefully designed features by domain experts given a task at hand. The design of these features relies on the knowledge and skills of the designer. It is particularly difficult if a large number of variables must be considered, and there is few knowledge and no hypothesis to support this design. Visual Analytics, a branch of visualization that combines data-visualization with data-mining and data-manipulation, is a powerful tool that helps the domain expert in understanding the effectiveness of the designed feature spaces and formulate new hypotheses [17, 18]. However, since the groundbreaking results achieved in the object recognition field by Deep Neural Networks (DNN), the focus of machine learning algorithms has shifted from feature design to neural network design (i.e., feature learning). Understanding which features are learned by DNN is challenging, as they often reside in very high-dimensional spaces that are difficult to visualize and understand. Visual Analytics is the key to better understand, design and improve DNN [16].

2.4 Part IV - Genes-Structure-Behavior: (Visual) Data Mining and Data Fusion for Neurosciences

Neuroscience is a highly data driven science accumulating large data resources on brain structure, function and underlying genetics. Imaging is one of the main methods for acquiring information across the whole range of spatial and temporal resolutions. However, efficiently mining and exploiting such big imaging data resources and putting them in context with related genetic information and own results from experiments remains a challenge. We will give insight in frameworks that host and provide efficient means to visualize and mine ten thousands of 3D imaging datasets of the central nervous system of drosophila larvae, drosophila and zebrafish as well as heterogeneous network data of mice and related gene expression and functional imaging data [3]: https://braingazer.org/. Besides of efficient fully automatic information extraction and highly efficient spatial data structures, visualization, interface and interaction design have shown to be major factor for user acceptance and the realization of effective data mining workflows. We will close with recent work showing how new brain parcellations can be inferred by fusing big drosophila imaging data resources and how functional target areas can be derived by fusing information on behavior related genes with structural or functional connectomes [7, 15].

2.5 Part V - Visual Analytics of Epidemiological Cohort Study Data

Large-scale longitudinal epidemiological studies such as the Study of Health in Pomerania (SHIP) [20] investigate groups of people with common characteristics or experiences (a cohort) including socio-demographic and biological factors. Their goal is the characterization of health by identifying risk factors and their relations to diseases and the indication of a per subject risk of developing a disease. Carried out in waves over many years, they comprise thousands of individuals and ten thousands of variables. Unique for SHIP is the inclusion of medical image data acquired via an extensive full-body MRI protocol.

In this tutorial part, we describe our research at the intersection of visualization and data mining in exemplary SHIP projects dedicated to the characterization and differentiation of age-dependent and pathologic variance in spinal curvature and its relation to lower back pain and the identification of predictive features on hepatic steatosis. We present interactive visual analysis approaches that fit into the epidemiological workflow and facilitate hypothesis validation and generation [1, 2, 10]. The approaches support a bi-directional analysis of non-image and image data [10], data-driven subpopulation discovery [1], regression analysis for identifying disease-specific features [9], and handling missing data [2].

2.6 Part VI - Visual Verification and Exploration of Causal Networks for Clinical Decision Support

Causal networks and in particular, causal Bayesian networks are employed in biomedicine and healthcare for diagnostic and prognostic reasoning, treatment selection, and discovering functional interactions among components of a system. They are modeled by domain experts or learned from huge data collections generated and maintained in both domains. While sophisticated learning techniques exist, a verification of the resulting networks by domain experts is indispensable, e.g., to identify causalities among the learned dependencies and remove wrong dependencies resulting from data quality issues. The experts must verify 1) how the network operates and 2) what the network produces. Different information need to be highlighted in this course. In this tutorial part, we define requirements on a graphical network presentation supporting a visual verification. Moreover, we compare existing network representations with respect to these requirements [11].

As an example for causal networks, we will present a therapy decision model for laryngeal cancer treatment that is represented as a Bayesian causal network and has been designed for supporting decision making in clinical expert meetings [5, 19]. Bayesian networks have the advantage of comprising the complexity of multidisciplinary decisions and supporting clinicians with a transparent and reproducible decision inference. The model consists of over 900 variables with over 1300 causal dependencies representing patient information, clinical examinations, medical knowledge, clinical expertise as well as the respective interrelations. We will present a visual analysis component, which supports physicians in a model-based verification of clinical information, e.g., the Tumor, lymph Node, and Metastasis (TNM) staging, prior to the expert meeting as well as supports an understanding of the model's therapy recommendations during the meeting [4, 14].

2.7 Part VII - Closing Words

We will conclude the tutorial by final thoughts on unsolved problems and future trends in blending visualization with data mining and machine learning for biomedical data analysis. According to our experience from previous tutorials, instead of at the end of the tutorial, we will have sufficient room for questions and discussions with the audience at the end of each tutorial part.

3 Learning Objectives

The tutorial attendees will learn about:

1. the difference between feature design and feature learning,
2. visualization and visual analytics techniques assisting in
   (a) understanding, designing, and improving neural networks,
   (b) exploring and verifying causal networks,
   (c) understanding the parameter space of data ming methods and interpreting their results,
3. typical tasks in biomedical machine learning and data mining by several application examples,
4. existing frameworks and their integration in expert workflows,
5. and open problems and current research trends.

4 History/Background

The tutorial has been specifically designed for the MICCAI conference. Tutorial part V (Sec. 2.5) was part of two tutorials presented at the IEEE VIS conference in 2015 [13] and 2017 [12]. It has been updated for MICCAI and extended by a discussion of how data quality, in particular missing data, influences reasoning.

5 Duration and Format

The tutorial is planned as a half-day tutorial (4h30min) and will be given in a lecture-based format. All tutorial parts will comprise demonstrations. Tutorial part IV (Sec. 2.4) may include a hands-on session of the BrainGazer framework (https://braingazer.org/).

6 Preliminary Tutorial Schedule

- I Introduction, *Oeltze-Jafra* (10min)
- II Image Analysis Meets Deep Learning and Visualization in Cardiology and Cardiac Surgery, *Engelhardt* (45min)
- III Visual Analytics from Feature Design to Deep Neural Networks Understanding, *Vilanova and Pezzotti* (45min)
- — Break (30min)
- IV Genes-Structure-Behavior: (Visual) Data Mining and Data Fusion for Neurosciences, *Bühler* (45min)
- V Visual Analytics of Epidemiological Cohort Study Data, *Preim* (45min)
- VI Visual Verification and Exploration of Causal Networks for Clinical Decision Support, *Oeltze-Jafra* (40min)
- VII Closing Words, *All Presenters* (10min)

7 Target Audience and Number of Attendees

As a target audience, researchers and practitioners from both, the MIC and the CAI community are addressed since the tutorial comprises parts related to medical imaging and surgery planning. The tutorial should be both, inspiring for early career researchers, who explore new opportunities, as well as experts in the field, who are interested in catching up with new trends.

The tutorial is new and thus the number of participants involves a lot of uncertainty. The tutorial should be quite interesting based on the number of high-profile speakers. It will be advertised not only by the organizers but also by distribution lists of the German societies CURAC and BVM. We expect 40-50 people.

8 Proceedings

All slide decks of the speakers will be made available on the tutorial website. We aim to provide the material before the tutorial.

9 Specific Requirements

We require a projector, loud speakers, and microphones.

10 Other Comments

Anna Vilanova is a general chair of the VCBM workshop that will be co-located with MICCAI. In case our tutorial gets accepted, we would deeply appreciate if this dual commitment could be considered in the planning phase. If overlap is inevitable, Nicola Pezzotti will take over Anna's share of tutorial part III (Sec. 2.3).

11 Terms and Conditions

All members of the organizing team confirm that they have read the details regarding their duties, budgets and management process.

References

[1] S. Alemzadeh, T. Hielscher, U. Niemann, L. Cibulski, T. Ittermann, H. Völzke, M. Spiliopoulou, and B. Preim. Subpopulation discovery and validation in epidemiological data. In *Proceedings of the EuroVis Workshop on Visual Analytics*, 2017. in print.

[2] S. Alemzadeh, U. Niemann, T. Ittermann, H. Völzke, D. Schneider, M. Spiliopoulou, and B. Preim. Visual analytics of missing data in epidemiological cohort studies. In *Eurographics Workshop on Visual Computing for Biology and Medicine (VCBM)*, volume 4, pages 43–52, 2017.

[3] S. Bruckner, V. Solteszova, E. Gröller, J. Hladuvka, K. Buhler, Y. Y. Jai, and B. J. Dickson. BrainGazer–visual queries for neurobiology research. *IEEE Transactions on Visualization and Computer Graphics*, 15(6):1497–1504, 2009.

[4] M. Cypko, J. Wojdziak, M. Stoehr, B. Kirchner, B. Preim, A. Dietz, H. Lemke, and S. Oeltze-Jafra. Visual verification of cancer staging for therapy decision support. *Comp Graph Forum*, 36(3):109–120, 2017.

[5] M. A. Cypko, M. Stoehr, M. Kozniewski, M. J. Druzdzel, A. Dietz, L. Berliner, and H. U. Lemke. Validation workflow for a clinical bayesian network model in multidisciplinary decision making in head and neck oncology treatment. *Int J Comput Assist Radiol Surg*, 12(11):1959–1970, 2017.

[6] P. M. Full, S. Engelhardt, B. Burkhardt, A. Tandon, M. N. V. Forte, G. F. Greil, I. Wolf, P. Lamata, and T. Hussain. Patient-specific shape modeling to predict response to pulmonary valve replacement in patients with repaired Tetralogy of Fallot. Submitted to: 52nd Annual Meeting of the Association for European Paediatric and Congenital Cardiology (AEPC), 2018.

[7] F. Ganglberger, J. Kaczanowska, J. M. Penninger, A. Hess, K. Bühler, and W. Haubensak. Predicting functional neuroanatomical maps from fusing brain networks with genetic information. *NeuroImage*, pages 1–39, 2017. `https://doi.org/10.1016/j.neuroimage.2017.08.070`.

[8] F. Isensee, P. Jaeger, P. M. Full, I. Wolf, S. Engelhardt, and K. H. Maier-Hein. Automatic Cardiac Disease Assessment on cine-MRI via Time-Series Segmentation and Domain Specific Features. In *Automated Cardiac Diagnosis Challenge (ACDC), 7th MICCAI Workshop on Statistical Atlases and Computational Modeling of the Heart (STACOM)*, 2017. Electronic preprint: arXiv:1707.00587.

[9] P. Klemm, K. Lawonn, S. Glaßer, U. Niemann, K. Hegenscheid, H. Völzke, and B. Preim. 3D Regression heat map analysis of population study data. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):81–90, 2016.

[10] P. Klemm, S. Oeltze-Jafra, K. Lawonn, K. Hegenscheid, H. Völzke, and B. Preim. Interactive visual analysis of image-centric cohort study data. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):1673–1682, 2014.

[11] J. Müller and S. Oeltze-Jafra. Visual verification of bayesian networks. Submitted to: 29th Medical Informatics Europe (MIE) conference, 2018.

[12] S. Oeltze-Jafra, U. Niemann, J. Bernard, and A. Perer. Visual analytics of medical cohort study data. Tutorial at IEEE VIS, 2017. `https://projects.iccas.de/dpm/tutorials/IEEEVis2017/`.

[13] S. Oeltze-Jafra, A. Ynnerman, S. Bruckner, and H. Hauser. Rejuvenated medical visualization. Tutorial at IEEE VIS, 2015. `http://www.vismd.de/lib/exe/fetch.php?media=teaching_tutorials:vis15-medvis-tutorial_proposal.pdf`.

[14] A. Oeser, J. Gaebel, A. Dietz, S. Wiegand, and S. Oeltze-Jafra. Information architecture for a patient-specic dashboard in head and neck tumor boards. Submitted to: Computer Assisted Radiology and Surgery (CARS) Congress, 2018.

[15] K. Panser, L. Tirian, F. Schulze, S. Villalba, G. S. Jefferis, K. Bühler, and A. D. Straw. Automatic segmentation of Drosophila neural compartments using GAL4 expression data reveals novel visual pathways. *Current Biology*, 26(15):1943–1954, 2016.

[16] N. Pezzotti, T. Höllt, J. Van Gemert, B. P. Lelieveldt, E. Eisemann, and A. Vilanova. DeepEyes: Progressive visual analytics for designing deep neural networks. *IEEE Transactions on Visualization and Computer Graphics*, 24(1):98–108, 2018.

[17] R. G. Raidou, H. J. Kuijf, N. Sepasian, N. Pezzotti, W. H. Bouvy, M. Breeuwer, and A. Vilanova. Employing visual analytics to aid the design of white matter hyperintensity classifiers. In *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 97–105. Springer, 2016.

[18] R. G. Raidou, U. A. van der Heide, C. V. Dinh, G. Ghobadi, J. F. Kallehauge, M. Breeuwer, and A. Vilanova. Visual analytics for the exploration of tumor tissue characterization. *Comp Graph Forum*, 34(3):11–20, 2015.

[19] M. Stoehr, A. Dietz, M. Cypko, and H. Lemke. Development of the Digital Patient Model "Laryngeal Cancer" to Support the Decision-Making Process. *Int J Comput Assist Radiol Surg*, 11(Suppl 1), 2016.

[20] H. Völzke, D. Alte, C. O. Schmidt, D. Radke, R. Lorbeer, N. Friedrich, N. Aumann, K. Lau, M. Piontek, G. Born, and et al. Cohort Profile: The Study of Health in Pomerania. *International Journal of Epidemiology*, 40(2):294–307, 2011.