



FOURTH ISC WORKSHOP ON HPC APPLICATIONS IN PRECISION MEDICINE

July 2, 2021

(8 a.m. – 12 noon, All times listed are Eastern Time)

Program

- 8:00 a.m.–8:05 a.m. **Welcome — Fourth ISC Workshop on HPC Applications in Precision Medicine (HAPM21)**
Eric Stahlberg, *Frederick National Laboratory for Cancer Research*
- 8:05 a.m.–8:45 a.m. **Keynote**
Collective effects in cancer invasion and progression: Insights from mathematical models and analysis
Andreas Deutsch, *Technische Universität Dresden*
- 8:45 a.m.–9:05 a.m. *Simulating evolution in heterogeneous tumors and the influence of the microenvironment*
Jakob Rosenbauer, *Research Center Jülich*; Marco Berghoff, *Karlsruhe Institute of Technology*; Alexander Schug, *Research Center Jülich*
Abstract: In heterogeneous tumors, cell types of different properties compete over the available resources, which are nutrients and space. Rapid expansion leads to solid stress in in-vivo tumors that can collapse blood vessels, which together with angiogenesis leads to fluctuations in nutrient availability. Here, we observe the influence of such fluctuations on tumor evolution.
We developed a 3D computational model that simulates the evolutionary trajectories of an evolving tumor. Cell motility and cell-cell adhesion are observed as two free evolving parameters in tumor cells that grow in a medium of surrounding cells. A nutrient-dependent cell

cycle is introduced, and constant and dynamic nutrient surroundings are compared.

We find a clear evolutionary advantage of low adhesion cells independent of the surrounding. Furthermore, we find a dependency between the evolution speed and the frequency of the nutrient fluctuations, with a significant increase of evolutionary speed for a frequency domain.

Presenter: Jakob Rosenbauer

9:05 a.m.–9:25 a.m.

A digital twin dyad with deep Q-learning in head and neck cancer treatment

Elisa Tardini, *The University of Illinois at Chicago*; Xinhua Zhang, *The University of Illinois at Chicago*; Guadalupe Canahuate, *University of Iowa*; Andrew Wentzel, *The University of Illinois at Chicago*; Abdallah S. R. Mohamed, *The University of Texas MD Anderson Cancer Center*; Lisanne Van Dijk, *The University of Texas MD Anderson Cancer Center*; Clifton D. Fuller, *The University of Texas MD Anderson Cancer Center*; G. Elisabeta Marai, *The University of Illinois at Chicago*

Abstract: Each year, over 60,000 people in the United States are diagnosed with head and neck cancers. Treatment strategies are dynamic and complex and can involve multiple courses of therapy with different potential outcomes. With head and neck cancer numbers increasing every year, partly due to viral infections, clinicians aim to develop customized treatment strategies tailored to each individual patient, under a healthcare model termed “precision medicine.” Precision medicine proposes the customization of healthcare, with medical decisions, practices, and products being tailored to the individual patient. The customization is based on individual factors collected from cohorts of patients who are similar to the patient under consideration. These factors may include disease markers; treatment options; demographics; and genetic, lifestyle, environmental, laboratory, or quality of life data. These data are heterogeneous, large scale, and locally sparse. Making similarity-based precise recommendations for a specific patient requires novel approaches, which blend biomedicine with complex quantitative methodology.

We describe a first application of deep-Q-learning (DQL) and simulation as a digital twin dyad to optimize multi-step treatment of head and neck cancer patients and to predict patient survival and toxicity outcomes. The treatment decision DQL digital twin and the patient’s digital twin were created, trained, and evaluated using Google Colab for the HPC solution on a dataset of 536 oropharyngeal squamous cell carcinoma (OPC) patients with the goal of, respectively, determining the optimal treatment decisions with respect to survival and toxicity metrics and

predicting the outcomes of the optimal treatment on the patient. The models were trained on a subset of 402 patients (split randomly) and evaluated on a separate set of 134 patients. The main outcomes and measures include accuracy in predicting the final outcomes of a treatment sequence and improvement in predicted outcomes rates of the DQL when compared to the observed physician outcomes. On the validation set, we obtained 87.09% mean and 90.85% median accuracy in treatment outcome prediction. Given the prediction accuracy and predicted improvement on medically relevant outcomes yielded by this approach, this digital twin dyad of the patient-physician dynamic treatment problem has the potential of aiding physicians in determining the optimal course of treatment and in assessing its outcomes.

Presenter: Elisa Tardini

9:25 a.m.–9:45 a.m.

Leveraging a hybrid and multiscale model to assist androgen deprivation therapy in recurrent prostate cancer patients

Mengdi Tao, Alokendra Ghosh and Ravi Radhakrishnan, *University of Pennsylvania*

Abstract: Localized prostate cancer (PCa) is commonly treated with radical prostatectomy (RP). However, 20%–25% of patients with RP have a biochemical recurrence (BR) within five years, with an increased level of prostate-specific antigen (PSA) in serum (≥ 0.2 ng/ml). PCa is an androgen-sensitive disease, and one standard treatment option for BR PCa is androgen deprivation therapy (ADT). Initially, ADT reduces cell proliferation and increases apoptosis, resulting in the suppression of PSA. However, within two years, the tumor progression of castration-resistant prostate cancer (CRPC) occurs.

In order to better predict the ADT effect and PSA level in prostatectomized patients, we extended a published multiscale PCa model [1], by adding a two tumor cells populations model to it. The combined hybrid and multiscale model captures the genetic mechanisms related to the development of CRPC and the Luteinizing Hormone-Releasing Hormone (LHRH) agonists or Combined Androgen Blockade (CAB) treatment effect on both PSA and testosterone production in patients.

Latin hypercube sampling (LHS) was performed on the published PCa model to assess sensitivities and robustness, with respect to model output – net cell growth (NCG). Extended Fourier amplitude sensitivity (eFAST) was performed on the two tumor cells populations model to examine sensitivities of model output, namely PSA levels. In-vivo and patient data were used to develop the two cells, population models. The extended combined model was validated with the EUREKA1 database, which contains a subset of prostatectomized patients considered high-

risk populations, with Gleason scores (GS) ≥ 8 and positive surgical margins. Different clinical scenarios were simulated with the extended model to predict PSA levels after the treatment of healthy controls and recurrent patients. Case studies are designed based on different levels of PTEN deletion and epidermal growth factor (EGF).

The extended combined model was developed in both Python and Matlab. LHS and case studies simulations were performed with high-performance parallel computing to overcome expensive computational cost.

Presenter: Mengdi Tao

9:45 a.m.–10:05 a.m.

Quantifying vulnerability of privacy attacks on MT-CNN models for information extraction from cancer pathology reports

Hong-Jun Yoon, Hilda B. Klasky, Christopher Stanley and J. Blair Christian, *Oak Ridge National Laboratory*; Eric B. Durbin, *University of Kentucky*; Xiao-Cheng Wu, *Louisiana State University Health Sciences Center School of Public Health*; Antoinette Stroup, *Rutgers Cancer Institute of New Jersey*; Jennifer Doherty, *University of Utah*; Linda Coyle, *Information Management Services, Inc.*; Lynne Penberthy, *National Cancer Institute*; Georgia Tourassi, *Oak Ridge National Laboratory*

Abstract: The presence of private information in a machine learning model's training data poses two problems: first, it degrades the task performance of the model, and second it increases its vulnerability to privacy attacks. However, there is no trivial way to distinguish between common information and private information in a dataset—machine learning training algorithms employ all available information to maximize classification accuracy. We apply an algorithm for membership inference attack (MIA) to machine learning models to quantify security vulnerability, which indicates how much private information is involved in the inference of the models.

This paper describes the application of the Shokri algorithm to multi-task convolutional neural network models for automatic information extraction from cancer pathology reports, which consists of six tasks that include many class labels with a severe class imbalance. Our study reviews the classification accuracy scores of an MIA for each class label. In the results, we observe higher MIA accuracy for rare cancer labels, up to 80%, whereas the model scores 55% for high-incidence cancers. This result clearly indicates that under-represented class labels are more likely to overfit and employ private information in model coefficients.

Presenter: Hong-Jun Yoon

10:05 a.m.–10:20 a.m.

Break

10:20 a.m.–10:55 a.m. **Panel: *Digital Twins for Cancer Care***

Moderator:

Eric Stahlberg, *Frederick National Laboratory for Cancer Research*

Panelists:

Anastasia Christianson, *Johnson & Johnson*

Tina Hernandez-Boussard, *Stanford University*

Emily Greenspan, *National Cancer Institute*

Marieke Kuijjer, *Centre for Molecular Medicine Norway*

10:55 a.m.–11:15 a.m. ***Medical cybernetics for continuous risk assessment and value of information analysis of treatments***

Zsolt Ori, *Ori Diagnostic Instruments LLC*

Abstract: Individualized precision methods are needed for continuous non-invasive monitoring of state variables (SVs) of risk indicators not just for early detection of subclinical derangements but also to monitor progress of effect of lifestyle modification and medical therapies throughout lifespan.

For purposes of “precision medicine” we propose here a cloud-based Cyber-Physical System (CPS), a mobile technology to integrate sensory data from various mobile devices of a user into individualized dynamic mathematical models of physiological processes to draw life-course trajectories in four domains with major implications for morbidity/mortality: 1. Cardiometabolic, 2. Cardiorespiratory, 3. Cardio vegetative, and 4. Cardiovascular functioning. The following state variables (SVs) are assessed and predicted daily: Ad1. insulin resistance measuring R_w ratio, lean mass, fat mass, fat vs carbohydrate oxidation ratio, respiratory quotient, de novo lipogenesis, and adaptive thermogenesis; Ad 2. maximum oxygen uptake, exercise capacity, heart rate reserve; Ad3. heart rate variability; and Ad 4. blood pressure, total arterial compliance. The sensor data are picked up from wearable sensors of the fitness industry such as Garmin and from our patented bioimpedance Body-Composition Hydration-Analyzer Photo-plethysmography equipped stand-up scale. All measurement methods can be calibrated to widely available standard office-based measurements.

We were able to prove the feasibility of our modeling concept of the insulin resistance by finding strong correlation $\rho = -0.6745$, $P = 0.000024$ between changes of R_w -ratio and insulin resistance HOMA-IR in 12 clinical studies with 39 clinical study arms. Based on these results we found that CPS is a suitable concept to indirectly measure and predict the otherwise very difficult- or impossible-to-measure slow change of SVs of the metabolism capture them for the first time noninvasively in the user’s natural environment. Serial fat and weight measurements and

energy calculations can help unmask changes of insulin resistance in response to user's diet and exercise habits.

We envision the possibility that the SVs can be plugged into already published risk calculation formulas derived from analysis of cumulative incidence in competing risks and analysis of competing regression. The derived metrics allow for assessing individually calculated risk of "end-point," quality-adjusted life-year (QALY), incremental cost-effectiveness ratio (ICER), and information analysis (VOI) of therapies. Having the effectiveness and cost-saving data could be essential for payers and policy makers to consider incentives or other major structural changes to health care delivery. The continuously provided value metrics especially if showing positive results could incentivize users and facilitate the creation of a sustainable business cycle for social entrepreneurship framework by informing all financial stakeholders in a community program.

The feedback of individualized metrics using tools of the digital health era may amount to channeling focus also to patient-centered individualized care and to accelerating nutrition research.

In conclusion, a CPS with machine learning using principles of optimal control theory supervised by physician can provide a truly individualized strategy for estimation, continuous monitoring, and prediction of physiological state variables for self-therapy, guided therapies, and mobile health interventions or cyber-therapy. CPS facilitated interventions allow for improving health, fitness, resilience, and the chance of survival of an acute illness.

11:15 a.m.–11:35 a.m.

Security standards and ease of use of HPC systems by clinical researchers

Matthew R. Link, Anurag Shankar, David Hancock, Robert Henschel, Scott Michael and Craig A. Stewart, *Indiana University*

Abstract: Precision health research and personalized health therapies involve analysis of protected health information. In 2007, Indiana University established the ability to analyze protected health information (HIPAA alignment) as the minimal and default security level for its research high performance computing (HPC) systems and research storage systems. This resulted in a dramatic increase in the use of IU HPC systems by clinical researchers. Security levels were later upgraded to FISMA Low as a default. We recommend that, within the US, FISMA (Federal Information Security Modernization Act) Low compliance be the default minimal level of security for large-scale HPC systems. This would facilitate precision medicine research and enable higher education HPC resources to be used in response to future civil health emergencies.

Presenter: Craig A. Stewart and Anurag Shankar

11:35 a.m.–11:55 a.m.

Longitudinal deep learning study on MIMIC-III dataset

Xin Dai, Ji Hwan Park, Nicholas D'imperio, Shinjae Yoo, *Brookhaven National Laboratory*

Abstract:

1. Background

In clinical research, the goal of the longitudinal study is to exploit the multi-modal electronic health records (EHR) and reveal the temporal dynamics of patients' health status.

Recurrent-Neural Network (RNN) and Transformer are two exceptional deep learning models in handling sequence data. Hence, they and their variants constitute good candidates for the clinical longitudinal study.

Our goal is to have a comprehensive evaluation to identify the pros and cons of the existing deep learning longitudinal models to guide future model development.

We use the freely accessible MIMIC-III dataset, one of the most used clinical datasets for model development and benchmark. The data includes information such as demographics, vital sign measurements, and laboratory test results.

We focus on the in-hospital mortality task, which uses the first 48 hours of 17 physiological data during an ICU stay to predict if a patient will die afterward.

2. Method

After conducting an extensive literature search, we selected four models, GRU-D, StageNet, Transformer, and Set-Att, and implemented them on the MIMIC-III dataset. GRU-D and StageNet are RNN-based models, while Set-Att is a simplified variant of the Transformer.

We followed the data processing pipeline outlined in the Ref., resulting in a total of 21,139 ICU stay records, with a 13% mortality rate. The sizes of train, validation and test data are 14,681, 3222, 3236, respectively.

3. Result

Due to the data imbalance, RO-AUC is a better evaluation metric than accuracy. To ensure a fair comparison, we performed hyperparameter tuning to find the best possible parameters for each model.

The RO-AUC results shown below are from the test data.

Model	GRU-D	StageNet	Transformer	Set-Att
-----	-----	-----	-----	-----
RO-AUC	0.857	0.844	0.821	0.839

4. Conclusion

While Transformer-base models have achieved superior performance in many applications, for the MIMIC-III dataset, RNN-based models perform better. To our surprise, the simplest model among the four, GRU-D, gets the highest score.

Compared with the regular time series sequence, the two most prominent challenges present in the clinical dataset are the irregular time interval between measurements and the high missing data rate. GRU-D introduces the learnable decay parameters to both the input data and hidden states to mitigate time irregularity and improve data imputation quality.

We conjecture the reasons for the inferior performance of Transformer-based models are two-fold. The temporal information is critical for the longitudinal study. While self-attention, the core idea of Transformer, can capture the non-local interactions between different parts of the sequence. It is a permutation-invariant operation. Hence, the Transformer is less sensitive to the time order.

Another possible explanation is the lack of pre-training. All successful applications of Transformer in Natural Language Processing (NLP) require large-scale unsupervised pre-training. The moderate size of the MIMIC-III dataset renders pre-training unfeasible.

Presenter: Xin Dai

11:55 a.m.–12:00 p.m. **HAPM21 Wrap-Up**

Keynote



Andreas Deutsch, Dresden University of Technology

Andreas Deutsch is head of the department of Innovative Methods of Computing at the Centre for Information Services and High Performance Computing (Dresden University of Technology). His research is focused on mathematical biology, especially cellular automata and agent-based modelling, cancer invasion and collective phenomena in the life sciences.

Organizing Committee

Marco Berghoff – Karlsruhe Institute of Technology

Stephen Litster – Amazon Web Services

Jan Nygard – Cancer Registry of Norway

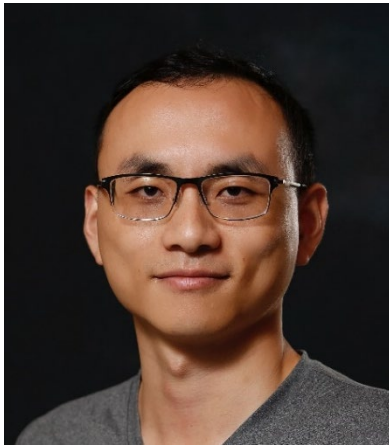
Eric Stahlberg – Frederick National Laboratory for Cancer Research

Thomas Steinke – Zuse Institute Berlin

Lynn Borkon – Frederick National Laboratory for Cancer Research

Petrina Hollingsworth – Frederick National Laboratory for Cancer Research

Presenter Bios



Xin Dai, Brookhaven National Laboratory

Xin Dai is a Research Associate in the Computational Science Initiative, Brookhaven National Laboratory. His research focuses on applying machine learning methods to various domain scientific problems. Before joining BNL, he was a postdoc researcher at Ohio State University. He obtained a PhD in physics from Tsinghua University in 2018.

Emily Greenspan, National Cancer Institute

Dr. Emily Greenspan is a biomedical informatics program director in the Center for Biomedical Informatics and Information Technology (CBIIT) at the National Cancer Institute (NCI). She serves as the NCI federal program lead for the NCI-Department of Energy (DOE) Collaborations focused on applying advanced computing and artificial intelligence (AI) technologies to specific areas of cancer research. She has supported and promoted predictive oncology and AI strategies across NCI.





Tina Hernandez-Boussard, Stanford University

Dr. Hernandez-Boussard is an associate professor in medicine (biomedical informatics), biomedical data science and surgery at the Stanford University School of Medicine. Dr. Hernandez-Boussard's background and expertise is in the field of computational biology and epidemiology, with concentration on clinical informatics, population health and health policy. A key focus of her research is the application of novel methods and tools to large clinical datasets for hypothesis generation, comparative effectiveness research and the evaluation of quality healthcare delivery. (boussard@stanford.edu)



Marieke Kuijjer, Centre for Molecular Medicine Norway

Marieke Kuijjer is a Group Leader in Computational Biology and Systems Medicine at the Centre for Molecular Medicine Norway (NCMM), a Nordic EMBL partner, University of Oslo, Norway. She also has a 20% affiliation as Assistant Professor at the department of Pathology, Leiden University Medical Center, the Netherlands. Before starting her research group at the NCMM, she completed a postdoctoral fellowship in computational biology in the laboratory of John Quackenbush at the Department of

Biostatistics and Computational Biology of the Dana-Farber Cancer Institute and the Department of Biostatistics of Harvard T.H. Chan School of Public Health, Boston, MA. She obtained her PhD in 2013 from Leiden University, working at the Department of Pathology of Leiden University Medical Center in the field of cancer genomics. Dr. Kuijjer's research interests include computational tool development and their applications in personalized cancer network genomics. She serves on the editorial board of Cancer Research and is associate editor of Bioinformatics Advances.



Zsolt P. Ori, Ori Diagnostic Instruments LLC

Zsolt P. Ori is a practicing internist, primary physician and hospitalist with previous training in Bio-Medical Cybernetics (Ilmenau University of Technology, Germany) before entering medical school (Albert Szent-Györgyi Medical School, Szeged, Hungary) and post-doctoral research fellowship in non-invasive cardiology (Northwestern University, Chicago). His engineering training together with his experiences in primary care in academic and non-academic environments have inspired his vision for using Cybernetics to improve cardiometabolic health with result driven predictive feedback control. His patented inventions can build a bridge between sensory data from wearables

and plug the data into physiological process models and make the otherwise undetectable slow changes of the energy metabolism observable and derive easily readable trend indicators allowing for stepwise dynamic behavior control for reaching cardiometabolic health and mental resilience. CPS derived data can provide metrics facilitating education about metabolic health as well as to reach community health, corporate health and public health goals. Principles of continuous non-invasive risk assessment and management of health in the user's home environment could be followed realizing individualized "precision medicine."

Jakob Rosenbauer, Forschungszentrum Jülich

Jakob Rosenbauer works in the group of Alexander Schug at Forschungszentrum Jülich and graduated in May. He obtained his MSc at Karlsruhe Institute of Technology. He currently works on the implementation and application of large-scale tissue modeling using supercomputing architectures. His main research interests are the simulation of emergent behavior from single cell mechanical properties and the translation to findings in tumor development.



Anurag Shankar, Indiana University

Anurag Shankar manages a small group of cybersecurity professionals within Indiana University's Center for Applied Cybersecurity Research (CACR), a Pervasive Technology Institute unit. He oversees HIPAA compliance for IU's Office of the Vice President for IT and runs a campus-wide service that assists IU researchers with cybersecurity and compliance. Shankar is a computational astrophysicist by training. He was part of the inaugural batch of users at the newly established US supercomputer centers in the 80s such as NCSA, SDSC and

PSC. After his research career, he transitioned to IT and research computing, spending nearly two decades developing, delivering, and managing IU's central research cyberinfrastructure,

including services such as Unix support, distributed storage, computational grids, and HIPAA aligned solutions for clinical researchers. He switched to research cybersecurity and regulatory compliance in the late 2000s and developed IU's present, NIST-based Risk Management Framework and leveraging it to align central research systems with HIPAA. He routinely assists other institutions tackling HIPAA and other types of compliance and conducts compliance assessments.



Eric Stahlberg, Frederick National Laboratory for Cancer Research

Dr. Eric Stahlberg is the director of Biomedical Informatics and Data Science (BIDS) at the Frederick National Laboratory for Cancer Research. Dr. Stahlberg is a founding co-organizer of the Computational Approaches for Cancer Research Workshop (CAFCW), held in conjunction with the SC conference since 2015. He has been instrumental in establishing the Frederick National Laboratory's high-performance computing initiative and in assembling scientific

teams across multiple, complex organizations to advance predictive oncology. In 2017, he was recognized as one of FCW's Federal 100.

Craig A. Stewart, Indiana University

Craig A. Stewart has been a leading innovator and strategist in advanced computing and cyberinfrastructure for research and development for decades. He was the executive director of the Indiana University Pervasive Technology Institute from 2008 to 2020. IU PTI is Indiana University's flagship organization for research and development in cyberinfrastructure, informatics and computer science in support of science and engineering research, artistic creativity and cybersecurity. Stewart's group coined the now standard definition of

cyberinfrastructure – see hdl.handle.net/2022/21589. Stewart has been involved in research computing since writing Fortran-based statistical applications while completing his PhD in biology in the 1980s. Stewart was the founding PI of Jetstream, the first cloud computing system funded by the National Science Foundation for use by the science and engineering community of the US. In addition to his work at IU, Stewart has been a visiting faculty member in computer science at the Universität Stuttgart and a Fulbright Senior Scholar at the Technische Universität Dresden. Stewart has been on staff and on several committees for the National Science Foundation as well as on advisory committees and in leadership of a number of other organizations related to advanced computing. Stewart is a past chair of the Coalition for Advanced Scientific Computation and in this role testified before the House Science and Technology Committee.





Mengdi Tao, University of Pennsylvania

Mengdi Tao is a PhD student of Dr. Ravi Radhakrishnan's lab, in the department of bioengineering at the University of Pennsylvania. She received her BS and MS degrees in biomedical engineering from Drexel University. A major focus of her ongoing research is to understand quantitative mechanistic characterization of structural biology and systems biology. The ultimate aim is to build hybrid and multiscale quantitative models of signaling networks, while retaining sufficient molecular specificity, for predicting the interactions of therapeutic agents with biochemical signaling mechanisms.



Elisa Tardini, University of Illinois at Chicago

Elisa Tardini is a Research Specialist at the University of Illinois at Chicago, where she obtained her master's of science degree in computer science. She also obtained a bachelor's and master's of science degrees in computer science and engineering from Politecnico di Milano. Her research interests are in machine learning, reinforcement learning, and medical informatics.

Hong-Jun Yoon, Oak Ridge National Laboratory

Hong-Jun Yoon is a research scientist in the Computational Sciences and Engineering Division at Oak Ridge National Laboratory. He is involved in the Joint Design of Advanced Computing Solutions for Cancer (JDACS4C), a research collaboration between the Department of Energy (DOE) and National Cancer Institute (NCI).

