AI4b.io Symposium 9 & 10 April 2024

You are invited to participate in the Artificial Intelligence Lab for Bioscience (AI4b.io) symposium. This symposium will take place on April 9 & 10. We are organizing this meeting for 100 participants who are active in Artificial Intelligence and Bioscience. Your participation is appreciated because of your expertise in this area, and we are looking forward to your contributions in vibrant discussions and see this symposium as a start of a new community on AI for bioscience. The program is included in this document and covers topics ranging from large-scale manufacturing to microbiome-based precision nutrition, going from large to small scale. Experts active in these topics will present their in-depth insights.

Details

Type of meeting: physical
Location: Vakwerkhuis, Professor Snijderstraat 2, 2628 RA, Delft
Date: April 9 (10:00 registration, program starts at 10:30, program ends at 17:00 with dinner afterwards) - April 10 (program starts at 9:15, program ends at 17:00

Latest program version

https://www.ai4b.io/assets/documents/AI4bio2024_Program_Titles_Abstracts.pdf

Organizers

Kim van den Houten
Joery de Vries
Paul van Lent

AI4b.io Steering Committee

Marcel Reinders
Henk Noorman
Hans Roubos
Jana Weber
Renger Jellema

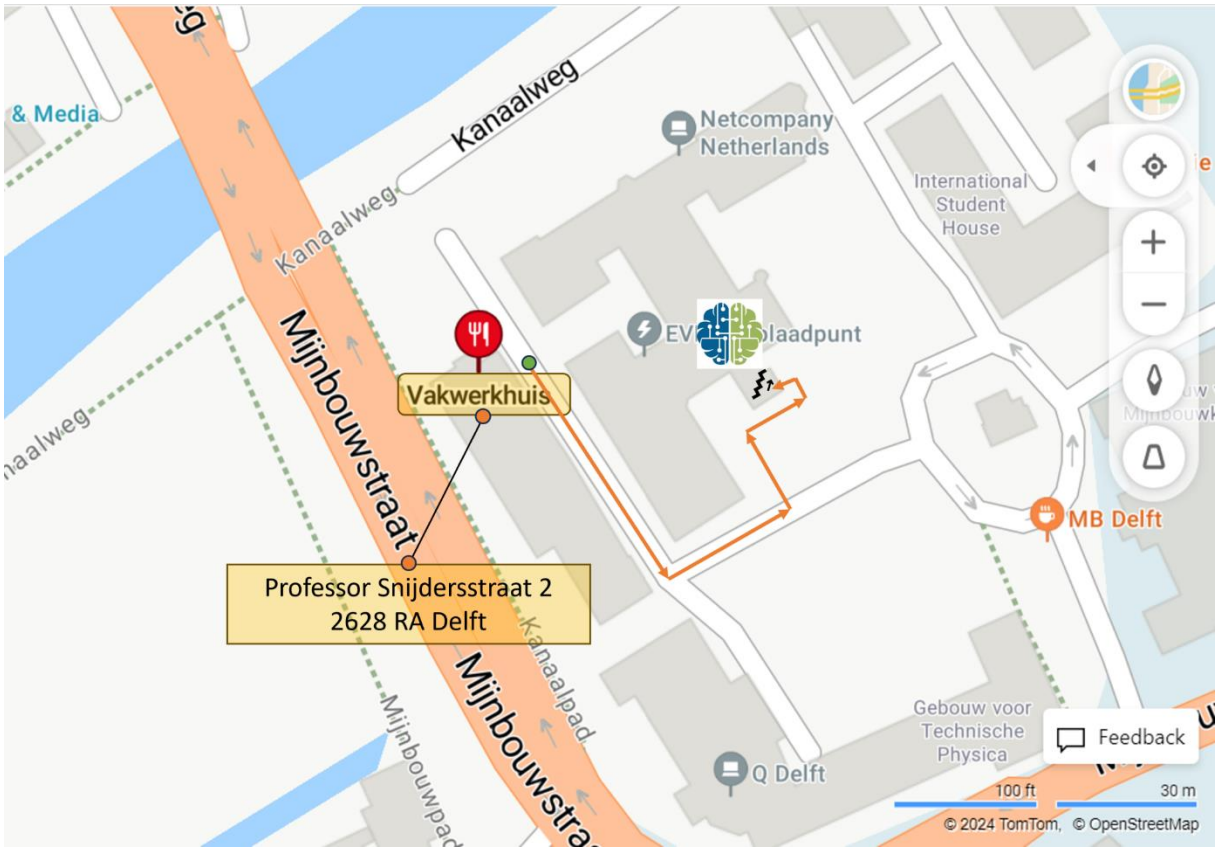
If you have questions about lodging and travel, please contact info@ai4b.io, mentioning AI4b.io symposium

Location & Parking Facilities

Vakwerkhuis is located at Professor Snijderstraat 2 in Delft.

The venue can be reached both by car and public transport. For those driving to the event, parking is available at the nearby Zuidpoort garage. If you travel by public transport, go to Delf Central station and either walk or cycle the 950 meters.

Please use the map below to find the exact location and the Vakwerkhuis through the link: https://maps.app.goo.gl/GBzkaxP7eSYFQD4ZA



Hotels

If you're considering staying overnight, there are several nice hotels nearby, including the BW Signature Collection Grand Museum Hotel, Ibis, The Social Hub, Hampshire Hotels, and Hotel de Koophandel.

Timetable – Day 1

Presentation	Speaker	Affiliation	Day	Time
Arrive, register, coffee			April 9th	10:00-10:30
		AI4b.io		10:30-10:45
1	Sara Moreno Paz	WUR		10:45-11:10
2	Thomas Abeel	TU Delft		11:10-11:35
3	Chengyao Peng	TU Delft		11:35-12:00
Lunch				12:00-13:00
Invited 1	Sanne Abeln	Utrecht University		13:00-14:00
4	Michiel Busschaert	KU Leuven		14:00-14:25
5	Jeroen van de Laar, Chris McCready	aBioPQ, Sartorius Data Science		14:25-14:50
Break				14:50-15:10
Invited 2	Jana Weber	TU Delft		15:10-16:10
6	Dominik Goldstein	TU Delft		16:10-16:35
7	Tanuj Karia	TU Delft		16:35-17:00
Dinner				17:00-20:00

Timetable – Day 2

Presentation	Speaker	Affiliation	Day	Time
Arrive, register, coffee			April 10th	8:45-9:15
8	Konstantina Tzavella	VU Brussel		9:15-9:40
9	Giaulia Crocioni	Netherlands eScience Center		9:40-10:05
10	Roel Leenhouts	KU Leuven		10:05-10:30
Break				10:30-11:00
Invited 3	Jan-Willem van de Meent	UvA		11:00-12:00
Lunch				12:00-13:00
11	Giocomo Lastrucci	TU Delft		13:00-13:25
12	Maximilian Theisen	TU Delft		13:25-13:50
13	Jie Yang	TU Delft		13:50-14:15
Poster Session + Coffee				14:15-15:45
Invited 4	Tias Guns	KU Leuven		15:45-16:45
Closing remarks				16:45-17:00
Closing drinks				17:00-18:00

Day 1, April 9th, 10:00 – 17:00

Session 1 Chair: Esteban Freydell | dsm-firmenich

10:00 – 10:30 | Arrive, register, coffee

10:30 – 10:45 | Welcome Note

1 | 10:45 – 11:10 | Machine learning-guided optimization of p-coumaric acid production in yeast

Sara Moreno Paz
Wageningen University and Research

Industrial biotechnology uses Design-Build-Test-Learn (DBTL) cycles to accelerate the development of microbial cell factories, required for the transition to a bio-based economy. To use them effectively, appropriate connections between each phase of the cycle are crucial. Using p-coumaric acid production in Saccharomyces cerevisiae as a case study, we propose the use of one-pot library generation, random screening, targeted sequencing, and machine learning (ML) as links during DBTL cycles. We showed that the robustness and flexibility of ML models strongly enable pathway optimization, and propose feature importance and SHAP values as a guide to expand the design space of original libraries. This approach allowed a 68% increased production of p-coumaric acid within two DBTL cycles leading to a 0.52 g/l titer and a 0.03 g/g yield on glucose.

2 | 11:10 – 11:35 | Computational microbial genomics

Thomas Abeel
TU Delft

Imbalances in the gut microbiome have been linked to various health problems, such as inflammatory bowel disease and certain types of cancer. While metagenomics is often used to study microbial communities, it does not fully capture their functionality. Other meta-omics modalities, like meta-transcriptomics, meta-proteomics, and metabolomics, could provide insights, but they face several challenges, including high costs and reliability issues.

The increasing availability of public multi-layered microbiome data presents an opportunity to develop machine-learning models capable of inferring connections between metagenomic data and other forms of meta-omics data, with the eventual goal of predicting the latter from the former.

We show proof-of-concept machine learning models that can accurately predict transcript, protein, and metabolite abundances in microbial community samples. We show that these predictions can be used in disease detection, providing insights beyond just metagenomics.

3 | 11:35 – 12:00 | Phytogenic feed additives and antibiotic growth promoters show differing poultry microbiome and transcriptome profiles

Chengyao Peng
TU Delft

In this study, we investigated the effects of a common antibiotic growth promoter (AGP) and a phytogenic feed additive (PFA) on the chicken cecum microbiome through a multi-omics approach.

We collected microbial metagenomic, metabolomic, and host transcriptomic samples from the cecum of chickens fed with control diet, AGP diet, or PFA diet at different time points. To pinpoint the potentially affected cecum microbial species/genes and host RNAs, we identified differentially abundant microbial species, gene clusters, and host transcripts associated with AGP and PFA treatments. Meanwhile, through a time-course analysis, we found treatment-specific changes across time in digesta microbial species, gene clusters and host transcripts. Our findings also revealed correlations between host mucosa transcriptome and digesta microbiota compositions. These results provide valuable insights into the mechanisms by which AGPs and PFAs influence the chicken gut ecosystem, paving the way for future development of improved poultry production strategies.

Break

Session 2 Chair: Ali May | dsm-firmenich

Invited 1 | 13:00 – 14:00 | Learning to understand life

Sanne Abeln
University of Utrecht

Recent advances in machine learning have opened new paths for making predictions based on large sets of complex life science data. However, is it also possible to gain an understanding of the complex underlying systems? And, is this possible when there is only a small number of labelled samples available?
Firstly, we demonstrate how to analyse the impact of a genomic alteration using system level response data. We employed machine learning methodology to explore the problem and subsequently developed a statistical approach to analyse the impact of genomic alterations using as few as 20 samples. Applying this approach, we identified several novel structural variants that are likely to have a significant impact on the development of colorectal cancer.
In a second example, we demonstrate that it is possible to predict which proteins are likely to be found in extracellular vesicles (EVs). By using meaningful input features, we can gain an understanding of both the vesicle-sorting mechanism, as well as understanding why specific proteins are likely to be found in vesicles. In addition, we can shed light on which experimental EV extraction protocols are most reliable. Finally, I will illustrate how we can tackle questions in the life sciences - that have a limited number of examples to train the model - by learning over related tasks.

4 | 14:00 – 14:25 | Application of heterogeneous FBA to microalgae

Michiel Busschaert
KU Leuven

In a biological culture, not all cells are created equal, and variability between individual cells influence the overall behavior of the cell culture. For example, in the microalgae species Chlamydomonas reinhardtii, cell size plays a role in cell composition, uptake rates of metabolites and light conversion efficiency in case of photosynthetic growth. This presentation discusses a novel scheme based on an extension of Flux Balance Analysis (FBA), which, under quasi steady-state growth, simulates intercellular metabolic rates as function of cell size, as well as the distribution of different cell sizes in the overall cell population. The photosynthetic growth of C. reinhardtii is used as case study, for which experimental data on cell size distribution has been collected.

5 | 14:25 – 14:50 | Hybrid deep learning model for in-silico optimization of a dynamic perfusion cell culture process

Jeroen van de Laar, Chris McCready
aBioPQ, Sartorius Data Science

Determining optimal process conditions for CHO cultures for biopharmaceuticals remains a significant challenge. Limited resources and time constraints often make it difficult to fully explore the optimal range of all process parameters, such as temperature, pH, dissolved oxygen, basal and feed culture media, and additives. Model-based optimization offers a potential solution by systematically screening selected conditions for the cell culture process to find the optimum. The hybrid model presented in this work, trained on several datasets of small scale and pilot scale experiments can be used to explain cell culture growth dynamics. Modeling of viable cell density (VCD), cell viability, productivity (using

the NN approach) allowed the optimization of the media exchange rate, in order to optimize the overall product output of the perfusion process.

Break

Session 3 Chair: Henk Noorman | dsm-firmenich

Invited 2 | 15:10 – 16:10 | Molecular machine learning for property prediction and molecular design

Jana Weber
TU Delft

The discovery and development of novel functional molecules impacts the sustainability transition through a multitude of aspects, e.g. by discovering improved (bio)catalysts, novel energy materials, biodegradable products, or more sustainable reaction pathways. Communities from computational chemistry, material design, and bioinformatics are jointly supporting such endeavours through the development of machine learning algorithms on molecular data: molecular machine learning. In this talk, we will see recent computational concepts such as self-supervised learning, language models, and generative AI applied on molecular data. I will highlight selected building blocks for molecular design, such as finding a suitable molecular representation, developing molecular property predictors, and guidance strategies for inverse molecular design.

6 | 16:10 – 16:35 | Automatic prediction of process control structure using generative AI

Dominik Goldstein, Lukas Schulze Balhorn, Artur M. Schweidtmann
TU Delft

Piping and instrumentation diagrams (P&IDs) are important depictions of chemical and biochemical processes and are derived from process flow diagrams (PFDs) in tedious, manual workflows. Recently, we have proposed a methodology for the prediction of control structures in PFDs, which is inspired by transformer-based natural language translation models. To extend our existing approach, we present four avenues for future research. Firstly, more data beyond the process topology could be included, while other suitable model architectures could be investigated. Further, the automation more steps in P&ID creation could be researched. Finally, we hope that more training data will enhance the model's performance.

7 | 16:35 – 17:00 | Teaching artificial intelligence to (bio)-chemical engineers

Tanuj Karia, Maximillian F. Theissen, Qinghe Gao, Hector Maldonado de Leon, Ramon van Valderen Jana M. Weber, Nancy de Groot, Jolanda Quak, Cees Haringa, and Artur M. Schweidtmann.
TU Delft

Recent advancements in machine learning (ML) and artificial intelligence (AI) are transforming the disciplines of engineering and science. In chemical and biochemical engineering, novel paradigms for engineering chemical and biochemical processes have been developed, necessitating the need to train students with the latest advancements in AI and ML. While a plethora of resources are available to learn AI and ML techniques, few are tailored to (bio-)chemical Engineering. This paper outlines how a new elective course on the latest AI/ML techniques applied to (bio)-chemical engineering has been developed and incorporated into the curriculum of the Master's level degree programme at the Department of Chemical Engineering and Department of Biotechnology at TU Delft.

Day 2, April 10th, 8:45 – 17:00

Session 4 Chair: David Tax | TU Delft

8:45 – 9:15 | Arrive, register, coffee

8 | 9:15 – 9:40 | D2D: a foundation model applied to predict the impact of mutations on protein function

Konstantina Tzavella
Vrije Universiteit Brussel

The remarkably quick evolution of highly flexible, reusable artificial intelligence (AI) models is transforming healthcare and medicine research. Our D2D model is, through the combination of a self-supervised large language model with protein-specific evolutionary data, capable of carrying out a diverse set of tasks using very little to no task-specific labelled data. We illustrate its potential for high-impact applications on the prediction of single and multiple mutation pathogenicity and protein stability. The D2D model is, for example, able to infer the impact of multiple mutations on a protein's fitness without being specifically trained on the task. The D2D model also significantly outperforms the state-of-the art predictors when it is fine-tuned merely on scarce labelled data, such as in the context of passenger and driver mutations in cancer. We expect that D2D-enabled applications can shift current strategies from solitary AI models with exclusive problem-settings to synergistic implementations leveraging non-annotated information.

9 | 9:40 – 10:05 | DeepRank2: Exploring 3D protein structures through geometric deep learning

Giulia Crocioni
Netherlands eScience Center

Machine learning, and in particular geometric deep learning (GDL), can be leveraged to extract valuable biological insights from molecular data for applications in protein engineering and cancer vaccine design. Despite notable advancements in computational structural biology, the translation of these breakthroughs into practical applications remains challenging. To facilitate this, we have developed DeepRank2, a Python package that removes the daunting steps of data processing and physicochemical feature calculations on millions of molecular structures and allows the user to easily train GDL algorithms to predict biologically relevant patterns. The package has already been used to tackle challenges in cancer immunotherapy, in particular to predict peptide binding to rare major histocompatibility complex proteins, which play a crucial role in effective cancer immunotherapy through T-cell receptor recognition. DeepRank2's user-friendly interface, comprehensive documentation, and pre-implemented features position it as a versatile tool empowering researchers to explore critical questions in structural medical biology independently

10 | 10:05 – 10:30 | Graph neural networks for molecular mixture properties

Roel J. Leenhouts
KU Leuven

Physicochemical properties are crucial in the fast design of new bioprocesses. In our recent work, we have investigated the extension of molecular property prediction from pure compounds to mixture properties. For this purpose of molecular mixture property prediction, a new neural network

architecture is created. We create databases based on quantum chemistry (QC) simulations, mixing rules, and experimental data. By applying transfer learning, we can leverage both the diversity of the simulations and the accuracy of the experimental data. The novel proposed architecture makes fast and accurate mixture property prediction possible. Enabling transfer learning to mixtures can be an important help for problems where data availability is an issue. Mixture properties such as solubility or viscosity have a major influence on the phase behaviour and transport properties of the process. This information is crucial for optimizing conditions, improving efficiency, and ensuring the success of bioprocesses in various applications, from pharmaceutical production to biofuel manufacturing.

Break

Session 5 Chair: Mathijs de Weerdt | TU Delft

Invited 3 | 11:00 – 12:00 | What can AI do for science, and what can science do for AI?

Jan-Willem van de Meent
University of Amsterdam

Recent advances in AI have shown that near-miraculous examples of generalization can emerge when large models are pre-trained on massive datasets. Beyond the horizon of these highly visible breakthroughs, which have predominantly focused on images and text, there lies a tremendous potential for AI in domains where the data modalities and prediction problems are much more diverse. In this talk, I will discuss where I see opportunities to apply AI to the modeling and design of physical systems in science and engineering. One opportunity is the application of probabilistic programming to infer structure and model parameters from data. A second is using AI methods to train fast phenomenological methods to approximate slower low-level simulations. I will also provide perspectives on emerging problems in science that inform methods development in AI, such as the methods for systems described by differential equations, and methods for inverse design.

Lunch

Session 6 Chair: Willi Gottstein | dsm-firmenich

11 | 13:00 – 13:25 | Physics-informed neural networks for modeling of methanol reactors

Giacomo Lastrucci[1], Abhinav A. Verma [2], Stanislav Jaso [2], Ankit A. Tyagi [2], and Artur M. Schweidtmann [1]

[1] TU Delft

[2] Shell PLC

Multiscale modeling of catalytic reactors often leads to complex PDE or ODE systems that, despite advancements, pose computational challenges in large-scale process simulations and optimizations. We develop a PINN-based model for packet-bed reactors, integrating physical insights and data, and enhance it to predict across varied initial conditions with an adaptive loss balancing mechanism. We systematically compare the developed model with standard ODE solvers and a vanilla MLP in terms of accuracy and runtime for potential application in complex plant-wide simulation and optimization. While adjusting tolerances reduces ODE solver runtime, the MLP and PINNs dramatically outpace it, being 35 times quicker and maintaining near-perfect accuracy (99.5%). This outcome indicates a potential for utilizing MLP and PINNs models within larger optimization or simulation studies.

12 | 13:25 – 13:50 | Topology-aware soft sensor modeling leveraging graph neural networks

Maximilian F. Theisen, Lynn Luderer, Gabrie M.H. Meesters, and Artur M. Schweidtmann
TU Delft

Soft sensors play an important role in modern (bio-) chemical process operations. Approaches to soft sensor modelling are divided into two categories, data-driven and first-principle based. As data becomes abundant in modern plants, data-driven approaches also become increasingly attractive. However, they are black boxes and neglect the underlying physics. To overcome this gap, we propose a hybrid framework to soft sensor modeling using both the process topology and data. We use graph neural networks to model unit operations as nodes and streams as edges. To demonstrate the effectiveness of our novel method, we applied it to one industry-relevant case study. Overall, we contribute towards closing the gap between data-driven and first-principle driven soft sensor.

13 | 13:50 – 14:15 | Generative AI with human in the loop

Jie Yang
TU Delft

Generative AI is used more and more in science, engineering, and operational tasks. Along with the rapid increase of their adoption come increased concerns about the inherent robustness issues of such technologies (e.g., hallucination) and the associated social, and ethical implications. To create Generative AI systems that can properly serve humans, it is crucial to put humans at the center of the process such that the outcome system behaves in a way that fits the cognition and value of people in the contexts of use. This poses new technological challenges: how to build Generative AI systems that can be understood by humans and that can align their behavior with human values? Tackling these challenges requires new ways of looking at the computational roles humans can and should play in both developing and using Generative AI. In this talk, I will present our recent work on human-centered computing for understanding and improving the robustness of Generative AI systems by connecting AI with human reasons and values.

14:15 – 15:45 | Poster Pitches, Poster Session & Coffee

Poster abstracts are included at the end of this document.

Session 7 Chair: Mathijs de Weerdt

Invited 4 | 15:45 – 16:45 | Prediction + Optimization problems

Tias Guns
KU Leuven

Combinatorial optimisation is used in industry, research and society to solve scheduling, assignment, allocation problems and more. Combinatorial optimisation problems can efficiently solve hard computational problems that involve finding a satisfying or optimal assignment to a set of decision variables, subject to a range of structural constraints. Increasingly, the constraints and objective that make up the problem are no longer fully determined by an expert. Instead, part of the values (demand, duration, heat production, energy use, ...) must be estimated and predicted from data. This means that there are now two components, an 'upstream' machine learning/forecasting task and a 'downstream' combinatorial optimisation task. A classic approach would be to do the two independently, but integrated approaches can do better. Interestingly we will show that for different types of data and predictions, different types of integrations are beneficial. We will review the motivations, assumptions and solution approaches to decision-focused learning (input from historic data) and preference-based solving (input from historic solutions). We close the talk by sketching how this fits in our larger vision of conversational human-aware technology for optimisation.

14 | 16:45 – 17:00 | Closing remarks

Poster Session | 14:15-15:45 | Abstracts

Learning from scenarios for repairable stochastic scheduling

Kim van den Houten
TU Delft

When optimizing problems with uncertain parameter values in a linear objective, decision-focused learning enables end-to-end learning of these values. We are interested in a stochastic scheduling problem, in which processing times are uncertain, which brings uncertain values in the constraints, and thus repair of an initial schedule may be needed. Historical realizations of the stochastic processing times are available. We show how existing decision-focused learning techniques based on stochastic smoothing can be adapted to this scheduling problem. We include an extensive experimental evaluation to investigate in which situations decision-focused learning outperforms the state of the art, i.e., scenario-based stochastic optimization.

Combinatorial pathway optimization of p-coumaric acid producing yeast strains

Paul van Lent
TU Delft

DNA libraries are often used to perform combinatorial pathway optimization (CPO) of strains, to increase the production of a compound of interest. In this work, we have performed a transformation experiment for improving production of p-coumaric acid, a compound that is a precursor to many downstream products with bioactive properties. We used a mechanistic model yeast8 to choose genes for engineering, performed a large screening of up to 3000 strains and sequenced a representative set of 600 strains. Screening results show that the best producer has increased production by 330% over the parent strain. Preliminary results of sequencing data suggest that several genes that were perturbed have an impact on production, with most of these genes being part of the aromatic amino acid biosynthesis pathway.

Bayesian meta-reinforcement learning with Laplace variational recurrent networks

Joery de Vries, Mathijs de Weerdt, Matthijs Spaan
TU Delft

In model-based optimization we require accurate uncertainty quantification to find actions (experiments) that optimally trade-off the returns with model correction. Learning world-models is often done with point-estimate based methods due to simplicity and stability. We show how the Laplace approximation can freely extend point-estimate methods to a Bayesian model, allowing us to exploit epistemic uncertainty.

Machine learning accelerated flow field prediction for stirred vessels

Mahdi Naderibeni[1], Liang Wu[2], David Tax[1]
[1]TU Delft
[2]dsm-firmenich

Accelerated simulation of fluid flow inside stirred vessels is a key step towards developing digital twins for industrial scale (bio-)chemical processes. In this work we leverage Physics-Informed Neural Networks to learn the velocity and pressure field of stirred vessels for a range of operating conditions.

Uncovering the Potential of Applying PCA on Raman Spectra of Biochar

Raymond Chen and Ewa J. Marek
University of Cambridge

Raman spectroscopy is a powerful tool, which can be used for analysing nanostructure features of biochar, but the plethora of methods for band fitting and analysis makes finding the right approach for uncovering all relevant characteristics difficult, and questions what information remains undiscovered in Raman spectra. To address both problems, we applied principal component analysis (PCA) on Raman spectra from biochar produced by pyrolysing dried Sargassum algae in N2 at 250-450°C. The loadings of the first two PCs showed an unexpected feature with a decreasing signal at a commonly neglected area in Raman spectra of around 800 cm 1, possibly indicating new, earlier unrevealed information. Furthermore, the pyrolysis temperature of the biochar samples could be distinguished by solely considering the values of the first two PCs, even if PCA is only applied on a commonly neglected region of biochar Raman spectra (500 1000 cm 1), further indicating unexplored information. These results demonstrate the capabilities of PCA in analysing Raman spectra of biochar samples with different characteristics.

Self-supervised graph neural networks for polymer property prediction

Qinghe Gao, Tammo Dukker, Artur Schweidtmann, and Jana Weber
TU Delft

The estimation of polymer properties is of crucial importance in many application areas such as energy and healthcare, where graph neural networks (GNNs) have recently shown promising results. However, the supervised learning approach for training GNNs is hindered by the extensive need for annotated data, which is expensive and time-consuming to collect. We explore the potential of self-supervised learning to pretrain GNNs on this unlabeled data, which can then be fine-tuned for specific property predictions with a smaller set of labelled data. By implementing a novel polymer graph representation and investigating various self-supervised learning setups, our approach significantly enhances performance in data-scarce scenarios, achieving substantial improvements in R2 for predicting key polymer properties over supervised methods.

Automated detection and segmentation of intracranial hemorrhage suspect hyperdensities in non-contrast-enhanced CT scans of acute stroke patients

N.Schmitt[1], Yahia Mokli[1,2], C.S. Weyland[1], S. Gerry[3], C. Henweh[1], P.A. Ringleb[1], and S. Nagel[1]
[1]Heidelberg University Hospital
[2]Giessen University Hospital
[3]University of Oxford

Artificial intelligence (AI)-based image analysis is increasingly applied in the acute stroke field. Its implementation for detecting and quantifying hemorrhage suspect hyperdensities in non-contrast-enhanced head CT (NCCT) scans may facilitate clinical decision-making and accelerate stroke management. In our study based on a challenging NCCT dataset of 160 patients with suspected acute stroke, the AI-based algorithm - developed by the company Brainomix - reliably detected acute intracranial hemorrhages and precisely quantified the volumes of intraparenchymal hemorrhages.

Autocorrection of process flowsheets using large language models

Lukas Schulze Balhorn, Marc Caballero, and Arthur M. Schweidtmann
TU Delft

In the (bio)process engineering domain, Process Flow Diagrams (PFDs) and Process and Instrumentation Diagrams (P&IDs), collectively known as flowsheets, are susceptible to errors that can lead to safety hazards and unnecessary expenses. Current manual correction processes, relying on expert knowledge, are both time-consuming and expensive. Addressing this issue, we propose a proof-of-concept for automatic correction (autocorrection) of flowsheets, leveraging Large Language Models (LLMs). Our approach treats flowsheet correction as a translation task, using the encoder-decoder model T5 and SFILES 2.0 notation. To train the LLM, we generate a synthetic dataset with 500,000 flowsheet pairs, including both erroneous and error-free patterns. The model demonstrates remarkable top-1 (82.1%) and top-5 (84.1%) accuracies, showcasing its ability to detect and correct design errors in synthetic flowsheets.

Designing reduced genomes using machine learning and whole-cell modelling

Ioana Gherman[1], Joshua Rees-Garbutt[1], Wei Pang[2], Zahraa Abdallah[1], Thomas Gorochowski[1], Claire Grierson[1], and Lucia Marucci[1]
[1]University of Bristol
[2]Heriott-Watt University

Whole-cell models are mathematical models designed to capture the function of all genes and core processes within a cell. They have been successfully used for applications such as minimal genome design, identifying novel functionalities of a biological system and guiding experimental pipelines. Here we show how machine learning can be used to reduce the computational complexity of whole-cell models for applications such as minimal genome design. The results demonstrate that using machine learning algorithms to emulate the behaviour of whole-cell models it is possible to speed up simulations by up to 16 times and to generate a reduced E. coli genome with 40% of the in silico modelled genes removed.

Reward functions for machine-learning generated experimental designs

Helena H. Thygesen
Independent statistical consultant

Traditional Optimal Design Theory aims at accurate parameter estimates. But when the design isn't based on a parametric model, or the goal of the study is to provide accurate outcome predictions or classifications, a "utilitarian" approach such as the Mean Objective Cost of Uncertainty (MOCU) may be more appropriate. This poster gives examples of the use of utilitarian reward functions in the optimization of in vitro experiments and adaptive clinical trials.

Towards model-based design of experiments to optimize bioprocesses

Ana Helena V. Caetano, Julian Kager, and Krist V. Germaey
Danmarks Tekniske Universitet

The pharmaceutical industry has shifted towards a Quality by Design (QbD) approach and a cornerstone of this strategy is the Design of Experiments (DoE) methodology. However, when applied to biological processes, traditional DoE can be of very limited information. To address the challenges of non-linear and dynamic processes, model-based design of experiment (MBDoE) approaches are developed, which leverage in silico models to dynamically design experiments. Besides significantly reducing the number of experimental iterations, MBDoE improves process understanding and efficiency and is thus of interest to test and develop further this framework.

Guided antibody sequence and structure design using developability properties

Amelia Villegas-Morcillo, Jana M. Weber, and Marcel J.T. Reinders
TU Delft

Recent advances in deep generative methods have allowed antibody sequence and structure co-design. This study addresses the challenge of tailoring the highly variable complementarity-determining regions (CDRs) in antibodies to fulfill developability requirements. We introduce a novel approach that integrates property guidance into the antibody design process using diffusion probabilistic models. This approach allows us to simultaneously design CDRs conditioned on antigen structures while considering critical properties like solubility and folding stability. Our property-conditioned diffusion model offers versatility by accommodating diverse property constraints, presenting a promising avenue for computational antibody design in therapeutic applications.