Virtual 2nd Annual Meeting of the de.NBI Industrial Forum

Abstracts of scientific presentations

November 25th 2021, 01:00 – 04:25 pm (CET)
Video-System: ZOOM
Login data: Send to you after Registration

Artificial Intelligence projects of the Industrial Forum members Part I

Higher-order Machine Learning Models Act as an Approximation of Biological Regulatory Mechanisms

S. Schultheiss, Computomics GmbH, Tübingen

Plant breeding needs to accelerate to supply new varieties for a growing population and a rapidly changing climate. New breeding technologies like gene editing and genomic prediction help bring about this acceleration, but are often used independently without sharing useful preexisting knowledge. Here, we present a method for discovering both new gene editing targets and higher-accuracy predictions. By using interpretable machine learning models specifically developed for genomic data, complex genetic mechanisms can be rapidly understood and visualized. Multi-genic traits show up in the visualization of feature importance and positional genomic importance. We apply this method to a dataset derived from a shelf-life experiments for 200 Capsicum varieties. Genotypes, manual scoring and plant image data are correlated to train a regression machine learning algorithm that identifies an ethylene-linked gene cluster responsible for shelf life and plant senescence. New breeding technologies require these kinds of insights into biological regulation to identify new editing targets quickly and reliably.

Open Ecosystem for integrated Machine-Learning Workflows

B. Schwarz, Carl Zeiss Microscopy GmbH, Jena

Today's imaging systems can create huge amounts of large dataset automatically and the main challenge is now to extract "Actionable Information" from those images and therefore an important topic for ZEISS.

One of the main toolsets to tackle those computer vision challenges is the usage of Artificial Intelligence to process, segment and classify images. The challenge however is to create an flexible and open technology stack to keep up with this very fast developing field and to be able to incorporate new tools from various open-source software (OSS) projects, which are an integral part of the AI computer vision community.

Our approach is to use an open and flexible ecosystem of different AI-based tools, which heavily rely on OSS tools, standard data formats and the usage of Python, Cloud and Docker™ technologies. This ecosystem allows us to re-use code for many different projects while giving us the option to still customize the end-user applications depending on the specific needs and requirements.

Promises and Progress in Applications of AI in R&D Sanofi-Aventis Deutschland

T. Klabunde, Head Data & Data Sciences Germany, R&D, Sanofi-Aventis Deutschland, Frankfurt

In Sanofi R&D, the Datal & Data Sciences (DDS) platform leverages artificial intelligence and modern methods of data analysis to make decisions smarter, easier and faster. Our goal is to translate data into knowledge in order to find the right drug for the right patient and to accelerate the identification of “first-in-class” and “best-in-class” drugs. DDS is linked to our partners from all areas of R&D via a “hub-and-spoke” organizational model in order to provide support in the application of innovative methods and artificial intelligence. Examples are the generation of virtual patients to perform in-silico clinical trials, the AI-supported analysis of medical images and the model-based design of clinical studies.

Artificial Intelligence projects of the Industrial Forum members Part 2

Adventitious Virus Testing – Exploring the Machine Learning’s Potential

H. Tran, IDT Biologika GmbH, Dessau

IDT Biologika is a leading biologics CDMO providing development and manufacturing of viral vaccines, viral vectors for gene and immune therapeutics and other biologics. Our products have to be free of unwanted contaminations like viruses and microorganisms in order to ensure human health. The testing for adventitious agents based on Next Generation Sequencing is a key step within our product quality control. We developed a bioinformatics workflow using sequence alignment and database queries based on k-mers which shows high sensitivity and specificity in virus detection. Previously, we successfully applied machine learning techniques to identify chemical compounds showing antibacterial and antiparasitic activities. Next, we will exploit the potential of machine learning in classification for the detection of adventitious virus and microorganisms.

AI or not AI – That is the Answer!

C. Stephan, Kairos GmbH, Bochum

Artificial intelligence faces a problem when it comes to the artificial simulation of intelligent structures. This is because an artificial intelligence for which the desired independent learning is actually desired, is fundamentally different from statistical evaluations [1]. Although these seem "intelligent" at first glance, they rather represent a different form of aggregated data presentation without any learning processes.

The basis for the creation of such a learned AI model is the use of information in digital form. However, this development faces a number of hurdles: The lack of IT infrastructure in healthcare institutions is accompanied by outdated data processing, some of which is still based on paper. At the same time, already digitized data stocks are not shared due to technical hurdles and remain untapped for medical research. Besides, they are often not interoperable and therefore useless for research.

With the help of AI, new knowledge can be harvested. After all, it is important that AI does not only grow on the basis of mere data stocks to knowledge that is already known. Rather, it is only through active learning of the system that value can be added to research and thus to the patient. In this way, AI can revolutionize the healthcare system by gaining new insights about the diseases of our time.

These developments result in initiatives like SepsisDataNet. The project combines multidisciplinary institutions to analyze the samples collected along with clinical data, evaluates the resulting Big Data findings, and translates the results into personalized treatments. This is realized by the development of a decision support module, which not only increases the sensitivity and specificity of the newly developed sepsis bioassay by means of classification models, but also supports physicians in making personalized therapy decisions based on the analyzed immune status. The success of this project prompted the creation of CovidDataNet to support personalized treatments of COVID-19 based on the combination of sample and intensive care data.

[1] Wikipedia, https://de.wikipedia.org/wiki/K%C3%BCnstliche_Intelligenz, 2018

Artificial Intelligence in a plant breeding company

J.-C. Richter / A. Menze, KWS SAAT SE, Einbeck

The field of AI, specifically Machine Learning, has garnered a lot of renewed interest in recent years. Both unsupervised as well as supervised learning approaches have long been core assets of professional plant breeders. Respective examples would be dimensionality reduction and clustering for the evaluation of genetic diversity within a given material and modeling of Genotype-Phenotype relationships to predict the performance of material that was not tested in the field.

The enormous increase in data availability, combined with recent developments in deep learning and high performance computing enables various new applications of ML in plant breeeding and yields the potential to improve existing ones. Using supervised learning methods especially in the area of image analysis for high quality and high throughput phenotyping is only one such example. Predictive breeding is strongly expected to benefit from the incorporation of environmental covariates alongside genotypic and phenotypic data, especially in complex quantitative traits such as yield.

This talk gives an overview about the potential of new and improved applications of ML in plant breeding.

NFDI projects supported by de.NBI

The National Research Data Infrastructure (NFDI) and NFDI4Biodiversity

F O Glöckner
MARUM - Center for Marine Environmental Sciences, University of Bremen
Alfred Wegener Institute - Helmholtz Center for Polar- and Marine Research
GFBio e.V. - German Federation for Biological Data e.V.
de.NBI e.V. - German Network for Bioinformatics Infrastructure e.V.

In November 2018, the German Science Minister Conference GWK agreed on a 10-year initiative to structure research data services across different scientific domains in Germany. Now, 3 years later, the National Research Data Infrastructure NFDI has become operational, with a Directorate established in Karlsruhe and a first cohort of nine domain-oriented consortia started on October 1st, 2020. A second set of consortia have been funded and a third round is planned for 2022. The final NFDI is expected to unite roughly 30 thematic consortia, covering all research domains in Germany.

NFDI4Biodiversity (Biodiversity, Ecology & Environmental Data) is a consortium of 49 partners dedicated to deliver practical tools for researchers and improve data availability for environmental studies, ecology as well as decision making in politics and industry. At the core of NFDI4Biodiversity are data on the biodiversity of animals, plants and microorganisms. NFDI4Biodiversity builds on previous work by the German Federation for Biological Data (GFBio), which has been funded by the DFG since 2014. As part of this project, the data centres of natural history museums and established data archives such as the Data Publisher for Earth and Environmental Science PANGAEA and the European Nucleotide Archive ENA were connected to a common portal, workflows for data submission were standardized, and common (meta)data standards as well as tools and services for research data management (FDM) were established. NFDI4Biodiversity extends the range of activities by incorporating use cases from the broader biodiversity community, including associations, governmental agencies and citizen science under the common vision of a Research Data Commons.

NFDI4Micobiota – FAIR data, analytical workflows and much more for the microbiology community

K. Förstner, ZB MED, Cologne

As part of the German NFDI (National research Data Infrastructure), NFDI4Microbiota aims to enable researchers to translate research data easily into a deep understanding of microbial species and their interactions on a molecular level. To achieve this NFDI4Microbiota will support the microbiology community with access to data, analytical services, data/metadata standards and training. A diverse set of use cases with medical, agricultural and biotechnological topics will be covered to ensure that the broad needs of the community are met.

de.NBI Coordinator: Prof. Dr. A. Pühler
Administration Office:
c/o Center for Biotechnology
Bielefeld University
33594 Bielefeld
Germany
contact@denbi.de
www.denbi.de