

Programme

2nd Symposium on Applied Deep Learning in Bioinformatics

16 Aug 2024

10:00 - 10:30 **Arrival and Coffee**

10:30 - 10:35 **Welcome by Jan Gorodkin and Lars Juhl Jensen, University of Copenhagen**

10:35 - 11:15 **Long Talk. Katerina Nastou, SSI. Title: "Extracting protein-protein interactions from the literature with deep learning-based text mining"**

A bstract :

STRING is a database with billions of protein-protein interactions, systematically collected from diverse sources, including automated text mining of the scientific literature. In this presentation, I will discuss the use of deep learning-based language models to enhance the "physical interactions" mode in STRINGv12. Additionally, I'll outline our current progress and plans toward introducing a new mode in STRINGv12.5, focusing on interaction type, sign, and direction, to improve the utility of existing networks for mathematical modeling of biological systems and logical inferences of upstream events based on downstream effects observed in biological datasets.

B io :

Katerina Nastou is a postdoctoral researcher at Statens Serum Institut. Previously, she was Assistant Professor at JensenLab, where she contributed to the influential STRING molecular interaction database, known for its over 100,000 monthly users and 30,000 citations, and recognized as a Core Data Resource by ELIXIR. Her focus lied in enhancing the database with deep learning-based text mining. She also specialized in co-occurrence-based relation extraction, deciphering connections between genes, proteins, chemicals, and diseases from the biomedical literature, where she also explored the potential for enhancing this method with deep learning approaches. Her expertise extends to named entity recognition for organisms and lifestyle factors, all aimed at aiding scientists in the synthesis of mechanistic hypotheses, pertinent to their scientific questions."

11:15 - 11:35 **Short Talk. Ying Sun, University of Copenhagen. Title: "Enhanced CRISPR base editing design from data generation and deep learning"**

Abstract: TBA

11:35 - 11:55 **Short Talk. Alberto Santos Delgado, DTU. Title: " Constructing Microbial Association Networks for Community Analysis"**

Abstract:

Microbial communities serve as the essential foundation of ecosystems, playing pivotal roles in sustainability and health. Yet, their complex interactions and the emergent functions stemming from them remain largely misunderstood. Recent advances in high-throughput omics technologies have facilitated the study of microbiomes from complex environmental samples. We are developing a graph-based framework that allows seamless integration of meta-omic datasets together with their biological context. This framework infers potential interactions among microorganisms through Microbial Association Networks (MANs) in different environments. These MANs and their additional metadata are integrated into a knowledge graph that enables the identification of microbial communities and their possible roles in their environment. Moreover, we provide an analytical toolbox to query, investigate, and visualize MANs signatures to enable researchers to understand microbial communities' influence in environmental stability, ecosystems resilience, and disease mechanisms.

Bio:

Alberto Santos is a senior researcher at the NNF Center for Biosustainability at the Danish Technical University, leading the Multi-omics Network Analytics research group (MoNA) and the Informatics Platform. Alberto's research focuses on **data science** applied to multidisciplinary topics in **biology** and aiming to address complex challenges in **environmental and human health**. One of the main research areas is on **Knowledge Graphs (KG)** for omics data, enabling the analysis and interpretation of these complex datasets. MoNA group is providing the research community with standard frameworks to build, evaluate, and mine KGs in a reproducible manner.

12:00 - 12:30 **Lunch break**

Light Lunch will be provided for all signed up participants

12:30 - 13:10 **Keynote by Arne Elofsson, University of Stockholm. Title: "Machine learning for protein-protein interactions"**

13:10 - 13:30

Short Talk. Davide Placido, Rigshospitalet. Title: "Early detection of pancreatic cancer using deep learning"

Abstract:
TBA

Bio:

Davide Placido obtained a bachelor and master degree in biomedical engineering from the Polytechnic University of Turin. He conducted his master thesis project at Technical University of Denmark (DTU), where he worked on prediction models using high-frequency data for ICU patients. Following this experience, he obtained a Ph.D. at the Novo Nordisk Foundation Center for Protein Research, supervised by Søren Brunak. During this time, he worked on the development of different prognostic models for the hospital departments, particularly predicting mortality for ICU patients and clinical deterioration in the general wards. He also developed a prediction model for early detection of pancreatic cancer. During the PhD he was a visiting researcher at Harvard Medical School, where he worked on the validation of the pancreatic cancer model on external datasets to assess the model generalizability. He continued for a postdoc in the same group, focusing on the integration of new data modalities for early detection of cancer. He has now started a new position as Data Scientist at the Intensive Care Unit, Rigshospitalet, Denmark. In this position he is working on developing ETL pipelines for OMOP standardization and ML applications for improving and informing the clinical trials conducted at the ICU.

13:30 - 13:50

Coffee Break

13:50 - 14:10

Short Talk. Simon Rasmussen, University of Copenhagen. Title: "Multi-omics and modal data integration"

Abstract: TBA

14:10 - 14:50

Long Talk. Veit Schwammle, SDU. Title: "Deep learning in proteomics: Achievements and challenges"

14:50 - 15:00

Closing remarks by Jan Gorodkin and Lars Juhl Jensen, University of Copenhagen