PS13. Bioinformatics

Organised by Roald Forsberg (Raven biosciences and JADBio)

Session overview

Over the course of the last 20 years changes in assay and sensor technologies have transformed life-science to become one of the most data rich disciplines that exist. This development affects all areas of life-science from genetics to agriculture and has brought data to the centre stage of all activities.

In this session we have invited a broad group of experts from both academia and industry. The aim is to give you a flavour of the diversity and depth of activities involved in the modelling, analysis, and application of life-science data ranging from precision medicine and vaccine development to precision agriculture.

Speakers

- Michael Schantz Klausen and Christian Thygesen - both from Evaxion
- Mikkel Hovden – Molekylær Medicinsk Afdeling, Århus Universitetshospital
- Rasmus Froberg Brøndum, Center for Clinical Data Science, Aalborg University and Aalborg University Hospital
- Morten Birk, FieldSense A/S

Programme

10.15 –	Welcome
10.15 – 10.45	**Rasmus Froberg Brøndum** *Unsupervised clustering, from idea to clinical implementation*
10.45 – 11.15	**Mikkel Hovden** *Data science tools for detection of cancer in the blood, using a neural network model of the background errors in next-generation sequencing data*
11.15 – 11.45	**Michael Schantz Klausen and Christian Thygesen** *Increasing immunogenicity in subunit vaccines using in silico peptide grafting*
11.45 – 12.15	**Morten Birk** *Replacing gut-based farming with data-driven decision making*

Abstracts

Unsupervised clustering, from idea to clinical implementation

Rasmus Froberg Brøndum
Center for Clinical Data Science, Aalborg University and Aalborg University Hospital

Unsupervised clustering methods may be applied to -omics data to enable discovery of new biological subtypes. This talk introduces the concept, details some applications in the context of cancer treatment, and provides a framework for addressing the inherent mis-labelling problem in downstream analysis.

Data science tools for detection of cancer in the blood, using a neural network model of the background errors in next-generation sequencing data

Mikkel Hovden Christensen Ph.D.[1], Simon Drue M.Sc.[1], Mads Heilskov Rasmussen Ph.D.[1], Amanda Frydendahl M.Sc.[1], Iben Lyskjaer Ph.D.[1], Christina Demuth M.Sc.[1], Jesper Nors MD.[1], Kaare Andersson Gotschalck Ph.D.[2], Lene Hjerrild Iversen Ph.D.[3], Claus Lindbjerg Andersen Ph.D.[1], Jakob Skou Pedersen Ph.D.[1]

[1]Department of Molecular Medicine, Department of Clinical Medicine, Aarhus University
[2]Department of Surgery, Regional Hospital Horsens, Denmark
[3]Department of Surgery, Aarhus University Hospital, Denmark

Cancer can be detected in the blood by finding fragments of mutated DNA. Using next-generation sequencing (NGS) to generate data specifically for known cancer genes, mutated DNA can be captured and cancer relapse can, in principle, be revealed. However, the detection is complicated by a low signal-to-noise ratio arising from sequencing errors that occur at a frequency similar to the mutation signal. To increase detection performance and the accuracy of individual mutation calls, we have developed DREAMS (Deep read-level modelling of sequencing errors) – a background error model for NGS data, that uses a neural network to exploit detailed information about the context of errors. Based on DREAMS and a probabilistic expectation-maximization (EM) approach, we have developed tools for detection of individual cancer mutations and for providing a combined cancer score, with a performance that exceeds state of the art methods.

Replacing gut-based farming with data-driven decision making

Morten Birk
FieldSense A/S

Farming is a practice where actions are often based on gut-feelings, intuition and experience. At FieldSense, we provide farmers with valuable agricultural data gathered from various sources such as hyper-local weather observations, soil sampling and satellite monitoring, to support and encourage actions based on a data-driven approach.

However, farmers generally follow a pragmatic problem solving approach, and thus one of our key priorities is to make data available at the proper level of abstraction, and in an intuitive manner that allows farmers to adopt our technologies and implement them in their day-to-day decision making.

I will talk about how we use deep-learning to interpret our data and extract high level features usable for farmers. Furthermore, I will demonstrate how we pipeline our real-time data ingestion to ensure that recent data used for modeling is quickly available without too much noise.