Master thesis in Bioinformatics

Professor Jan Komorowski

Department of Cell and Molecular Biology

Project proposal

Systemic Lupus Erythematosus subtyping

The machine learning field continues to develop new applications for predicting outcomes and explaining phenomena in the domain of biology and medicine. Deep learning and neural networks (NNs) have had success in different subfields for their adaptability to different kinds of inputs and deployment on available hardware. However, NNs offer limited help in explaining the features and their values for the models they produce. In order to address this problem other kinds of machine learning methods are available, such as rule-based models, which provide transparency, and thus interpretability.

Rheumatological diseases constitute a large heterogenous group of pathologies that involve miss-regulation of the immune functions at different levels. Within this group, Systemic Lupus Erythematosus (SLE) is a devastating autoimmune disease with a wide clinical range that might include neurological, renal, hematological, cardiological and/or pulmonary damage. Our previous research by Yones *et al*. has successfully shown how ruled-based machine learning (RBML) models and rule networks (RN) can be used to reveal key gene sets distinguishing pediatric SLE patients with low and high disease activity, DA1 and DA3, respectively. A distinguishing property of that work is a creation of an RBML workflow developed by Yones *et al* that helps establish clinical and therapeutic stratification of patients.

We propose to use the approach of Yones *et al.* to characterize Lupus nephritis patients in comparison to low disease activity patients (DA1). Lupus nephritis is a disease caused by SLE and is a serious renal condition in which the tissues of the kidney become inflamed and may lead to kidney failure. To fulfill the aims of the proposal the workflow defined by Yones *et al* is to be made into a form of a pipeline with well-defined and documented inputs and outputs, and then applied to new transcriptomic data on SLE. The results from this project may deepen our understanding of SLE and rheumatological diseases in general.

Candidates are required to possess strong programming skills in languages commonly used for data science, R, Python or Julia, as well as knowledge of machine learning algorithms. Experts in computation, molecular biology and machine learning will support and supervise the proposed project.

Contact:
Daniel Rivas: daniel.rivas@icm.uu.se
Jan Komorowski: jan.komorowski@icm.uu.se