ROMAN SCHULTE SASSE

schulter.github.io

I am a passionate data scientist and machine learning engineer with experience in deep learning, robotics, and computational biology. I care about applying advanced ML models to complex problems to gain knowledge and care about interpretability of predictive systems.

Email: schultesasse@posteo.de
Phone: +49 177 216 10 37
LinkedIn: linkedin.com/in/roman-schulte-sasse
Twitter: @SchulteRoman
Github: github.com/schulter

EDUCATION

PhD student - Max Planck Institute for molecular genetics
JANUARY 2017 – JANUARY 2021

During my PhD, I developed and adapted a deep learning model to learn about molecular mechanisms leading to disease. I used graph deep learning to integrate different molecular experimental data of patients to predict cancer-related genes. Using feature interpretation methods for neural networks, we could identify the causes leading to the predictions of individual genes.

Master's degree Computer Science – Freie Universität Berlin
OCTOBER 2014 – OCTOBER 2016

took focused courses on machine learning during this time and reinforced linear algebra and computer vision. Master thesis about detecting patterns in DNA sequences using convolutional Restricted Boltzmann Machines.

Bachelor's degree Computer Science – Freie Universität Berlin
OCTOBER 2009 – JUNE 2014

I studied computer science at the free university of Berlin and started working with the FUmanoids, the soccer playing robot team of the university with a Bachelor thesis on robot localization based on landmarks identified using computer vision.

SKILLS

Machine Learning	●●●●●●●●●
Data Science	●●●●●●●●○
Deep Learning	●●●●●●●●○
Bioinformatics	●●●●●●○○○
Network Analysis	●●●●●○○○○
Computational Cancer Biology	●●●●●●●○○

LANGUAGES

- **German**: Mother tongue
- **English**: Fluent
- **French**: Fluent

WORK EXPERIENCE

Max Planck Institute for molecular genetics — **Student Research Assistant**
JANUARY 2016 – NOVEMBER 2016

I worked on the publication of our method to identify transcription factor binding sites with convolutional restricted Boltzmann machines and extended the framework to multiple layers (a deep belief network), producing an unsupervised deep learning architecture.

Humboldt Universität zu Berlin — **Student Research Assistant**
AUGUST 2015 – DECEMBER 2015

I worked on the automatic tracking of tendons in ultrasound videos for research at the sports faculty. In this work, I continued developing a MATLAB framework for semi-automated tracking of the tendons.

FUmanoids — **Student Developer**
JANUARY 2013 – SEPTEMBER 2015

I worked on modeling and computer vision in C++. For my bachelor thesis, I developed a localization framework based on particle filters (see publications) and afterwards developed an SVM-based ball recognition. I also implemented a strategy model.

PUBLICATIONS

Integration of Multi-Omics Data with Graph Convolutional Networks Identifies New Cancer Genes and their Associated Molecular Mechanisms, — **Nature Machine Intelligence**
Accepted 2021

We predicted cancer genes by integrating several molecular data types such as mutation rates, DNA methylation, gene expression and protein-protein interaction data. To integrate these data types, we made use of graph convolutional networks and use gradient-based a posteriori feature interpretation methods to disentangle the molecular alterations of our classifications.

TriPepSVM: de novo prediction of RNA-binding proteins based on short amino acid motifs, 2019 — **Nucleic Acid Research**

We used support vector machines with string kernels to predict RNA-binding proteins from the protein sequence of the protein in question for human and bacteria. We find tri-peptides (sequences of three amino acids) allow to distinguish well between RNA binders and non-binders and furthermore find that several RBP-enriched tri-peptides occur more often in structurally disordered regions of RBPs.

Graph Convolutional Networks Improve the Prediction of Cancer Driver Genes, 2019 — **Lecture Notes of Computer Science**

We presented the work on data integration to predict cancer genes (mentioned above) at the ICANN conference. This paper has a more methodological focus and provides fewer biological insights.

Unsupervised learning of DNA sequence features using a convolutional restricted Boltzmann machine, 2017 — **BioRxiv**

We used unsupervised convolutional architectures to identify transcription factor binding sites. Transcription factors (TFs) bind preferentially to genomic regions, thereby regulating the expression of other genes. Our method identifies TF binding sites from experimental data using a convolutional restricted Boltzmann machine.