

Analysis of DNA sequences using artificial intelligence

All living organisms are made up of cells that contain DNA. The DNA sequence, i.e. the order of the four bases A, C, G and T, is the blueprint of the molecular building blocks of the cell. Correct and accurate analysis of DNA is, therefore, a cornerstone in understanding life and central in many research fields, including medicine and evolution. However, the linguistic complexity of DNA is high – at least as high as natural languages – which makes it hard to correctly interpret. The meaning of the DNA sequence, even in the simplest of organisms, is to a large extent unknown.

In this PhD project, we will use artificial intelligence to develop novel state-of-the-art methods for the analysis of DNA sequences. We will initially work with transformers, a new class of AI algorithms originally developed for the interpretation and translation of natural languages. Transformers use a process called self-attention which enables the description of complex long-range dependences – a phenomenon common in DNA sequences. By combining the transformers with deep neural networks, it is possible to do both prediction and classification. This includes pattern recognition and identification of structures of biological relevance.

An important part of the project is to apply the new methods to real biological problems. Here we will work primarily with antibiotic-resistant bacteria. These bacteria have acquired specific genes ('antibiotic resistance genes') enabling the bacteria to survive antibiotic treatment, which make them a dangerous threat to human and animal health. The exact patterns of many antibiotic resistance genes are not known, which means that they cannot be detected. In this PhD project, we will apply our developed AI methods to study large volumes of bacterial DNA data to identify novel antibiotic resistance genes. These predictions will be further studied in detail and also experimentally verified by our collaborators working with experimental microbiology. The findings from our AI method, therefore, has the potential to directly impact the ability to perform accurate diagnostics of bacterial infections and our knowledge about the evolution of antibiotic-resistant bacteria.

This project combines research in artificial intelligence and bioinformatics. After completion, the student will have gained a highly competitive skillset including methods from AI, big data, large-scale data analysis, and statistics, all of which are currently in high demand from both academia and industry. The project is located in a highly successful interdisciplinary research group and interaction with scientists from other fields will be central. The group have a large number of active national and international collaborations within the fields of infectious diseases and the evolution of microorganisms. The group have also several connections to the biotech industry.

If you have any questions, do not hesitate to contact Erik Kristiansson at erik.kristiansson@chalmers.se.