AI for Genomics: From CNNs and LSTMs to Transformers

Course Overview

This course explores the application of modern AI architectures—Convolutional Neural Networks (CNNs), Long Short-Term Memory networks (LSTMs), and Transformers—to genomic and metagenomic data. Students will gain practical experience through hands-on coding labs and interactive notebooks, learning how to model sequence data, extract biologically meaningful features, and interpret results. Emphasis is placed on real-world applications, including prediction of genomic functional elements, sequence classification and source tracking, as well as biological sequence generation.

Requirements

- Basic knowledge of molecular biology and genomics (e.g., genetic variation)
- Familiarity with Python/R programming

Learning Outcomes

By the end of this course, participants will be able to:

- Understand and implement LSTM and CNN architectures for genomic sequence data
- Apply attention mechanisms to improve genomic feature extraction and prediction
- Train simple Transformer models for sequence classification or functional element prediction
- Use notebooks to run and modify ML & DL workflows for genomics research
- Interpret model outputs and assess performance using biological context

Program

Day 1:
A) Introduction to artificial neural networks and concepts of convolution and recurrence.
B) Encoding genomics data for CNNs and LSTMs in Keras and TensorFlow.

Day 2:
A) Training CNNs and LSTMs for sequence classification and functional element prediction.
B) Applying NLP concepts to genomics data: bag-of-words and Word2Vec models.

Day 3:
A) Metagenomic source tracking with CNNs: microbial community sequence annotation.
B) Biological sequence generation with attention, transformer model for genomics applications.