

UNIMORE
UNIVERSITÀ DEGLI STUDI DI
MODENA E REGGIO EMILIA



Lab Introduction

Marta Lovino, PhD
2023/2024



Outline

01 Course schedule

02 Lab topics

03 Lab setup - Live

Course schedule

AI in Bioinformatics

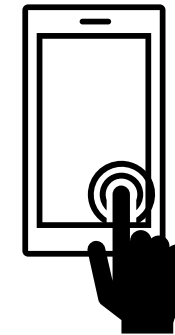
9 CFU – I Semester

Monday, 8-10 → to be rescheduled

Tuesday 9-11

Thursday 14-16

Scan me



<http://bit.ly/45Ui8kx>

Sept 18 – Dec 14, 2023

Always check the course website and Slack channel to stay up-to-date on the course schedule

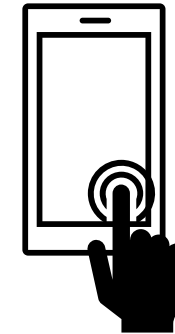
https://join.slack.com/t/2023aiinbioin-esm6867/shared_invite/zt-23e7dweoy-CyFqD5exlQZfoAgEfPnzMA

Course schedule

Consulting time:

- During Lab lessons
- Schedule an appointment with the Lab reference person, writing an email

Scan me



<http://bit.ly/45Ui8kx>

Always check the course website and Slack channel to stay up-to-date on the course schedule

https://join.slack.com/t/2023aiinbioin-esm6867/shared_invite/zt-23e7dweoy-CyFqD5exlQZfoAgEfPnzMA

Course Objectives

- **This course introduces advanced AI models and problems for biomedical imaging and data analysis.**
- **Students will learn to address challenges such as:**
 - Data scarcity and annotation
 - Reliability and fairness of AI models
 - Integration of multiple data sources
 - Continuous learning
- **Through practical examples and lab sessions, students will learn techniques for:**
 - Multiple-instance learning (MIL)
 - Bias correction
 - Multimodal learning and co-learning
 - Statistical methods for AI model evaluation and data analysis
 - 2D and 3D segmentation (e.g., using UNet and Visual Transformers)
- **Specific topics will be covered in thematic seminars, such as continual learning, etc.**

Lab Prerequisites

•Required:

- Solid foundation in computer science, including algorithms, data structures, and programming (e.g., Python, C, C++, or Java).

•Recommended:

- Completion of the course MACHINE LEARNING AND DEEP LEARNING (even without passing) or basic knowledge of machine and deep learning.

•Support:

- Students who do not meet the prerequisites will receive explanations and offline materials to understand the basics of the course.

Lab Prerequisites, details

Required:

- Expressions and Variables
- String Operations
- List and Tuples
- Dictionaries
- Sets
- Conditions and Branching
- Loops
- Functions
- Exception Handling
- Objects and Classes
- Reading Files with Open
- Writing Files with Open
- Pandas
- Numpy

Lab Topics

Course Introduction (0.5 CFU)

- Course Objectives, Course Structure and Exam Rules, Setup for Lab Sessions
- Introduction to Bioinformatics:
 - Definition
 - Relevant problems and opportunities in research, industry, and economy
 - Computational problems and requirements of AI models

Introduction to Biomedical Imaging (1 CFU)

- Introduction to biomedical images (e.g., WSI Gigapixels images);
- Focus on relevant research topics and challenges in the design of AI models and systems:
 - Data dimensionality
 - Scarcity of annotations and labels
 - Uncertainty in annotations and labels
 - Data containing multiple annotations and labels (i.e., multi-labels)
 - Data at different resolutions
 - Data bias
 - Progressive knowledge, and therefore continuous learning
- Introduction to AI topics that address the above problems

Lab Topics

Deep learning architectures (2.5 CFU)

- **Graph neural networks** (theory and laboratory)
- **Bayesian neural networks** (theory and laboratory)
- **Transformers and Visual Transformers** (theory and laboratory)
- Self-learning: examples of architectures (e.g., DINO)
- **Multimodal learning and Co-learning**: examples of architectures for integrating data from multiple sources (e.g., text, video, images, sound, etc.) with (partial/total) absence of data (theory and laboratory)

Statistical methods for data science (0.5 CFU)

Biomedical imaging (2.5 CFU)

- **Multiple-Instance Learning** (MIL):
 - definition
 - examples of MIL architectures (theory and laboratory)
- **Debiasing AI**:
 - examples of architectures for debiasing (with or without prior knowledge of the bias)
- 2D and 3D segmentation: UNet, Visual Transformers (theory and laboratory)

Lab Topics

Molecular Data: DNA and RNA-sequencing (2 CFU)

- Brief description of **genomic sequencing technologies, data formats, and most used tools for sequence analysis** (e.g., Botwie, Arriba, CNV caller, GenomeSpy/Genome Browser) (theory and laboratory)
- **Analysis of genomic sequences:**
 - Examples of architectures based on Transformers (BERT) and Perceivers
 - Use cases: prediction of genetic and protein expression from genomic sequences (theory and laboratory)
 - Explainability of results and model through inspection of attention mechanisms
- **Multimodal learning for molecular data:** examples of integration between genomic sequences and molecular data (theory and laboratory)

Project presentation for the final exam.

Material

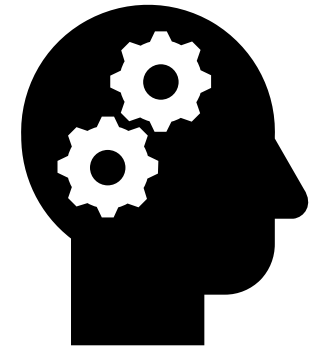
Lessons will be held in standard classrooms, not laboratories.

Bring your own laptop to classes! No particular hardware configuration is required. Please come back to us if you have problems.



Aim of lab classes:

practice in solving problems, improve programming skills,
work group



Local machine and Colaboratory



You can freely program on local machine or on your Colaboratory account.

However, a few genomic tools require local installation.

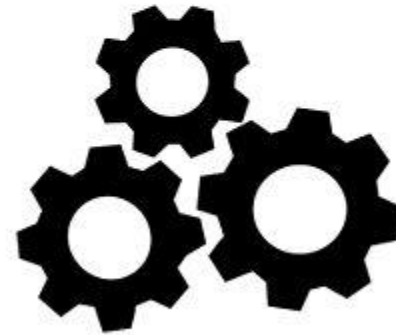
Github

Importance of version control and backup!



FAIR principles

F_{indable} A_{ccessible} I_{nteroperable} R_{eusable}



Implementation of FAIR principles

Rule #1—For every result, keep track of how it was produced

Rule #2—Avoid manual data manipulation steps

Rule #3—Archive the exact versions of all external programs used

Source: <https://dataconomy.com/2017/07/03/10-rules-results-data-science/>

Implementation of FAIR principles

Rule #4—Version control all custom scripts

Rule #5—Record all intermediate results, when possible in standardized formats

Rule #6—For analyses that include randomness, note underlying random seeds

Source: <https://dataconomy.com/2017/07/03/10-rules-results-data-science/>

Implementation of FAIR principles

Rule #7—Always store raw data behind plots

Rule #8—Generate hierarchical analysis output, allowing layers of increasing detail to be inspected

Rule #9—Connect textual statements to underlying results

Rule #10—Provide public access to scripts, runs, and results

Source: <https://dataconomy.com/2017/07/03/10-rules-results-data-science/>

Exam rules

The exam consists of a **choice between an individual or group project**, with an oral presentation of the project and some open-ended questions on theoretical topics related to the project subject.

Individual or group project (maximum score 30L):

- The course instructors will present the available projects in the second part of the course.
- The presented projects will be characterized by analytical, design, and implementation/optimization objectives, depending on the specific project.
- Depending on the specific complexity of the project, the number of students in each group will be limited.
- There will be no intermediate deadlines for completing the project.
- Projects can be developed without time constraints and presented at any time of the year.
- Projects will be discussed in an oral presentation (via slides); the correct date for the presentation will be agreed upon with the student.

Exam rules

Preparation for the presentation and discussion of the project:

- The oral presentation will cover the development of the project, the technological alternatives in the literature, the problems encountered, the solutions developed, and the results obtained (presentation via slides).
- The questions will cover the development of the project, the results, and the related theoretical topics covered during the course.
- A demo will also be required.
- A user manual on technical issues, as well as the developed code and results, will be provided to the teaching staff no later than three days before the project presentation.

Exam evaluation criteria

Project discussion evaluation

The evaluation is composed of the following:

- **24 points** for the design, implementation, documentation, and presentation of the **project**
- **2 optional points** for the laude if the project presents additional steps/analysis/objectives compared to those required in the specifications. Obviously, the additional contribution must be related to the assigned project and significant.
- **6 points** for open-ended questions on theoretical topics seen during the course and related to the project topic.

SLURM scheduler



Environment setup



Python basics for Bioinformatics



21/09/23

Questions?

*Better a stupid question in
class than a stupid answer
in the exam*