EFELIA
ÉCOLE FRANÇAISE DE
L'INTELLIGENCE ARTIFICIELLE
CÔTE D'AZUR

Coordonné par
UNIVERSITÉ
CÔTE D'AZUR

FRANCE 2030
Initiative d'Excellence

AI-Driven Approaches to Complex Biological structures

Structure : EUR LIFE	**Code de l'UE :** SMUSVAI	**Lieu d'enseignement :** Valrose
Niveau du cours : M1, M2	**Semestre :** Impair	**Langue :** Anglais

Professors :

- Salvish GOOMANEE (Salvish.GOOMANEE@univ-cotedazur.fr)
- Océane FIANT (Oceane.FIANT@univ-cotedazur.fr)

Description :

This course introduces students to the programming and conceptual tools required to apply machine learning (ML) to complex biological structures. A brief refresher on methods for ML in biology is first provided before diving into Python foundations for scientific computing and biological data handling. Students progressively build toward implementing ML approaches for problems such as biomarker discovery, drug target prediction, epigenetic pattern analysis, multi-omics integration and more. The course emphasises hands-on practice with real and synthetic biological datasets, computational modelling of biological networks and structures, and critical reflection on the ethical use of AI in biomedicine. By the end, participants will be equipped to preprocess data, apply ML workflows, and interpret AI-driven results in biological contexts.

Upon successful completion of this course, students will be awarded **6 ECTS**.

This course was co-developed by the EUR LIFE and the EFELIA team (École Française de l'Intelligence Artificielle).

Public :

Students in the first and second year of one of the following Masters : international BSCA, bio-information and computational biology.

Prerequisites :

- **General Python programming**
  - **Basic syntax & variables** (integers, floats, strings, booleans).
  - **Control flow** (if/else, for, while, break, continue).
  - **Functions** (definitions, arguments, return values, scope).
  - **Data structures** (lists, dictionnaires, tuples, etc…).
  - **Input/Output** (read/write to .csv/.npy/.pt files).
- **Machine learning tools with Python**
  - **NumPy** (Arrays vs lists, basic lin. alg., etc…)
  - **Pandas** (DataFrames, loading datasets, basic operations).
  - **Plotting tools** (Matplotlib/ Seaborn) - will be reintroduced as we go along.

Learning outcomes :

By completing this module, students will be able to :

- Use core libraries **(**NumPy**,** Pandas**)** to handle numerical and tabular biological data.
- Apply basic data cleaning and preprocessing techniques.
- Generate informative visualizations of biological datasets (e.g., expression heatmaps, structural plots).
- Understand how biological structures (sequences, networks, images) can be represented in Python for ML applications (based on PyTorch*, PyTorch Geometric*).
- Understand the applications of specific ML/DL architectures for dealing with complex biological data and systems.
- Design a simple geometric deep learning model of choice (CNNs, GNNs, LLMs, etc.…)_ _for a given task by the end of the course.
- Discuss the ethical use of AI in biomedicine and overall biomedical applications.

Teaching hours :

- **CM** (12 hrs.)
- **TD/TP** (28 hrs.)

Programme :

Session	Date	Duration	Classroom	Topics
	9/09 Christophe Becavin			Computer Programming Refresher (optional).
	16/09 Christophe Becavin			Computer Programming Refresher (optional).
1	23/09 Océane Fiant	3hr 14h-17h	Online	• Introduction to AI for Biology • History and Concepts of AI: - Origins - Symbolic approaches - Connectionist approaches
2	30/09 Océane Fiant	3hr 14h-17h	PV302 du petit valrose	• Challenges in AI: - Bias - Opacity - Generalizability • Generative AI: - Overview - Use cases in biology - Key issues (bias, « hallucinations », etc.)
3	7/10 Océane Fiant	4h 13h-17h	Salle C22	• AI in Biology: - Big Data and biology - Precision medicine - Case studies • Assesment
4	14/10 Salvish Goomanee	2hr	Salle C22	• Basics of python programming for biology. - Review of tools for data cleaning and manipulation. - Introduce examples with NumPy/Pandas. Generation of informative plots from real or synthetic data.

5	21/10 Salvish Goomanee	2hr	PV302 du petit valrose	• Towards ML for biological systems. - Understanding computational irreducibility and the need for ML in biology. - Review of some ML methodologies for biology (SL, UL, RL, etc.…)_. - Introduction to some classical DL architectures. (CNNs, GANs and/or GNNs).
6	28/10 Salvish Goomanee	2hr		• Dealing with the complexity of big data in biology. - Dealing with complex data structures and how to tame them (dimension reduction methods, …)_._ _ - AI response to omics-scale complexity. - Applications across life sciences & predictive analysis. - FAIR data principles & challenges.
7	4/11 Salvish Goomanee	2hr		• Classical ML/DL architectures for multi-omics integration and image analysis. - Understanding multi-omics analysis for ML. - RF for gene-expressions matrices (e.g.: used for biomarker discovery or disease classifications) - SVMs for ingestion of high dim datasets (e.g.: for predicting cancer subtypes from transcriptomics). - CNNs with image data: histopathology, cell microscopy, MRI/CT scans. -

8	18/11 Salvish Goomanee	2hr		• Deep learning (DL) architectures for complex biological systems. - When classical ML fails: dealing with more complex biological structures using GNNs (e.g.: structural biology: predicting protein contact maps – _CNNs v/s GNNs) - GNNs/GANs for drug docking/molecular property prediction and/or gene regulatory networks. - Towards LLMs in gene and epigenetic analysis*.
9	25/11 Salvish Goomanee	2hr		• Ethical use of AI in biomedical engineering. - Epistemic shortcomings of AI-Derived evidence. - Transparency, Fairness, and Algorithmic Bias. - Accountability and Traceability.

Assessment :

Theory questions (50%) + short project (50%)

Bibliography (books) :

- **ML for biological systems**

[1]. Ghosh, S., & Dasgupta, R. (2022). *Machine Learning in Biological Sciences: Updates and Future Prospects*. Springer Singapore. https://doi.org/10.1007/978-981-16-8881-2

[2]. Moses, A. M. (2017). *Statistical Modeling and Machine Learning for Molecular Biology* (1st ed.). Chapman & Hall/CRC.

[3]. Bassi, S. (2017). *Python for Bioinformatics* (2nd ed.). Chapman & Hall/CRC. https://doi.org/10.1201/9781315268743

- **Big data and omics**

[1]. Xiong, M. (2018). *Big Data in Omics and Imaging: Integrated Analysis and Causal Inference* (Chapman & Hall/CRC Computational Biology Series; xxix, 736 pp.). CRC Press.

[2]. Yona, G. (2011). *Introduction to Computational Proteomics*. Chapman & Hall/CRC.

- **Computational genomics and sequencing**

[1]. Akalin, A. (2020). *Computational Genomics with R*. Retrieved from https://compgenomr.github.io/book/

[2]. Korpelainen, E., Tuimala, J., Somervuo, P., Huss, M., & Wong, G. (2014). *RNA-seq Data Analysis: A Practical Approach* (1st ed.). Chapman & Hall/CRC.

[3]. Ismail, H. D. (2023). *Bioinformatics: A Practical Guide to Next Generation Sequencing Data Analysis* (1st ed.). Chapman & Hall/CRC.

- **Systems biology** *(with a focus on graph-based representations of networks!)*

[1]. Raman, K. (2023). *An Introduction to Computational Systems Biology: Systems-Level Modelling of Cellular Networks* (1st ed.). Chapman & Hall/CRC.

- Ethics for use of AI in biomedicine

[1].

Bibliography (articles) :

[1]. Wolfram, S. (1985). Undecidability and intractability in theoretical physics. *Physical Review Letters, 54(8), 735–738.* https://doi.org/10.1103/PhysRevLett.54.735

[2]. Lawrence, E., El-Shazly, A., Seal, S., Joshi, C. K., Liò, P., Singh, S., Bender, A., Sormanni, P., & Greenig, M. (2024). Understanding biology in the age of artificial intelligence. *arXiv*. https://doi.org/10.48550/arXiv.2403.04106

[3]. Sapoval, N., Aghazadeh, A., Nute, M. G., Antunes, D. A., Balaji, A., Baraniuk, R., Barberán, C. J., Dannenfelser, R., Dun, C., Edrisi, M., Elworth, R. A. L., Kille, B., Kyrillidis, A., Nakhleh, L., Wolfe, C. R., Yan, Z., Yao, V., & Treangen, T. J. (2022). Current progress and open challenges for applying deep learning across the biosciences. *Nature Communications, 13(1), 1728.* https://doi.org/10.1038/s41467-022-29268-7

[4]. Goshisht, M. K. (2024). Machine learning and deep learning in synthetic biology: Key architectures, applications, and challenges. *ACS Omega, 9(9), 9921–9945.* https://doi.org/10.1021/acsomega.3c05913

Resources for success :

- TUT'TOP : peer tutoring on methodological, social, administrative or logistical issues.
- écri+ : to improve your written French.
- Centre de ressources en langues : to improve your foreign language skills (French or other).
- METODA : to improve your documentary research skills.
- S'orienter / Se réorienter : to be advised by the university's career counsellors.
- Centre de santé et aide sociale : to look after your physical and mental health, and to seek support in the event of social hardship.
- Cellule Handicap : support for students with disabilities.
- Plateforme de signalement : to report acts of violence, harassment or discrimination (sexual and gender-based violence, LGBTphobia, racism, xenophobia, etc.) you have witnessed or experienced at the university, and to get support.

Important : This syllabus has no binding value. Its content may change during the course of the year.

Ce travail a bénéficié d'une aide de l'Etat gérée par l'Agence Nationale de la Recherche (ANR) au titre de France 2030 pour le projet EFELIA Côte d’Azur portant la référence ANR-22-CMAS-0004.