Deep Learning for scaling up microalgae cultivation from the lab to industrial scale

Laboratoire Méthodes Formelles (LMF), ENS Paris-Saclay

November 21, 2025

M2 internship description

Context Microalgae are unicellular photosynthetic organisms capable of converting ambient carbon dioxide into chemical energy via light-driven photosynthesis. They produce valuable compounds for both academic research and industrial applications. Promising uses include lipid production for biofuels in aerospace, wastewater purification through ammonium absorption, and plant-based protein for food applications.
Scaling up microalgal biomass for industrial purposes requires large culture volumes. However, as volume increases, challenges arise that create a gap between laboratory-scale observations and performance in photobioreactors or open ponds exceeding 1,000 L. Key issues include light heterogeneity caused by self-shading, which reduces photosynthetic efficiency, temperature gradients and overheating risks, limitations in gas and nutrient mass transfer, and the risk of culture contamination.

Proposed Approach Mechanistic modeling based on kinetic equations or mass balances can describe certain aspects of microalgal cultures [1], but it quickly reaches its limits due to the complexity of biological processes and heterogeneities at the process scale. Data-driven approaches, whether using machine learning methods [2] or classical statistical models, offer a promising alternative.
In particular, recurrent neural networks such as Long Short-Term Memory (LSTM) networks have been shown to effectively capture temporal dynamics in biological growth processes, including microalgae, by learning patterns from time-series data [3]. In the context of upscaling, where culture conditions become more variable and difficult to capture with deterministic equations, these approaches therefore hold significant potential. Initial tests have been performed to scale up microalgae cultures from 96-well plates (200 µL) to flasks (50 mL). Additional data at larger volumes have also been generated, and the complete dataset is described below. The current strategy for predicting microalgae growth relies on a two-phase LSTM-based framework:

- *Phase 1 – Pre-training (Autoencoder):* Microplate growth data with contextual parameters (e.g., initial light intensity L_0, inoculum concentration C_0) are used as input. A two-layer LSTM encoder with 16 units and dropout extracts meaningful temporal features, and the network is trained to reconstruct the input growth curves in a self-supervised manner.
- *Phase 2 – Final Model (Transfer Learning):* The pre-trained encoder is frozen, and the contextual parameters are repeated along the time sequence and concatenated with the encoded features. A decoder composed of dense layers then predicts growth curves at the Erlenmeyer scale.

Internship Objectives: The internship will focus on improving this framework by:

- Enhancing the incorporation of contextual parameters (e.g., light distribution, temperature, pH) for better predictive accuracy.
- Exploring advanced LSTM architectures or attention-based models to capture complex temporal dependencies.
- Assessing model generalization to larger volumes and variable culture conditions.
- Providing interpretable predictions to support scaling strategies from laboratory to industrial scale.

This project offers hands-on experience in deep learning for bioprocess modeling and the application of transfer learning to biological time-series data.

Dataset The training dataset has been prepared by part of the supervising team and pre-processed for immediate use. It currently contains 2,223 microalgae growth curves measured over time across multiple culture volumes (200 µL, 50 mL, 1 L, 3 L, 5 L, and 285 L). For laboratory-scale volumes (200 µL, 50 mL), various nutrient conditions and light intensities were tested to provide diverse contexts. Previously developed mechanistic models of microalgae growth by the supervising team can also be leveraged for data augmentation if needed.

Place and duration

The internship will last 4 to 6 months in LMF laboratory at ENS Paris-Saclay at 4 Avenue des Sciences, Gif-sur-Yvette, 91190.

Education and required skills

- Master 2 in artificial intelligence, applied mathematics, computer science or electrical engineering
- Proficient in Python for AI applications, in particular experience with scikit-learn and PyTorch
- Interest in biological topics and experience in data analysis is highly appreciated

Application details

You can send your CV and motivation letter at the following contacts.

Contacts: Mélanie Pietri, PhD student melanie.pietri@ens-paris-saclay.fr
Thomas Nowak, Professor thomas.nowak@ens-paris-saclay.fr
Mathias Fugger, CNRS Researcher matthias.fugger@ens-paris-saclay.fr

Lab presentation

The Laboratoire Méthodes Formelles (LMF) is the computer science laboratory of the École normale supérieure Paris-Saclay, Université Paris-Saclay, and the French National Centre for Scientific Research (CNRS) as joint research unit UMR 9021. The internship will take place within the Interactions Biology group.

In this team we have been developing predictive population-level models for bacterial and microalgal systems, mathematically analyzing stochastic effects among competing cell populations, and developing a simulation tool for microbiological systems. Our group members and collaborators mainly come from AI, synthetic biology, distributed computing, and control theory. We were fortunate enough to be able to equip an experimental wet lab for our research group in the premises of ENS Paris-Saclay. Our research agenda includes the study of quantitative and predictive modeling of intra-cellular and cell-population dynamics, simulation techniques, automated model discovery of bioprocesses, automated design of plasmids for bacterial hosts, optimization of bioproduction, and the design of synthetic circuits towards applications in bioproduction and medical diagnosis.

Activities

Weekly team and individual meetings are organized to discuss progress and provide guidance. Monthly lab activities, such as informal social gatherings and movie sessions, also take place. Interns will have the opportunity to visit the biology laboratory where the dataset was generated.

References

[1] Eunyoung Lee, Mehregan Jalalizadeh, and Qiong Zhang. Growth kinetic models for microalgae cultivation: A review. *Algal Research*, 12:497–512, 2015. ISSN 2211-9264. doi: https://doi.org/10.1016/j.algal.2015.10.004. URL `https://www.sciencedirect.com/science/article/pii/S2211926415300783`.

[2] Tehreem Syed, Felix Krujatz, Yob Ihadjadene, Gunnar Mühlstädt, Homa Hamedi, Jonathan Mädler, and Leon Urbas. A review on machine learning approaches for microalgae cultivation systems. *Computers in Biology and Medicine*, 172:108248, 2024. ISSN 0010-4825. doi: https://doi.org/10.1016/j.compbiomed.2024.108248. URL `https://www.sciencedirect.com/science/article/pii/S0010482524003329`.

[3] Daniel Boiar, Nils Killich, Lukas Schulte, Victor Moreno, Jochen Deuse, and Thomas Liebig. *Forecasting Algae Growth in Photo-Bioreactors Using Attention LSTMs*, pages 26–37. 02 2023. ISBN 978-3-031-26235-7. doi: 10.1007/978-3-031-26236-4_3.