

Lamine TOURE

laminetoure626@gmail.com - linkedin.com/in/lamine-toure - github.com/LamineTourelab

EDUCATION

University Paris Est

Créteil, France

Master in Bioinformatics and Multi-Omics Engineering

September 2020 - June 2022

- **Main courses :** Multi-Omics data analysis, Data Mining, statistical modeling, machine learning, Biotechnology, disease understanding

University Paris Est

Créteil, France

Bachelor of Science

September 2018 - June 2020

- **Main courses :** Biostatistics (descriptive statistics, statistical testing), Molecular biology, microbiology, biotechnology, enzymology, metabolism

TECHNICAL SKILLS

Programming Languages: Python, R, Matlab(bases)

Libraries and Tools in Python : Sklearn, Tensorflow, Keras, PyTorch, Pandas, Numpy, OpenCV, Matplotlib, Seaborn, Plotly, Dash ...

Libraries and Tools in R: Caret, dplyr, ggplot2, edgeR, DESeq2, tidyverse, biomaRt, magrittr...

Tools : Snakemake, Jira, Git, Docker, SQL...

Web App : Shiny (R and Python), flexdashboard, MkDocs, Streamlit, Plotly Dash...

ML Architectures: NN, CNN

Bioinformatics and Biostatistics skills : Predictive modeling, Data mining, OMICS analysis, Early drug discovery, Statistical modeling, image analysis, explainable AI, biological network analysis

Operating System : Windows, Mac OS, Linux

WORK EXPERIENCE

Bioinformatics Engineer

Institut Necker Enfants Malades, Paris, France

October 2024 – present

- Single cell data analysis (data preparation, tSNE, UMAP, batch effect correction)
- Analysis of gene regulation networks (Bayesian methods, metaVIPER)
- RNA velocity and trajectory analysis (scVelo, Cellrank)
- Biologically informed deep learning for single-cell reference mapping (ExpiMap)
- Shiny App for genomics data (differential expression visualization, Data visualization, Statistical analysis ...).
- ...

Data Scientist

OncoDesign Precision Medicine, Dijon, France

September 2022 - June 2023

Goal : Implement machine learning methods to discover new therapeutic targets

- Integration and analysis of multi-omics data using GCN, AE, factorial methods for (early, intermediate and late) integration methods
- Variable selection (LASSO, Group LASSO)
- Machine learning (xgboost, lightGBM, Random Forests, SVM)
- Longitudinal data analysis with clinical data at different time-point
- Deep learning (Autoencoder, neural networks, Graph Convolution Network) for variables selection.
- Setting up bioinformatics pipelines for interpreting NGS data with Snakemake (scalability and reproducibility)
- Post-hoc explicability methods (SHAP and LIME) or integrated into models such as the attention mechanism, gradient-based methods (e.g: GradCAM) ..
- Medical Images preprocessing and analysis (Segmentation, classification)
- Communities Detection in graphs using igraph, networkX, node2vec.
- Shiny App for differential expression visualization.

Data Scientist intern

OncoDesign Precision Medicine, Dijon, France

March 2022 - August 2022

Subject: Discovery of new therapeutic targets using multi-omics data

- Variable selection methods(LASSO, Group LASSO, Overlap Group LASSO)
- Performed two types of variables grouping used in Group LASSO: "statistic" with the group construction based on statistical method and on "a priori biological knowledge" grouping approach based on pathways
- Co-expressed gene network analysis
- Statistical analysis (Descriptive, clustering, outliers detection)
- Post-hoc explicability methods (SHAP and LIME)

Bioinformatics intern

ICM – Genomic and Sequencing Platform, Paris, France

April 2021 - August 2021

- Diagnosis of glioblastoma using NGS data analysis
- Data quality control
- Responsible for the SARS-CoV-2 survey of ICM staff
- Paraneoplastic disease: immunohistochemical analysis

PROJECTS

- **Interpretable AI review** : Explain machine learning algorithms applied in healthcare in collaboration with [Salvatore RAIELI](#). The source code can be found at <https://github.com/SalvatoreRa/explainableAI>
- **Tutorial repository** : The goal here is to build a huge tutorial repository in the field of AI and bioinformatics. The source code can be found at <https://github.com/LamineTourelab/Tutorial>
- **Docs Deployment** : The goal here is to make resources available for every one by deploying all the documentations. Take a look on my website for details <https://laminetourelab.github.io/>
- **School project: Multi-omics Research Project**
 - Goal: Link between gut microbiota and the development of Alzheimer's disease in transgenic mice with multi-omics approaches
 - Integration of lipidomic and proteomic data from different years (over 5 years)
 - Data analysis, preprocessing and integration
 - Find common marker that explain the link between the microbiota and the development of Alzheimer's disease
 - Report in scientific article format

EXTRACURRICULAR ACTIVITIES

- **Student Coordinator** during the master degree - September 2020 - June 2022
- **Other Interests** Football, Traveling, Cinematography, reading, volunteering. -