

MAHAN BALOOEI

Modena, Italy · mahan.balooei@studio.unibo.it · LinkedIn · GitHub

PROFILE

MSc Bioinformatics student at the University of Bologna with a background in Medical Laboratory Sciences and a focus on computational protein biology, protein language models, and machine-learning-based protein sequence analysis. Experienced with ESM-2 representations, sequence-function prediction, HMMER/MMseqs2 workflows, and reproducible machine-learning pipelines. Interested in developing learning-based methods for protein sequence, structure, function, and molecular interactions.

RESEARCH INTERESTS

- Protein language models and representation learning
- Protein sequence-structure-function modelling
- Protein-protein and protein-peptide interaction modelling
- Generative and structure-aware protein design
- Remote homology detection and functional annotation

EDUCATION

University of Bologna — MSc in Bioinformatics

Oct 2024 – Present

Coursework: sequence alignment, RNA-seq, microarrays, statistical analysis in R; ML/DL with scikit-learn & TensorFlow; SQL/graph queries, Docker, HPC/Slurm.

Iran University of Medical Sciences — BSc in Medical Laboratory Sciences

Sept 2019 – Jul 2023

Relevant Coursework: Molecular Diagnostics, Immunoassays, Clinical Chemistry, Haematology, Microbiology, Virology.

PUBLICATIONS

Mechanism-based Targeting of Lactate Dehydrogenase Tetramerization Process through Natural Derived Compounds: A Survey of Molecular Docking, Molecular Dynamics, and QSAR Analysis

Sep 2025

Computers in Biology and Medicine, 2025

Equal-contributing author

- Screened approximately 5,000 naturally derived compounds against four LDH tetramerization interfaces and evaluated top candidates using molecular docking, 100-ns molecular dynamics simulations, MM-PBSA, PCA/DCCM, pharmacophore, ADMET, and QSAR analyses.
- Identified candidate inhibitors with distinct effects on LDH conformational dynamics, highlighting compounds 3211 and 4436 as promising modulators of the A2 and B2 tetramerization interfaces, respectively.
- [Article link](#)

RESEARCH PROJECTS & COMPUTATIONAL EXPERIENCE

Protein Language Model-Based Signal Peptide Prediction

Sep 2025 – Nov 2025

MSc Bioinformatics Project - Independently Implemented

- Built an end-to-end protein sequence modelling pipeline for eukaryotic signal-peptide prediction using high-confidence UniProtKB/Swiss-Prot annotations.
- Applied MMseqs2 clustering for redundancy reduction and constructed biologically meaningful train/benchmark splits.
- Compared von Heijne PSWM, biochemical-feature SVM, CNN-BiLSTM, and ESM-2 embedding-based models.

- Achieved $F1 = 0.962$ and $MCC = 0.957$ on a held-out benchmark set using the ESM-2-based model.
- Analyzed amino-acid composition, signal-peptide lengths, kingdom-specific bias, and cleavage-site motifs for biological interpretation.
- [GitHub Repository](#)

Protein Domain and Functional Annotation Using HMMER and MMseqs2

Jun 2025 – Jul 2025

- Built an automated protein domain annotation workflow using profile HMMs, HMMER, and MMseqs2 for large-scale sequence classification.
- Processed protein structure-derived alignments to support domain-level annotation and sequence-based functional inference.
- Compared profile-based and sequence-similarity-based classification workflows, achieving $MCC = 0.997$ and $MCC = 0.991$, respectively.
- Connected classical homology and profile-HMM approaches to modern protein representation learning tasks such as remote homology detection and protein family classification.
- [GitHub Repository](#)

Computational Epigenomics Analysis of DNA Methylation (Illumina 450K)

Jun 2025 – Jul 2025

- Conducted an R/Bioconductor workflow for Illumina 450K DNA methylation array analysis.
- Performed quality control, normalization, PCA-based batch-effect assessment, and differential methylation analysis.
- Identified Differentially Methylated Regions (DMRs) and visualized methylation patterns using heatmaps, volcano plots, and exploratory statistical summaries.
- [GitHub Repository](#)

Chest X-Ray Pneumonia Detection via Transfer Learning

Mar 2026 – Apr 2026

- Compared ResNet50, EfficientNetB0, and DenseNet-121 XRV for pneumonia classification.
- Addressed class imbalance and evaluated threshold-selection strategies using clinically relevant metrics.
- Applied ensembling, test-time augmentation, and Grad-CAM-based interpretability.
- DenseNet-121 XRV achieved the strongest performance under the maximum-MCC threshold strategy.
- [GitHub Repository](#)

SKILLS

Protein Bioinformatics & Sequence Analysis: ESM-2 protein language model embeddings, protein representation learning, signal peptide prediction, sequence-function prediction, protein-domain annotation, HMMER, MMseqs2, BLAST, Clustal, UniProt, PDB.

Machine Learning & Deep Learning: PyTorch, scikit-learn, TensorFlow, SVMs, CNN-BiLSTM, transfer learning, representation learning, model evaluation, threshold optimization, class-imbalance handling, ensembling, and Grad-CAM-based interpretability.

Functional Genomics & Epigenomics: DNA methylation analysis (Illumina 450K arrays, minfi), differential methylation region (DMR) detection, PCA-based batch-effect assessment, R/Bioconductor.

Reproducible Workflows & Tools: Python, R, Bash/Linux, SQL, Git/GitHub, Docker, HPC/Slurm, R Markdown, Jupyter Notebooks, conda environments.

WORK EXPERIENCE

Clinical Laboratory Internships

Tehran, 2022–2023

- Rotations in haematology, microbiology, clinical chemistry, and biochemistry across hospital and diagnostic laboratory settings.

LANGUAGES

English: Proficient (TOEFL iBT 93, CEFR B2)

Italian: Basic communication

Persian (Farsi): Native