

Emil Cacayan

Bolingbrook, IL

Website: <https://www.emilcacayan.com>

LinkedIn: <https://www.linkedin.com/in/emil-andre-cacayan-797741129/>

GitHub: <https://github.com/cacayan2>

SUMMARY

Bioinformatics professional with experience developing computational workflows, machine learning pipelines, and analytical tools for biological and healthcare datasets. Skilled in Python, R, statistical modeling, and scientific software development with interests in computational genomics, transcriptomics, machine learning, and reproducible data engineering workflows.

EDUCATION

Loyola University Chicago

M.S. Bioinformatics

Expected Graduation: August 2026 | GPA: 3.97

Relevant Coursework: Machine Learning • Quantitative Bioinformatics • Statistical Modeling • Experimental Design • Computational Genomics

University of Illinois Urbana-Champaign

B.S Biochemistry

Graduation Date: August 2023

Relevant Coursework: Genomics and Human Health • Applied Statistics • Data Structures and Algorithms • Molecular Biology

TECHNICAL SKILLS

Programming: Python, R, Bash, SQL, Java, C++, TypeScript

Bioinformatics & Data Science: RNA-seq, Transcriptomics, GWAS/QTL Analysis, Statistical Genetics, Single-Cell Analysis, Machine Learning, Feature Engineering

Tools & Technologies: PLINK, pandas, NumPy, scikit-learn, Git/GitHub, Linux/Unix, Jupyter, Next.js, React, Vercel, PyMOL, NCBI Suite, Biopython, Seurat

Data & Systems: Workflow Automation, Reproducible Pipelines, ETL/Data Processing, Scientific Computing, Data Visualization

PROJECTS

Machine Learning Classification Pipeline (Python, scikit-learn)

- **GitHub:** <https://github.com/cacayan2/machinelearning-classification-pipeline>
- Designed end-to-end machine learning pipeline including **data preprocessing, feature engineering, model training, evaluation, and deployment**
- Implemented multiple classification models and validated performance using cross-validation

Fetal Health Classification (R)

- **GitHub:** <https://github.com/cacayan2/fetal-health-classification-r>
- Developed predictive models (**multinomial regression, k-NN, random forest**) to classify fetal health outcomes from cardiotocography data
- Performed **model selection (AIC/BIC)** and cross-validation to evaluate model robustness (>95% accuracy achieved)
- Interpreted model outputs to identify clinically relevant predictors and patterns in healthcare data
- Generated visualizations and summaries to communicate results clearly
- Interpreted quantitative findings and communicated results in a clinically relevant context

Genomic Colocalization Pipeline (R, Python, PLINK)

- **GitHub:** <https://github.com/cacayan2/genomic-colocalization-pipeline>
- Built automated pipeline integrating **GWAS** and **QTL datasets** for colocalization analysis
- Implemented **SNP harmonization, allele matching, and LD-based correlation matrix generation** using 1000 Genomes data
- Developed reproducible workflows for high-dimensional genomic data processing and analysis
- **Validated outputs** against expected statistical and biological behavior to **ensure correctness**.

COVID-19 Severity Clustering (R)

- **GitHub:** <https://github.com/cacayan2/covid-19-severity-clustering>
- Applied **PCA** and **K-means clustering** to **high-dimensional epidemiological data**

- Performed **dimensionality reduction** and evaluated clustering using **silhouette and elbow methods**
- Interpreted patterns in **global disease burden** and **outcome variability**

ALDH2 Structural Analysis Project (PyMOL, Structural Biology)

- **GitHub:** <https://github.com/cacayan2/aldh2-structural-analysis>
- Conducted **structural analysis** of ALDH2 wild-type and mutant protein conformations using **PyMOL** and **Protein Data Bank** datasets.
- Investigated **residue-level structural differences** associated with **altered enzymatic activity** and NAD⁺ binding.
- Applied **computational visualization techniques** to interpret biologically relevant structural changes.
- Presented a poster in an **academic research symposium setting**.

RNA-seq Transcriptomics Pipeline (Python, R, Bash)

- **GitHub:** <https://github.com/cacayan2/rna-seq-transcriptomics-pipeline>
- Developed reproducible transcriptomics workflows for **RNA-seq preprocessing, alignment organization, quantification, and downstream analysis**.
- Integrated commonly used bioinformatics tools and **automated handling of sequencing datasets and analysis outputs**.
- Structured **modular workflows** for scalable biological data processing and interpretation.

CLINICAL/PROFESSIONAL EXPERIENCE

Computational Genomics Research Intern

Loyola University Chicago, Stuart Lab

May 2026 – Present

- Contribute to **computational genomics** and **bioinformatics workflow** development supporting mitochondrial genome and comparative genomics research.
- Assist with **biological sequencing data organization, automation, genomic preprocessing, and scalable data handling pipelines**.
- Collaborate on **development of reproducible computational workflows** for **large-scale biological dataset processing and analysis**.
- Support **scientific infrastructure development** and **computational tooling** for **genomics-focused research applications**.

Emergency Department Technician

Advocate Good Samaritan Hospital – Downers Grove, IL

July 2025 – Present

- Utilize **Epic EHR** for documentation, order tracking, and clinical workflow documentation
- Assist physicians and nurses in patient care, including **phlebotomy, monitoring, and specimen collection**
- Communicate patient status and clinical observations to care team in a high-acuity environment
- Coordinate workflow across providers, supporting efficient patient throughput and care delivery
- Interacted with structured and unstructured clinical data within EHR systems

Emergency Medical Technician (EMT)

Superior Air-Ground Ambulance – Aurora, IL

December 2023 – August 2025

- Delivered prehospital care following clinical protocols, supporting patient stabilization and transport
- Communicated with hospital staff to ensure continuity of care and accurate information transfer
- Gained exposure to **real-world variability in patient presentation and outcomes**

LEADERSHIP, INVOLVEMENT, & CERTIFICATIONS

Graduate Student Advisory Council (GSAC)

Loyola University Chicago

August 2024 – Present

- Represent graduate student interests and contribute to program-level discussions and initiatives
- Collaborate with faculty and administration to improve academic and student experience
- Participate in planning and coordination of student-focused events and communications

Crisis Text Line Volunteer

February 2025 – Present

- Provided real-time crisis intervention using structured communication and de-escalation techniques
- Assessed risk and guided individuals toward actionable next steps