Merlin Dassanayake
Boston, MA | dassanayake.m@northeastern.edu | LinkedIn | GitHub

EDUCATION

Northeastern University, Boston, MA — January 2024 – Expected June 2026
Candidate for the Professional Master of Science in Bioinformatics — GPA: 3.96
Coursework: Introduction to Data Mining/Machine Learning, Bioinformatics Programming, Statistics for Bioinformatics, Genomics, and Information Design and Visual Analytics

University of California, Santa Barbara, Santa Barbara, CA — September 2017 – June 2021
Bachelor of Science, Biopsychology — GPA: 3.5
Coursework: Molecular Genetics, Principles of Biochemistry, Neuropharmacology, Organic Chemistry, Linear Algebra, Vector Calculus, Research Methods, Statistics

TECHNICAL SKILLS

Programming & Scripting: R (tidyverse, DESeq2), Python (pandas, numpy, scanpy, anndata), Bash, SQL
Bioinformatics Tools & Databases: TCGAbiolinks**,** BLAST, UCSC Genome Browser, FASTQC, IGV
Data Visualization: ggplot2 , Tableau Public, Matplotlib, Power BI
Environment & Workflow Management: Conda/mamba, Git/Github, Linux/Unix shell, Docker

ACADEMIC PROJECTS

TCGA Breast Cancer Transcriptomic Analysis (RNA-seq Differential Expression and Enrichment),
Personal Project — September 2025 – October 2025

- Processed and normalized 1,200+ TCGA-BRCA RNA-seq samples using TCGAbiolinks and DESeq2, identifying ~5,000 significant DEGs (FDR < 0.05) across tumor vs. normal tissues
- Integrated clinical survival and expression data to perform Kaplan-Meier and Cox regression analyses, identifying one gene with significant prognostic impact ($p < 0.05$)
- Automated workflow with modular R scripts and bash pipeline, generating reproducible volcano plots, enrichment maps, and survival curves

Soybean Grain Yield Prediction using Machine Learning in R, Northeastern University
Introduction to Data Mining/Machine Learning Signature Project — July 2024 - August 2024

- Applied data cleaning and PCA techniques in R on agricultural dataset handling missing values and reducing collinearity, improving RMSE and maintaining 80% variance explained
- Utilized three different machine learning models (Multiple Regression, Regression Tree, and SVM) in R to predict grain yield, improving accuracy by 47% when using k-fold cross validated homogeneous bagging model compared to SVM (RMSE of 571 vs. 297)

RESEARCH EXPERIENCE

Vision and Image Understanding Lab, Santa Barbara, CA — October 2019 – June 2021
Undergraduate Research Assistant

- Streamlined experiment enrollment while serving on executive team with two other research assistants by designing recruitment posters and developing standardized Google Sheets to track participant progress and compensation, improving efficiency of experiment deployment time
- Participated in initial data testing for cancer detection study using 3D breast tomography scans ensuring reliability of experimental program on MATLAB, allowing experiment to proceed in enrolling test subjects and eventual publication