Youssef Gehad

857-880-6875 | yag1@rice.edu | linkedin.com/in/youssef-gehad | github.com/yag1Rice

Education

Rice University Aug. 2022 - May 2026

Bachelor of Arts in Computer Science, Bachelor of Arts in Statistics, & Minor in Data Science (GPA: 3.96/4.0) *Houston, TX*

- Computer Science Coursework: Computational Thinking, Algorithmic Thinking, Introduction to Program Design, Reasoning About Algorithms, Introduction to Computer Systems, Tools and Models for Data Science, Compiler Construction, Practical Machine Learning, Introduction to Computer Vision, Deep Learning for Vision and Language, Concurrent Program Design, Reinforcement Learning, Deep Learning Systems*, Software Engineering Methodology*
- Statistics Coursework: Probability & Statistics, Linear Regression, R for Data Science, Quantitative Financial Analytics, Computational Finance, Applied Time Series/Forecasting, Crypto & Blockchain, Probability, Applied Machine Learning and Data Science*

*To be taken in Spring 2026

Research Interests

Computer Vision, Multimodal Learning, Trustworthy AI, and AI for Healthcare

Research Experience

Balakrishnan Laboratory - Rice University Sept. 2025 - May 2026

Research Assistant | PyTorch & Python *Houston, TX*

- Curated a new cross-modal dataset using MIMIC-IV for echocardiogram videos, patient EHR, & radiology and discharge notes to evaluate and pinpoint the weakness of current foundation models
- Building a multimodal CLIP-based model using echocardiogram videos with extracted ECG signals to improve generalizability and downstream task performance (e.g., left-ventricle ejection-fraction estimation) while providing a richer feature representation that addresses limitations in current SOTA models

Patel Laboratory - Harvard University May 2025 - May 2026

Research Intern | R, Bash, & Python *Boston, MA*

- Processed, cleaned, and visualized high-dimensional biological Omics datasets with over 100,000 entries from the UK Biobank, creating a reusable pipeline for data analysis
- Developed and validated over 300 statistical and machine learning models, including Logistic Regression with ElasticNet regularization, SVM, Random Forests, & Decision Trees, to evaluate predictive performance for cardiovascular disease using metabolomic & proteomics data
- Co-author on a paper submitted to *Nature Medicine* (Under Review)
- First author on a manuscript in preparation, expected submission in 2026

Nirmal Laboratory - Harvard University May 2024 - Sept. 2024

Research Intern | Docker Hub, Slurm, Podman, Bash, Python, & PyTorch *Boston, MA*

- Fine-tuned foundation model vision transformers and applied pretrained vision transformers to extract patch-level embeddings from H&E tissue images, gaining insights into cellular regions within the tissue
- Constructed a multilayer perceptron that employed the extracted embeddings for training and adjusted the loss function to combine related classes, leading to a notable increase in overall model performance metrics
- Co-author on a manuscript in preparation, expected submission in 2026

Kockara Laboratory - Rice University Sept. 2023 - Dec. 2024

Research Assistant | PyTorch & Python *Houston, TX*

- Built a Class Activation Mapping to visualize attention region heat maps for several CNN architectures (ResNet, MobileNet, VGG16, Inception V3), allowing for insights into image misclassifications and the saliency regions of the model
- Utilized CNN-based AI models for semantic segmentation of Dermoscopy images, accurately extracting hair follicle outlines and successfully segmenting over 1,000 images
- Demonstrated that obstructions on skin lesions reduce classification model F1 score by 45%, highlighting the need for better obstruction-removal techniques

Projects

OwlDB | *Go* Aug. 2025 - Oct. 2025

- Collaborated in a team of 3 to develop a robust, scalable document-based database system
- Developed a custom NoSQL backend database supporting a simple authentication flow, resource modification and retrieval, and subscriptions to resource updates through Server-Sent Events, all via a RESTful web API
- Leveraged Skip Lists to ensure concurrency and dependency injections to adhere to a modular and decoupled software design

Multimodal Leaf Health Analysis | *Python, PyTorch, IBM WatsonX, LangChain, & IBM Cloud* Feb. 2025 - Apr. 2025

- Fine-tuned two pretrained CNNs (Densenet-121) on 50,000 images collected from the PlantVillage Dataset
- Implemented SmoothedGradCAM to visualize the regions of highest importance used by the models when either classifying plant types or identifying specific diseases
- Leveraged IBM's cloud infrastructure to remotely run IBM's Granite-3-2 large language model, integrating it with a Retrieval-Augmented Generation (RAG) architecture using LangChain to minimize hallucinations and provide relevant context

PageRank | *SQL* Feb. 2024

- Implemented Google's PageRank algorithm using stored procedures, analyzing over 14,000+ citation relationships across 5,000+ papers
- Created an efficient SQL stored procedure to interpret the data as an undirected graph, identifying connected components and displaying related paper titles

Technical Skills

Programming Languages: Python, R, LaTeX, Bash, Java, SQL, Go, TypeScript, HTML, CSS, & C **Frameworks**: PyTorch, Cuda, Slurm, PySpark, Podman, Docker Hub, Cypress, & Hadoop **Developer Tools**: Git, GitLab, GitHub, VS Code, Excel, & RStudio

Publications

Assessing Type 2 Diabetes and GLP-1 agonist response trajectories with a Proteogenomic Atlas of disease pro- gression. Sivateja Tangirala, Shakson Isaac, **Youssef Gehad**, Filomene Roquefort, Pauline Gabrieli, Gary W Miller, Vidhu Thaker, Braden T Tierney, Chirag J Patel. *Nature Medicine* (Under Review).

Harnessing Histopathology Foundation Models for Single-Cell Type Inference from Hematoxylin and Eosin Stain Images. Tyler Jost, **Youssef Gehad**, Peter Sorger, Ajit Nirmal. In preparation: expected submission 2026.

Metabite: Advancing Data-Driven Prediction of Cardiovascular Disease from Metabolomic Profiles. **Youssef Gehad**, Braden T Tierney, Sivateja Tangirala, Chirag J Patel. In preparation: expected submission 2026.

Conferences & Presentations

Metabite: Advancing Data-Driven Prediction of Cardiovascular Disease from Metabolomic Profiles Oct. 2025

- Poster Presentation #12: *Innovation for Health Care Access Conference: from Global to Local,* (Houston, TX). Presenter: Youssef Gehad

Advancing Data-Driven Prediction of Cardiovascular Disease from Metabolomic Profiles Aug. 2025

- Department of Biomedical Informatics, Harvard University (Boston, MA). Presenter: Youssef Gehad

Predicting Cell phenotypes from Hematoxylin and Eosin images via multiplexed image data Aug. 2024

- Department of Dermatology, Harvard Medical School (Boston, MA). Presenter: Youssef Gehad

Awards

President's Honor Roll (Rice University. Institutional. Academic.)

- Received 3 times (Spring 2024, Fall 2024, Spring 2025)

Volunteer & Teaching Experience

Volunteer Peer Academic Fellow Sept. 2023 - May 2025

- Led weekly help sessions in introductory programming courses at Rice University

Volunteer Math Tutor Sept. 2021 - June 2022

- Mentored elementary school students and hosted weekly lessons in mathematics