Dona Hasini Vathsalya Gammune

Milwaukee, Wisconsin | hasinigammune@gmail.com | +1 (469) 4035895 | LinkedIn | GitHub

Summary

Ph.D. graduate in Data Science and Statistics with expertise in high-dimensional statistics, machine learning and deep learning, and computational biology. Proficient in developing innovative statistical, machine learning, and deep learning models to analyze high-dimensional omics data, novel computational tools, and analytical methods for understanding complex biological systems and modeling gene regulation. Adept at integrating multi-omics data and implementing cutting-edge predictive modeling and genomic analysis techniques. Skilled in Python, R, and HPC systems. Highly motivated to bridge the gap between emerging technologies and clinical applications.

Education

Ph.D. in Data Science and Statistics (GPA: 3.90/4.00) Aug. 2020 - May 2025
The University of Texas at Dallas - Texas, USA

Certification: **Graduate Certificate in Data Science**

Courses: Statistical Inference, Statistical Methods, Bayesian Methods, Machine Learning, Applied Multivariate Analysis

Master of Science in Statistics, Specialization in Data Science(GPA: 3.94/4.00) Aug. 2020 - Dec. 2023
The University of Texas at Dallas - Texas, USA

Courses: Big Data Management, Machine Learning, Data Structures and Algorithms, Graph Theory

Bachelor of Science in Mathematics, Specialization in Statistics (GPA: 3.86/4.00) Feb. 2015 - Feb. 2019
University of Kelaniya, Sri Lanka

Courses: Operational Research, Design of Experiments, Stochastic Processes, Survey Methods and Sampling Techniques, Statistical quality control, Functional Analysis

Professional Experience

Postdoctoral Research Fellow - Versiti Blood Research Institute Jun. 2025- Present

- Leading research efforts on integrative genomic and transcriptomic analysis in Acute Myeloid Leukemia (AML).
- Applying advanced statistical and machine learning models (e.g., Cox regression, PCA, SVM) to identify prognostic markers and regulatory variants.
- Analyzing multi-omics datasets to uncover gene expression patterns associated with AML progression and risk stratification.
- Developing reproducible pipelines using R, Python, and HPC systems to support collaborative research and translational discovery.

Graduate Teaching Assistant - The University of Texas at Dallas Jan. 2021- May 2025

- Mentoring undergraduate students in machine learning, providing support in R and Python programming.
- Supported predictive modeling on diverse datasets, including clinical and genetic datasets, machine learning techniques such as regression, classification, clustering, decision trees, and neural networks.
- Guiding high-dimensional data processing, machine learning, and advanced statistical techniques.
- Collaborating with multidisciplinary teams to address complex, data-driven challenges.
- Leading discussion and tutoring sessions in mathematics and statistics, instructing over 60 students per semester.

Research Experience and Projects

Statistical Learning in Transcriptome - Wide Association Studies (*Dissertation*) Jan. 2022 - May 2025

- Developed machine learning models, including Bayesian Additive Regression Trees (BART), and applied joint modeling techniques to analyze gene interactions and integrate long- and short-range SNP interactions, improving prediction accuracy.
- Constructed spatially-resolved gene regulatory networks using graph-theoretic approaches and GTEx data, identifying enhancer-gene relationships and disease-associated regulatory variants.
- Designed graph-based models leveraging Hi-C and other regulatory datasets to explore gene interdependencies and 3D chromatin interactions.
- Created and implemented recursive algorithms and depth-first search techniques to identify biologically meaningful subgraphs and co-regulated gene clusters.

A mathematical model to analyze the dynamics of Dengue transmission (*Bachelor's Thesis*) Jan. 2018 - Jan. 2019

- Developed a vector-borne compartmental model to study Dengue transmission dynamics in Western Province, Sri Lanka.
- Examined the effects of climate change on Dengue transmission by incorporating time-varying mosquito-biting rates, revealing a significant correlation between Dengue incidence and average temperature.
- Demonstrated that Dengue transmission is more sensitive to mosquito-biting rate variations than mosquito population size.
- Validated the model through comparison with a test dataset, accurately predicting future outbreaks using R and MATLAB.

Document Summarization using NLP Techniques (*GitHub Repository*) Jan. 2023 - Apr. 2023

- Developed an extractive text summarization algorithm for text dataset based on sentence scoring using PySpark in Databricks.
- Implemented TF-IDF calculation from scratch without using standard libraries, scoring sentences using two methods: the sum of TF-IDF values and the average TF-IDF of non-stop words in each sentence

Neural Network Model for Wine Quality Prediction (*GitHub Repository*) May 2023 - Jun. 2023

- Built a neural network from scratch without relying on mainstream libraries, for predicting wine quality using physicochemical properties in the Wine Quality dataset (red wine).
- Implemented Random Forest for feature selection and gradient descent with momentum for optimization using Python, showcasing proficiency in feature selection and model evaluation.

Publications and Presentations

- **Gammune, D.H.V., Chen, M.** Bayesian Additive Regression Trees in Transcriptome-Wide Association Studies (Manuscript under preparation).
- **2024 Joint Statistical Meetings (JSM)** - Presented research on Statistical learning methods for TWAS and genomic data analysis.
- **External Advisory Council Meeting** - Presented a poster on Statistical learning methods for TWAS and genomic data analysis, focusing on integrating gene interactions.
- **Advances in Statistical and Computational Methods for Analysis of Biomedical, Genetic, and Omics Data (ABGOD 2023)** - Presented a poster on TWAS and gene interaction analysis.
- **Asian International Conference on Multidisciplinary Research 2019 (AIMR'19)** - Presentation on A mathematical model to analyze the dynamics of Dengue transmission (*Best Presenter Award*)

Skills

Research Proficiency:	Machine learning, regulatory genomics, statistical genetics, high-dimensional data analysis, Bayesian methods, chromatin interaction mapping, Survival analysis in cancer genomics
Technical tools:	R, Python, Bash/Shell scripting, SAS, Linux systems, SQL, Keras, TensorFlow, MATLAB, LaTeX
HPC:	MZ, Sysbio (Genomics cluster), Ganymede (CIRC cluster), AWS
Other:	Technical writing, collaboration, independent and teamwork, problem-solving, simplifying complex concepts for diverse audiences, extensive usage of HPC environments, submitting batch run jobs using job scheduling systems

Scholarships and Awards

- **Graduate Studies Scholarship**: Offered by the School of Natural Sciences and Mathematics, University of Texas at Dallas, based on graduate-level academic performance. (Aug. 2020 - Present)
- **Julia Williams Van Ness Merit Scholarship**: Offered by the School of Natural Sciences and Mathematics, University of Texas at Dallas, based on graduate-level academic achievements. (Aug. 2024)
- **NSM Conference Travel Award**: Offered by the School of Natural Sciences and Mathematics, University of Texas at Dallas, to present research projects at professional conferences. (Jun. 2024)
- **ABGOD Travel Award**: Offered by the conference committee of Advances in Statistical and Computational Methods for Analysis of Biomedical, Genetic, and Omics Data. (Mar. 2023)

Volunteer Work

- Served as a volunteer for the ABGOD conference, assisting with event organization, participant coordination, and session support to ensure a smooth experience for attendees and presenters.

Affiliations

- American Statistical Association (ASA): Member Jan. 2024
- Caucus for Women in Statistics and Data Science (CWS): Member Sep. 2024