Ritvik Prabhu

Email: ritvik@ritvikprabhu.com | **GitHub:** RitvikPrabhu | **Google Scholar:** Ritvik Prabhu

Education

- **Virginia Tech, Blacksburg, VA (Graduate Degree)** **Expected Graduation May 2030**
  - **PhD in Computer Science (CS)**
  - **Areas of Research**: High-Performance Computing; Algorithms; Advanced Machine Learning; Bioinformatics
  - **Cumulative GPA**: 3.75

- **Virginia Tech, Blacksburg, VA (Graduate Degree)** **Expected Graduation May 2026**
  - **Master's (MS) in Computer Science (CS)**
  - **Cumulative GPA**: 3.75

- **Virginia Tech, Blacksburg, VA (Undergraduate Degree)** **May 2023**
  - **BS in Computer Science (CS)**
  - **BS in Computational Modeling & Data Analytics (CMDA)**
  - **Minor**: Mathematics
  - **Cumulative GPA**: 3.74

Publications

- **R. Prabhu**, E. Vatai, B. Moussad, E. Jeannot, W. Feng, M. Wahib, "Looking for (Genomic) Needles in a Haystack: Sparsity-Driven Search for Identifying Correlated Genetic Mutations in Cancer," IEEE International Parallel & Distributed Processing Symposium (IPDPS), 2026. **[To Appear]**
- **R. Prabhu**, W. Feng, "Balancing Performance and Productivity: A Comparative Study of Apache Arrow vs. MPI," IEEE High-Performance Extreme Computing Conference, 2025.
- **R. Prabhu**, B. Moussad, K. Youssef, E. Vatai, W. Feng, "Analyzing Scalable Genomic Diversity Study using SparkLeBLAST: A Feasibility Study," IEEE High-Performance Extreme Computing Conference, 2024. [***Outstanding Student Paper Award***]
- A. Chaturvedi, **R. Prabhu**, M. Yadav, W. Feng, G. Cao, "Impact of Loss Function on the Performance of COVID-19 CT Image Denoising," IEEE Transactions on Radiation and Plasma Medical Sciences, 2024.
- H. Han, **R. Prabhu**, T. Smith, K. Dhakal, X. Wei, S. Li, C. North. "Interactive Deep Learning for Sorting Plant Images by Visual Phenotypes," Earth and Space Science Open Archive, 2022.
- H. Han, R. Faust, B. Norambuena, **R. Prabhu**, T. Smith, S. Li, C. North. "Explainable Interactive Projections for Image Data" 17th International Symposium on Visual Computing (ISVC), 2022.
- S. Farzanehpour, S. Rajeev, H. Liang, **R. Prabhu**, C. Brown. "ResearchBot: Bridging the Gap between Academic Research and Practical Programming Communities" International Workshop on Bots in Software Engineering, 2025.

Research and Work Experience

- **Oak Ridge National Laboratory: Student Intern** **May 2025 – August 2025**
  Oak Ridge, Tennessee, USA
  - Built tiled GEMM kernels in CUDA and a profiling harness to learn/optimize GPU memory hierarchy, occupancy, and warp scheduling. Then used Triton for rapid iteration and auto-tuning of fused kernels.
  - Designed a fused GEMM+SiLU (SwiGLU partial fusion) kernel in Triton, delivering $\sim$ 1.2–1.35$\times$ single precision speedups on large matrices vs. PyTorch baseline on Ampere and Hopper generation Nvidia GPUs.
  - Prototyped a full SwiGLU fusion in CUTLASS (C++) which exposed a Python callable function of the CUTLASS kernel.
  - Preparing a workshop paper on fused SwiGLU kernels.

- **RIKEN Center for Computational Science: Student Research Intern** **October 2024 – December 2024**
  Tokyo, Japan
  - Developed a scalable algorithm to identify multi-hit genetic mutation combinations associated with specific cancer types, targeting full-scale execution on the Fugaku supercomputer.
    → Aimed to advance cancer research by qualitatively improving existing approaches, which are currently limited to analyzing three- to four-hit mutation combinations, enabling exploration of more complex genetic interactions.
    → Designed and implemented the Sparsity-Aware Weighted Set Cover algorithm to tackle the exponential growth in computational requirements, enhancing the feasibility of analyzing complex mutation combinations.
    → Working towards demonstrating the algorithm's scalability on Fugaku's 158,976 nodes, with the goal of a potential **Gordon Bell** submission.

- **Synergy Lab: Graduate Research Assistant** **March 2022 – Present**
  Blacksburg, Virginia, USA
  - Conducted a feasibility study of SparkLeBLAST on an A64FX cluster and actively improved the analytics pipeline efficiency for potential exascale deployment on the Fugaku supercomputer.
    → Aimed to accelerate research on viral evolution to help predict and respond to future outbreaks in a timely manner, potentially informing treatments and vaccine development.
    → Analyzed the scalability and estimated performance for full-scale execution on the Fugaku supercomputer, identifying fine-grained bottlenecks and optimizing the in-memory representation and columnar layout for SparkLeBLAST.
    → Collaborated with the team at RIKEN as a **RIKEN Student Trainee** to enhance system performance by reducing format-thrashing, implementing efficient vectorization methods, and boosting computational throughput.
    → Explored additional scalability options for SparkLeBLAST using Apache Arrow and Arrow Flight to optimize in-memory data handling and support high-speed data transfer, enabling efficient processing of large datasets for exascale computing.
  - Contributing to the enhancement of COVID-19 identification in low-dose CT images using 3D-DDNet and VGG models.
    → Aimed to improve diagnostic accuracy of low-dose chest CT scans for COVID-19 and other lung diseases by enhancing image quality through advanced deep learning techniques.
    → Collaborated with team members on the implementation of 3D-DDNet for volumetric training of the 2D slices of CT-scans.
    → Explored methods to further enhance image quality and diagnostic accuracy using VGG models on 3D image models.
  - Implemented power-of-two FFTs for large sequence sets, each of length 8192, on edge devices (Raspberry Pi 5) via an OpenCL-based "write once, run anywhere" pipeline.

- **Amazon Web Services: Software Development Engineer Intern** **May 2022 – August 2022**
  Boston, Massachusetts, USA
  - Collaborated within a team of software engineers to manage and optimize indexing and querying processes for millions of IoT devices across multiple databases.
  - Developed a tracking feature to monitor and report on the indexing progress of IoT devices from various databases, integrating the data into an Elastic Search database for streamlined access.
  - Utilized AWS CloudWatch to design and implement graphical representations of backfill progress metrics, enabling efficient access to numerical insights through SQL queries.

- **InfoVis Lab: Undergraduate Research Assistant** **May 2021 – May 2022**
  Blacksburg, Virginia, USA
  - Engineered an interactive system, Andromeda, to enhance human-machine interaction by enabling the exploratory sorting of edamame seed images.
  - Developed a graphing tool within Andromeda, allowing users to cluster similar images based on domain expertise. Leveraged transfer learning techniques using pre-trained algorithms, including VGG-16 and ResNet variants, for initial image projection.
  - Refined the weighted distance function within the feature space, enhancing sorting accuracy by optimizing feature weights based on user-driven projections.

Awards and Professional Recognitions

- **Pratt Fellowship (2025-26 Academic Year)**: Awarded for academic excellence and demonstrated research potential within the College of Engineering.
- **Outstanding Master's Student Award Nomination**: Nominated by the Computer Science Department as its representative to compete for the College of Engineering's Best Master's Student Award.
- **Outstanding Student Paper Award**: Awarded for the paper "Analyzing Scalable Genomic Diversity Study using SparkLeBLAST: A Feasibility Study" presented at the IEEE High-Performance Extreme Computing Conference, 2024.
- **Featured on Stony Brook Super Computing Center Website**: Highlighted by Stony Brook University's Supercomputing Center as a "Success Story" for achievements in high-performance computing research.

Professional Service

- **Web Chair**, IEEE International Conference on Cluster Computing (CLUSTER), 2026.
- **Student Volunteer**, International Parallel & Distributed Processing Symposium (IPDPS), 2026.
- **Student Volunteer**, IEEE International Conference on Parallel Processing (ICPP), 2025.

Technical Skills

- **Technical Tools**: Git, Linux, Unix, Keras, Pytorch, Tensorflow, CUDA, Docker, OpenCL, MPI, OpenMP
- **Programming Languages:** Python, Java, R, C, MATLAB, SQL
- **Technical Knowledge**: Machine Learning, Artificial Intelligence, Natural Language Processing, Computer Vision, Computer Organization, Apache Arrow