Sajal Dash

802 Cascade Ct., Blacksburg, VA

Email : sajal@vt.edu
Mobile : +1-919-597-1857

Research Interest

Big data analytics using HPC and algorithmic machine learning; Cancer genomics; Deep learning; Incremental and approximate algorithms; Bioinformatics;

Education

- **Virginia Tech** — Blacksburg, VA
  PhD in Computer Science — *August, 2020*
- **University of North Carolina at Chapel Hill** — Chapel Hill, NC
  MS in Computer Science — *December, 2012*
- **Bangladesh University of Engineering and Technology** — Dhaka, Bangladesh
  BSc in Computer Science and Engineering — *March, 2009*

Publications

Journal Papers

- [**Scientific Reports, a natureresearch journal**] Differentiating between Cancer and Normal Tissue Samples using Multi-hit Combinations of Genetic Mutations; S Dash, NA Kinney, RT Varghese, HR Garner, W Feng, R Anandakrishnan; Scientific reports 9 (1), 1005, 2019 . [**Top 100 in cancer**]
- [**Scientific Reports, a natureresearch journal**] Identifying Multi-hit Carcinogenic Gene Combinations: Scaling up a Weighted Set Cover Algorithm using Compressed Binary Matrix Representation on a GPU; S Dash*, QA Hajri*, W Feng, H Garner, R Anandakrishnan* (***Contributed equally**).
- [**Information Visualization, in preparation**] Visual to parametric interaction on big data using accelerated WMDS, algebraic optimizations, and incremental gradient computation; S Dash, M Dowling, S Leman, C North, W Feng
- [**PLOS One, under review**] iBLAST: intelligent BLAST of new sequences via automated e-value correction; S Dash, S Rahman, HM Hines, W Feng; bioRxiv, 476218, 2018.

Conference Papers

- [**HPCC, 2017**] Portable Parallel Design of Weighted Multi-Dimensional Scaling for Real-Time Data Analysis; S Dash, A Verma, C North, W Feng; Published at IEEE International Conferences on High Performance Computing and Communications (HPCC), 2017 (**Best Paper Finalist**).
- [**IPDPS, 2021, under review**] Scaling Out a Combinatorial Algorithm for Discovering Carcinogenic Gene Combinations to Thousands of GPUs; S Dash, QA Hajri, W Feng, H Garner, R Anandakrishnan
- [**IJCAI, 2021, in preparation**] Mitigating Catastrophic Forgetting Using Historical Summary; S Dash, J Yin, M Shankar, W Feng

Workshop Papers and Posters

- [**DLS, 2019**] Strategies to Deploy and Scale Deep Learning on the Summit Supercomputer; J Yin, S Gahlot, N Laanait, K Maheshwari, J Morrison, S Dash, M Shankar; 2019 IEEE/ACM Third Workshop on Deep Learning on Supercomputers (DLS).
- [**AGU, 2018**] A Machine Learning Approach to Estimate Multi-Aerosol Mixing State Metric; Z Zheng, S Dash, D Schmidt, J Yin, N Riem, M West, VG Anantharaj; American Geophysical Union's Fall Meeting, 2018.
- [**ICPP, 2018**] Identifying Carcinogenic Multi-hit Combinations using Weighted Set Cover Algorithm; S Dash, N Kinney, R Varghese, H Garner, W Feng, R Anandakrishnan; International Conference on Parallel Processing, 2018.
- [**BIBMW, 2011**] On the Energy of Bifurcated Hydrogen Bonds for Protein Structure Prediction; S Dash and J Snoeyink; IEEE International Conference on Bioinformatics and Biomedicine Workshops (BIBMW), 2011.

Ongoing Research

- **Training Deep Learning Models in a Streaming Setting:** Innovating ways to train deep learning models in streaming setting using historic summary reconstruction to tackle catastrophic forgetting. This work is being supported through a collaboration between ORNL's leadership computing facility and Virginia Tech.
- **Identifying Multi-hit Combinations for Cancer using Super Computer:** Trying to identify combinations of $4-8$ gene mutations responsible for different types of cancer by mining cancer genomics data using hundreds of V100 GPUs on Summit Super Computer in a day instead of projected days and months.

Professional Experience

- **Oak Ridge National Laboratory** — Oak Ridge, TN
  Postdoctoral Research Associate — *October, 2018 - Present*
- **Oak Ridge National Laboratory** — Oak Ridge, TN
  ASTRO Research Intern — *Summer, 2018 and 2019*
  - **Deep Learning Best Practices:** Helped developing best practices for deep learning applications on supercomputers for various data domains.
  - **Training Deep Learning Models in Streaming Setting:** Designed and developed a system for training deep learning models in streaming setting with historical summary construction.
- **Plexxi Inc. (Now part of HPE)** — Nashua, NH
  Algorithm Architect Intern — *Summer, 2016 and 2017*
  - **Incremental Network Routing Algorithms:** Designed and implemented routing algorithms with redundancy to handle fault tolerance in private cloud network.
  - **Parallel Network Routing Algorithms:** Implemented parallel routing algorithms for fast routing in private cloud network.
- **Qualcomm** — San Diego, CA
  Engineer I — *02/2013-06/2014*
  - **Software Development for Test Automation:** Developed software for automating the bring-up procedure of first-ever 64-bit Android chip by Qualcomm. Developed APKs and binary files to monitor devices' userspace status. Wrote shell scripts to facilitate test data collection, crash detection, background tests in the device (Android Shell Script).
  - **Test Automation using Robots:** Developed and maintained logic for test-robots to physically imitate recorded user action (Perl, OpenCV, Arduino, Raspberry Pi and proprietary robots).
- **Computer Science, Virginia Tech** — Blacksburg, VA
  Graduate Research and Teaching Assistant — *08/2014 - Current*
  - **Using High Performance Computing in Big Data Analytics.:** Developed parallel algorithms for cancer genomics, sequence similarity search tool, and training deep learning models.
  - **Graduate Teaching Assistant:** Administered lab sessions for "Software Design and Data Structures" course.
- **Computer Science, UNC Chapel Hill** — Chapel Hill, NC
  Graduate Research and Teaching Assistant — *08/2010 - 12/2012*
  - **Graduate Research Assistant:** Designed and developed an algorithm for Boolean Operations on 2D Polygons that requires a lower number of bytes than traditional algorithms. Researched nature of bifurcated hydrogen bond using the Quantum Mechanical model of molecular systems.
  - **Graduate Teaching Assistant:** Administered lab sessions for and helped to instruct "Foundation of Programming" focusing on JAVA.
- **Stochastic Logic** — Dhaka, Bangladesh
  Quantitative Software Developer — *03/2009 - 08/2010*
  - **Developing Financial Software and Analytic Tools:** Analyzed financial data, made predictions, managed a portfolio, and analyzed credit derivatives using data mining techniques, machine learning, and probabilistic analysis using C++, C#, MATLAB, SQL Server, and ASP.NET. I developed Support Vector Machine and Genetic Algorithm based buy/sell decision-maker using MATLAB.
  - **Anomaly Detection and Prediction Tools:** Developed a tool for anomaly detection in the stock market using data mining techniques. Developed a prediction tool for analyzing product demand for a leading chain shop "Swopno" using MATLAB, C#.

Professional Services

- **Journal Reviewer:** Information Visualization - SAGE Journals.
- **Student Volunteer:** SC 2019, Women in Computing'11 at Virginia Tech.
- **Conference Attendance:** SC'19, ICPP'18, HPCC'17, BIBM'11, SOCG'12, and FWCG'10.
- **Math Olympiad:** Volunteered as Judge and Exam Controller of Bangladesh Mathematical Olympiad competitions.

Computer Skills

- **Programming Languages:** C++, JAVA, OpenCL, CUDA, C, C#, MATLAB, Python, Kotlin, Perl, Prolog, and ML.
- **Web Programming:** ASP.NET, PHP, HTML, and JavaScript.
- **Database Management System:** MS SQL Server, MySQL, and Oracle.
- **Deep Learning:** TensorFlow, PyTorch, Horovod, and MLPerf.
- **High Performance Computing:** CUDA, OpenCL, OpenMP, and MPI.
- **Other:** Android app development and OpenGL.

Projects

- **US Online Voting System:** Designed an online voting system using design patterns concepts.
- **Panorama Stitching:** *Panorama Stitching* tool and *3D Reconstruction* tool from sequence of 2D images using MATLAB.
- **Protein Morphing Simulation:** Simulation of protein vibration and morphing using C++, OpenGL.
- **Business Solution System:** Developed *Business Solution System* for an architects´ farm (Vita Arch) using MS SQL Server, C#, and ASP.NET.
- **Robotic Mastoidectomy:** Path planning for *Robotic Mastoidectomy* (a surgical procedure) using C++.
- **Poker Game:** Online Poker playing site using JAVA(J2SE) and MySQL.

Awards

- **SMC Data Challenge:** Finalist, SMC Data Challenge 2018, 2019, and 2020 organized by Oak Ridge National Laboratory.
- **VT-General Dynamics Data Challenge:** Winner of General Dynamics and Virginia Tech Network Flow Data Analysis Competition, 2015.
- **BUET-Microsoft Inter-University Math Olympiad:** 4^{th} prize in Inter University Math Olympiad, Bangladesh organized by CSE, BUET and Microsoft, Bangladesh on 2008.

Talk

- **HPCC'17 Plenary Session:** "Portable parallel design of weighted multi-dimensional scaling for real-time data analysis".

References

- Prof. Wu-chun Feng (PhD Advisor)
  Director, Synergy Laboratory
  Professor, Department of Computer Science, Department of Electrical & Computer Engineering, Health Sciences
  Virginia Tech, Blacksburg, USA
  Email: feng@cs.vt.edu

- Prof. Ramu Anandakrishnan (PhD Thesis Committee Member and Research Collaborator)
  Assistant Professor for Biomedical Sciences
  Edward Via College of Osteopathic Medicine, Blacksburg, USA
  Email: ramu@vt.edu

- Dr. Junqi Yin (Mentor at Oak Ridge National Laboratory and Research Collaborator)
  Computational Scientist and research staff member,
  Advanced Data and Workflows group,
  National Center for Computational Science,
  Oak Ridge National Laboratory
  Email: yinj@ornl.gov