

Shubhankar Gahlot

Huntsville, AL · (224) 307-4789 · sgahlot@hawk.iit.edu · linkedin.com/in/shubhankargahlot

Education

Computer Science / Environment Science, Ph.D., UAH / Uni. Of Canterbury

Aug 2022-Present

Data Science, M.S., Illinois Institute of Technology, Chicago

Aug 2016 – May 2018

Product Management for AI & Data Science, 365datascience (Online)

Dec 2022

Design (HCI), B.Des., Indian Institute of Technology, Guwahati (India)

Jul 2008 – May 2012

Skills

Machine Learning: 4 years' experience with Generative AI, Classification, Segmentation, Regression using Deep Learning, XGBoost, SVM, K-Means, Decision Trees, Bayesian Inference, Dim. Reduction: PCA, SVD, KDE, Simulation, AB-testing, Geospatial data science.

Programming: 5 years' experience with Python (NumPy, Pandas, mpi4py, Matplotlib, etc.), SQL, HDF5, PyTorch, TensorFlow, JavaScript, Docker, NVIDIA Rapids, AWS, Cloud computing.

Data engineering: 4 years' experience with data engineering, distributed ML (Ansible, Airflow, Dask, Ray, Kubernetes, OpenMPI)

Visualization: 4 years' experience with data visualization (charts, dashboards, data tools)

Management: 3 years' experience managing projects, teammates. One of the managed projects Rutogo was acquired by ixigo.com. ([article](#))

Communication: 5 years' experience presenting, writing (articles, papers, blogs, software documentation), webcasts.

Experience

Research Assistant, School of Earth and Environment, University of Canterbury

2022-2025

· **DevOps and data analysis:** Develop global climate data preprocessing pipeline for machine learning based solutions.

· **Software Development:** Collaborated with university research teams to implement GPU-accelerated deep learning models for atmospheric data analysis, mentoring graduate students in CUDA programming and enabling a 3x improvement in processing speed.

Research Scientist, NASA IMPACT, The University of Alabama, Huntsville, AL

2020-2022

· **Machine Learning Research:** Research, develop and deploy various deep learning technologies focused on NASA's IMPACT project vision of sustaining life cycle and accessibility of Earth Data. Computer vision and NLP research.

· **Data Engineering & Software Development:** Create tools for easy deployment and continuous integration of machine learning projects.

· **Project Management:**

- Led a multi-disciplinary team of 8+ researchers and engineers to manage and execute machine learning and data engineering projects, ensuring alignment with NASA's objectives.
- Developed project timelines, assigned tasks, and monitored milestones, resulting in the timely delivery of key deliverables across multiple projects.
- Managed stakeholder communication between research teams, project sponsors, and leadership, improving project transparency and reducing bottlenecks.
- Oversaw budget allocation and resource management for technology and cloud infrastructure, optimizing expenditures by 10%.
- Applied Agile project management methodologies (Scrum, Kanban) to streamline research workflows, improving team collaboration and output quality.

Scalable Data Science & Deep Learning Deployment Research Associate, Oak Ridge National Lab, Oak Ridge, TN

2018-2020

· **Data Collection & Statistical Analysis:** Utilized regression analysis and similarity measures to evaluate and benchmark various neural network operations based on input parameters, developing best practice guidelines for deployment. Achieved a **scaling efficiency of 87%** for a deep learning run of ResNet50 across **1,024 nodes (6,144 V100 GPUs)**. Presented findings published in SC'19.

· **ML Modeling & Image Classification:** Designed a computer vision ensemble classifier using the ResNet network to classify crystallography space groups that achieved an **accuracy of 70%**. The project was recognized as **runner-up in the SMC Data Challenge 2019**. ([presentation](#)).

· **Feature Extraction & Statistical Analysis:** Conducted statistical analysis on the impact of urban weather on energy use, with the project selected as **runner-up at the SMC Data Challenge 2018**. ([presentation](#))

· **Software Development:** Engineered scripts and tools to streamline development, deployment, and benchmarking of Machine Learning & Deep Learning frameworks such as TensorFlow, PyTorch, and Keras. These tools have been adopted by internal teams and contributed to **reducing deployment time by 50%**. ([gitlab](#))

· **Workshop & Tutorial Organization:** Organized and presented a **webinar** on building Deep Learning libraries (TensorFlow and PyTorch) from source on High Performance Computing (HPC) machines, attended by over **100 participants** from various research institutions. ([webcast](#))

· **Project Management:**

- Managed the **DLWorkflowWithPytorch** and **FitnessTracker4Workloads** projects, overseeing the development lifecycle from design to deployment. Defined project scope, allocated resources, set timelines, and coordinated with cross-functional teams to ensure timely delivery of milestones.
- Led the team in the development of a proof-of-concept dashboard for workload tracking on the Summit supercomputer, collaborating closely with research scientists and engineers to deliver real-time analytics on system performance.
- Coordinated cross-functional efforts for scaling machine learning models on the Summit supercomputer, managing stakeholder communication and documentation for the **Supercomputing 2019 (SC'19)** presentation and publication.

Sponsored Data Science Practicum (Data Analyst), Prospect Resources Inc. Chicago, IL

2017

· **Statistical Analysis:** Implemented and tested a buying strategy based on **Moving Average (MACD)** technique against existing regression-based buying strategy which helped in improving the hedging predictions in the electrical and energy sector. Results compiled in a technical report. ([report](#))

UX/ UI Design Consultant, Multiple Clients, India

2012-2016

· **Software Design:** 3+ years of experience in designing software for multiple clients in India such as WIPRO Technologies, Asian Paints, etc.

· **Product Design & Management:** 2+ years of experience in conceptualizing, designing & delivering software products and handling teams of multiple sizes for various products. One of the products conceptualized and managed **Rutogo** was **acquired** by ixigo.com. ([article](#))

Publications

- Gahlot, S.** (2024). Eco-Drive Revolution: Reinforcement Learning-Enhanced Cruise Control for Fuel Efficiency and Climate Impact. *Neural Information Processing Systems (NeurIPS), Ver.1*. <https://tinyurl.com/4js2s8b4>
- Gahlot, S.,** Gurung, I., Maskey, M., & Molthan, A. (2022). Flood extent data for machine learning. *NASA-IMPACT, Ver.1*. <https://doi.org/10.24432/C50P62>
- Gahlot, S.,** Kaulfus, A., Priftis, G., & Ramasubramanian, M., et.al. (2022). Time Series Machine Learning Methods for Surface PM2.5 Estimations Using Geostationary Satellites and Numerical Weather Models. *American Meteorological Society Annual Meeting*. <https://ntrs.nasa.gov/citations/20220000473>
- Gahlot, S.,** Gurung, I., Maskey, M., & Ramasubramanian, M., et.al. (2021). Leveraging citizen science and artificial intelligence for monitoring and estimating hazardous events. *American Geophysical Union*. <https://ntrs.nasa.gov/citations/20210025322>
- Gahlot, S.,** Gurung, I., Khatri, M., Maskey, M., & Ramasubramanian, M., et.al. (2021). Application of Artificial Intelligence for Surface PM2.5 Estimations from Geostationary Satellite and Atmospheric Numerical Model Data. *American Meteorological Society*. <https://ntrs.nasa.gov/citations/20205011654>
- Gahlot, S.,** Kaulfus, A., Priftis, G., & Ramasubramanian, M., et.al. (2021). A novel machine learning method for surface PM2.5 estimations from geostationary satellites. *American Geophysical Union*. <https://ntrs.nasa.gov/citations/20210024721>
- Gahlot, S.,** Ramasubramanian, M., Gurung, I., Hänsch, R., Molthan, A., & Maskey, M. (2022). Curating flood extent data and leveraging citizen science for benchmarking machine learning solutions. <https://doi.org/10.1002/essoar.10511103.1>
- Bollinger, A., **Gahlot, S.,** Gurung, I., Maskey, M., Ramachandran, R., & Ramasubramanian, M. (2021). Machine learning pipeline for Earth Science using Sagemaker. *American Geophysical Union*. <https://ntrs.nasa.gov/citations/20210024815>
- Acharya, A., Davis, C., **Gahlot, S.,** Koehl, D., & Ramasubramanian, M., et.al. (2021). Verb sense disambiguation for densifying knowledge graphs in Earth science. *American Geophysical Union*. <https://ntrs.nasa.gov/citations/20210025330>
- Gahlot, S.,** Shankar, A., & Yin, J. (2020). Data optimization for large batch distributed training of deep neural networks. *Computational Science & Computational Intelligence*. <https://doi.org/10.48550/arXiv.2012.09272>
- Dash, S., **Gahlot, S.,** Laanait, N., Maheshwari, K., Morrison, J., Shankar, M., & Yin, J. (2019). Strategies to deploy and scale deep learning on the Summit supercomputer. *Supercomputing*. <https://doi.org/10.1109/DLS49591.2019.00016>
- Dash, S., **Gahlot, S.,** Maheshwari, K., Morrison, J., Shankar, A., & Yin, J. (2019). Performance evaluation and best practice recommendations for extreme-scale machine learning and deep learning on Summit supercomputer. *AI Expo, Oak Ridge National Lab Postdoctoral Association Research Symposium*.
- Gahlot, A., & **Gahlot, S.** (2019). Changing the state of literacy in the digital age in India. *LINC 2019 Conference, MIT*. <https://doi.org/10.29007/qbpr>

Grant Sponsorship

- Gahlot, S.,** Katurji, M., & Zhang, J. (2024). Global atmospheric dynamics through Lagrangian coherent structures: A dataset for climate research and machine learning applications. *Climate Change AI NeurIPS*.
Amount: 150k USD

Mentorship & Workshop organization

- Enabling Analysis in the Cloud Using NASA Earth Science Data** Dec 2021
American Geophysical Union (AGU'21)
- Scaling Machine Learning for Remote Sensing using Cloud computing** June 2021
IEEE-GRSS Summer school on High-performance and Disruptive Computing in Remote Sensing
- Global Flood Detection Challenge** April 2021 – June 2021
NASA-IMPACT
Received participation from **137 participants, 460 submissions**.

Projects

- TraderGPT:** A Ollama based trader recommendation chatbot. Oct 2024

Benchmarking

- **Machine Learning:**
 1. Designed tests to benchmark in-house Natural Language Processing (NLP) model (**bert-e**) against the current state of the art (SOTA) models (**scibert** and **bert**) for extracting Earth Science terms from their definitions. Feb 2022
 2. Generated visualizations and animations to compare inferences from different DL segmentation models for smoke detection. Feb 2022
- **Cloud Computing:** Test and benchmark data pipeline ingress capability for Harmonized Landsat Sentinel 2 (**HLS-2**) Oct 2021

Machine Learning

- **Hurricane eye detection:** Developed DL point detection models for locating coordinates of hurricane eye in GOES-16 data. Sep 2021
- **Flood extent detection:** Developed benchmark DL segmentation models for detecting flood extent in open water bodies for **Global Flood Detection Challenge** organized by NASA-IMPACT. May 2021
- **PM2.5 estimation:** Developed a gradient boosted tree model to estimate the amount of PM2.5 in the atmosphere. Dec 2020

DLWorkflowWithPytorch ([github](#))

- Oak Ridge National Lab Oct 2019

- **Software Development:** Workflow tool for doing single/multi-GPU deep learning with ability to track multiple hyperparameters, code checkpointing and save the best model definition.