

# Shubhankar Gahlot

Huntsville, AL • +1 (224) 307-4789 • sgahlot@hawk.iit.edu • www.tinkerer.in

## EDUCATION

<b>Comp. Science / Env. Science, Ph.D.</b> UAH / University of Canterbury	Aug 2022 - Present
<b>Data Science, M.S.</b> Illinois Institute of Technology Chicago	Aug 2016 - May 2018
<b>Product Management for AI &amp; Data Science</b> 365datascience ( <a href="#">Online</a> )	Dec 2022
<b>Design (HCD), B.Des.</b> Indian Institute of Technology Guwahati	Jul 2008 - May 2012

## EXPERIENCE

<b>Research Assistant, School of Earth and Environment, University of Canterbury</b> • <b>ML and climate research:</b> Developed global climate data solutions using computer vision and LLM techniques. • <b>Software Development:</b> 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.	2022 - 2025
<b>Research Scientist, NASA IMPACT, University of Alabama Huntsville</b> • <b>Machine Learning Research:</b> Developed and deployed deep learning technologies for NASA's IMPACT project, focusing on Earth Data accessibility via computer vision and NLP. • <b>MLOps, Data Engineering &amp; Software Development:</b> Built tools for seamless deployment and continuous integration of ML projects. • <b>Project Management:</b> Led an 8+ member team to execute ML and data engineering initiatives, ensuring alignment with NASA objectives. Managed timelines, tasks, and milestones, achieving on-time delivery. Improved transparency through stakeholder communication and optimized budgets by 10%. Applied Agile methodologies to enhance team collaboration and output quality.	2020 - 2022
<b>Scalable Data Science &amp; Deep Learning Deployment Research Associate, Oak Ridge National Lab, Oak Ridge, TN</b> • <b>Data Analysis &amp; Benchmarking:</b> Conducted analysis to evaluate neural network operations, achieving 87% scaling efficiency for ResNet50 on <b>6,144 GPUs</b> ; findings published in SC'19. • <b>ML Modeling:</b> Designed a ResNet-based classifier for crystallography space groups (70% accuracy), earning runner-up in the SMC Data Challenge 2019. • <b>Software Development:</b> Built tools to streamline ML framework deployment (TensorFlow, PyTorch), reducing deployment time by 50%. ( <a href="#">gitlab</a> ) • <b>Outreach &amp; Training:</b> Presented a webinar on building DL libraries for HPC systems, attended by 100+ researchers. • <b>Project Management:</b> <ul style="list-style-type: none"><li>• Led a proof-of-concept dashboard for real-time workload tracking on Summit supercomputer.</li><li>• Coordinated efforts for scaling ML models on Summit, overseeing SC'19 presentation and publication.</li></ul>	2018 - 2020
<b>Sponsored Data Science Practicum (Data Analyst), Prospect Resources Inc. Chicago, IL</b> • <b>Statistical Analysis:</b> Implemented and tested a buying strategy based on <b>Moving Average (MACD)</b> 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. ( <a href="#">report</a> )	2017
<b>UX/ UI Design Consultant, Multiple Clients, India</b> • <b>Software Design:</b> 3+ years of experience in designing software for multiple clients in India such as WIPRO Technologies, Asian Paints, etc. • <b>Product Design &amp; Management:</b> 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 <i>Rutogo</i> was <b>acquired</b> by ixigo.com. ( <a href="#">article</a> )	2012 - 2016
<b>PROJECTS</b>	
<b>Eco-Drive Revolution:</b> Reinforcement Learning-Enhanced Cruise Control for Fuel Efficiency and Climate Impact (NeurIPS 2024)	
<b>Text to image diffusion model:</b> Trained a U-Net based diffusion autoencoder and CLIP for text input using Fashion MNIST dataset. ( <a href="#">github</a> )	
<b>Flood extent detection:</b> Designed and trained a multi-modal vanilla U-Net and Feature Pyramid Network models for flood extent segmentation achieving an IOU score of 0.77 ( <a href="#">presentation</a> )	
<b>Deep learning deploying and scaling strategies:</b> Conducted analytical tests and developed strategies to deploy and scale Deep learning on Summit supercomputer, achieving 87% scaling efficiency on 6,144 GPUs ( <a href="#">IEEE/ACM Workshop</a> )	
<b>Data optimization for large batch distributed training:</b> Developed a novel data optimization technique for distributed training of neural nets of up to 20k batch size ( <a href="#">DOI</a> )	
<b>Verb sense disambiguation for densifying knowledge graphs in Earth science</b> ( <a href="#">NASA</a> )	

**Large Language Model BERT-e:** Fine-tuned and benchmarked in-house LLM (bert-e) against scibert and bert for extracting Earth Science terms from their definitions. ([huggingface](#))  
**Towards a Universal Classifier for Crystallographic Space Groups:** Developed a computer vision model to classify crystallography space groups that achieved an **accuracy of 70%**. *SMC Data Challenge 2019* runner up ([presentation](#))

**Impact of Urban Weather on Energy Use:** Conducted statistical analysis on the impact of urban weather on energy use. *SMC Data Challenge 2018* runner up ([presentation](#))

**Back-testing hedging strategies:** Implemented and back-tested MACD and LSTM based hedging strategies against existing regression-based strategy which improved prediction by 6% ([report](#))

**PEER-REVIEWED  
CONFERENCE &  
JOURNAL  
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/qjs2s8b4>
- Gahlot, S.,** Gurung, I., Maskey, M., & Molthan, A. (2022). Flood extent data for machine learning. *NASA-IMPACT*, Ver.1. <https://doi.org/10.24432/C5oP62>
- 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>

**GRANTS &  
SPONSORSHIP**

- Funding for Global atmospheric dynamics through Lagrangian coherent structures: A dataset for climate research and machine learning applications *Climate Change AI NeurIPS*. (150k USD) 2024
- Recipient of New Zealand Ministry of Business Innovation and Employment Research funding *University of Canterbury* (35k USD) 2024

**WORKSHOP  
ORGANIZATION**

- [Enabling Analysis in the Cloud Using NASA Earth Science Data](#) Dec 2021
- [Scaling Machine Learning for Remote Sensing using Cloud computing](#) June 2021
- [Global Flood Detection Challenge NASA-IMPACT](#) (137 participants) April 2021 - June 2021
- [Webinar](#) on building Deep Learning libraries (TensorFlow and PyTorch) from source on High Performance Computing (HPC) machines Mar 2019

**AWARDS &  
HONORS**

- Appreciation for contribution to the GRSS Summer school on High-performance and Disruptive Computing in Remote Sensing 2021
- Sharing is Caring Newcomer Award NASA IMPACT 2020
- Smoky Mountains Computational Sciences & Engineering Conference (SMC) Data Challenge runner up 2018 and 2019

**MENTORSHIP**

- Ilboudo Dieudonne (Burkina Faso), NeurIPS mentee Aug 2024
- Ekansh Chauhan, IIIT Hyderabad MS student Mar 2023 - May 2023
- Shuto Araki, DePauw University BS student Jun 2019 - Aug 2019
- Emily Costa, FIU BS student Jun 2019 - Aug 2019

**SKILLS**

**Machine Learning:** 5+ years in Generative AI, Classification, Regression, Segmentation, XGBoost, SVM, Decision Trees, Bayesian Inference, PCA, SVD, KDE, A/B testing, and geospatial data science.  
**Programming:** Proficient in Python (NumPy, Pandas, PyTorch, TensorFlow, etc.), SQL, JavaScript, Docker, NVIDIA Rapids, and AWS. **Data Engineering:** 4+ years in distributed ML and data engineering with tools like Ansible, Airflow, Dask, Ray, Kubernetes, and OpenMPI.  
**Management:** 5+ years leading teams and projects, including Rutogo, acquired by ixigo.com. ([article](#))