Shubhankar Gahlot

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

Education

Data Science, M.S., *Illinois Institute of Technology, Chicago* **Aug'16 – May'18**
Design (HCI), B.Des., *Indian Institute of Technology, Guwahati (India)* **Jul'08 – May'12**
Product Management for AI & Data Science, *365datascience (Online)* **Dec'22**

Skills

Machine Learning: 3+ years' experience with Feature Engineering, Classification, Regression, XGBoost, SVM, K-Means, Decision Trees, Deep Learning, Bayesian Inference, Dim. Reduction: PCA, SVD, KDE, Simulation, Visualization, AB-testing, Geospatial data science.
Programming: 4+ years' experience with Python (NumPy, Pandas, mpi4py, Matplotlib, etc.), SQL, HDF5, PyTorch, TensorFlow, JavaScript, Docker, NVIDIA Rapids, AWS, Cloud computing.
Data engineering: 2+ years' experience with data engineering, distributed ML (Ansible, Apache Airflow, Prefect, Kubernetes, GitHub CI/CD, mpi4py, OpenMPI)
Visualization: 3+ years' experience with data visualization (charts, dashboards, data tools).
Management: 2+ years' experience managing projects, teammates. One of the managed projects Rutogo was acquired by ixigo.com. (article)
Communication: 2+ years' experience presenting, writing (articles, papers, blogs, software documentation), webcasts.

Coursework

Probability, Statistical Learning, Machine Learning, Big Data Technologies, Data Structures, Geospatial Vision and Visualization, Neural Networks and Deep Learning, Improving Deep Neural Networks: Hyperparameter tuning, Regularization and Optimization (*Coursera*)

Experience

Software developer & Data Engineer, *Kavita Antiques* **2022-2023**
· **Data Analysis:** Research, develop and deploy various strategies for antique products based on demand and supply modeling techniques.
· **Software Development:** Developed the online business tools including blog, e-commence store, and online payment system.

Research Scientist II, *NASA IMPACT, The University of Alabama, Huntsville, AL* **2020-2022**
· **Machine Learning Research:** Research, develop and deploy various ML / DL technologies focused on NASA's IMPACT project vision of sustaining life cycle and accessibility of Earth Data.
· **Data Engineering & Software Development:** Create mechanisms for easy deployment and continuous integration of NASA's IMPACT projects.
· **Mentor:** NASA Openscapes (2021 cohort) Worked together with other mentors to create cloud learning resources.
Tools: Python, TensorFlow, PyTorch, Keras, Ansible, Apache Airflow, Docker, AWS

Scalable Data Science & Deep Learning Deployment Research Associate, *Oak Ridge National Lab, Oak Ridge, TN* **2018-2020**
· **Data Collection & Statistical Analysis:** Used regression analysis and similarity measures to evaluate & benchmark various neural network operations based on input parameters and developed best practice guidelines to deploy different ANNs. Presented baseline examples of scaling efficiency 87% for a DL run of ResNet50 running on 1024 nodes (6144 V100 GPUs). Work **published** in **SC'19.** (ppt, publication)
· **Modeling & Image Classification:** Built an ensemble classifier using Resnet for classifying crystallography space groups. Project selected as runner up in SMC Data Challenge 2019. (presentation).
· **Feature Extraction & Statistical Analysis**: Statistical analysis of Impact of Urban Weather on Energy Use. Project selected as runner up at SMC Data Challenge 2018. (presentation)
· **Software Development:** Develop scripts and tools for easy development, deployment, and benchmarking of Machine Learning & Deep Learning frameworks like TensorFlow, PyTorch and Keras etc. (gitlab)
· **Workshop & Tutorial Organization:** Webinar on building Deep Learning libraries (TensorFlow and PyTorch) from source on High Performance Computing (HPC) machines. (webcast)
Tools: Python, Scipy, TensorFlow, PyTorch, Keras, HDF5, REST API, Docker, Horovod, NCCL, NVPROF, CUDA C, Shell, Git

Sponsored Data Science Practicum (Data Analyst), *Prospect Resources Inc. Chicago, IL* **2017**
· **Modeling:** Implemented a buying strategy based on **Moving Average (MACD)** technique and tested it against existing regression-based buying strategy which helped in improving the hedging predictions. Results compiled in a technical report. (report)
Tools: R (ggplot2, MACD), Python (Scipy), Google charts, Plotly

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

Flood Extent Data for Machine Learning, v1.0 **Mar'22**
NASA-IMPACT
***Shubhankar Gahlot**, Iksha Gurung, Andrew Molthan, Manil Maskey*

Leveraging Citizen Science and Artificial Intelligence for Monitoring and Estimating Hazardous Events **Dec'21**
American Geophysical Union (AGU'21)

***Shubhankar Gahlot**, Muthukumaran Ramasubramanian, Iksha Gurung, Manil Maskey*

A Novel Machine Learning Method for Surface PM2.5 Estimations from Geostationary Satellites **Dec'21**
American Geophysical Union (AGU'21)
*Georgios Priftis, Aaron Kaulfus, Muthukumaran Ramasubramanian, **Shubhankar Gahlot**, et.al.*

Curating flood extent data and leveraging citizen science for benchmarking machine learning solutions **Oct'21**
***Shubhankar Gahlot**, Muthukumaran Ramasubramanian, Iksha Gurung, Ronny Hänsch, Andrew L. Molthan, Manil Maskey*

Machine Learning pipeline for Earth Science using Sagemaker **Dec'21**
American Geophysical Union (AGU'21)
*Iksha Gurung, Muthukumaran Ramasubramanian, **Shubhankar Gahlot**, Andrew Bollinger, Manil Maskey, Rahul Ramachandran*

Verb Sense Disambiguation for Densifying Knowledge Graphs in Earth Science **Dec'21**
American Geophysical Union (AGU'21)
*Ashish Acharya, Carson Davis, Derek Koehl, Muthukumaran Ramasubramanian, **Shubhankar Gahlot**, et.al.*

Data optimization for large batch distributed training of deep neural networks **Dec'20**
Computational Science & Computational Intelligence (CSCI'20)
***Shubhankar Gahlot**, Junqi Yin, Arjun Shankar*

Strategies to Deploy and Scale Deep Learning on the Summit Supercomputer **Nov'19**
Supercomputing 2019 (SC'19)
*Junqi Yin, **Shubhankar Gahlot**, Nouamane Laanait, Ketan Maheshwari, Jack Morrison, Sajal Dash, Mallikarjun Shankar*

Performance Evaluation and Best Practice Recommendations for Extreme-Scale Machine Learning and Deep Learning on Summit Supercomputer **Aug'19**
AI Expo 2019, Oak Ridge National Lab
Oak Ridge Postdoctoral Association Research Symposium 2019, Oak Ridge National Lab
***Shubhankar Gahlot**, Junqi Yin, Ketan Maheshwari, Jack Morrison, Sajal Das, Arjun Shankar*

Changing the state of literacy in Digital Age in India **June'19**
LINC 2019 Conference, MIT
*Aanandita Gahlot, **Shubhankar Gahlot***

Workshops and competitions

Enabling Analysis in the Cloud Using NASA Earth Science Data **Dec'21**
American Geophysical Union (AGU'21)
Cynthia Hall, Makhan Virdi, Alexis Hunzinger, Amy Steiker, Catalina M Oaida, Aaron M Friesz, et.al.

Scaling Machine Learning for Remote Sensing using Cloud computing **June'21**
IEEE-GRSS Summer school on High-performance and Disruptive Computing in Remote Sensing
*Iksha Gurung, Muthukumaran Ramasubramanian, **Shubhankar Gahlot**, et.al.*

Global Flood Detection Challenge **April'21 – June'21**
NASA-IMPACT
***Shubhankar Gahlot,** Iksha Gurung, Muthukumaran Ramasubramanian*

Projects

Benchmarking

· **Machine Learning**:

1. Designed tests to benchmark in-house Natural Language Processing (NLP) model (bert-e) against current state of the art (SOTA) models (scibert and bert) for extracting Earth Science terms from their definitions. **Feb'22**
   Tools: Python
2. Generated visualizations and animations to compare inferences from different DL segmentation models for smoke detection. **Feb'22**
   Tools: Python, QGIS, iMovie

· **Cloud Computing**: Tested and benchmarked data pipeline ingress capability for Harmonized Landsat Sentinel 2 (HLS-2) **Oct'21**
Tools: Python, AWS CDK, PostgreSQL

Machine Learning

· **Hurricane eye detection:** Developed DL point detection models for locating coordinates of hurricane eye in GOES-16 data. **Sep'21**
· **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'21**
· **PM2.5 estimation:** Developed a gradient boosted tree model to estimate the amount of PM2.5 in the atmosphere. **Dec'20**
Tools: Python, PyTorch, Numpy, Scikit-learn, JupyterLab, Bash

DLWorkflowWithPytorch (github) **Oct'19**
Oak Ridge National Lab
· **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.
Tools: Python, Pytorch, Numpy, JupyterLab, Bash