SUBHASHIS HAZARIKA

Email: *hsubhashis@gmail.com* | **Website:** *subhashis.github.io* | *Google Scholar* | **Phone:** *+1 614-462-9957*

RESEARCH INTEREST

Statistical Data Modeling | **Machine Learning** | **Explainable AI** | **Uncertainty Quantification** |**Visualization**

EDUCATION

M.S. and Ph.D. | ***Computer Science and Engineering*** Aug. 2013 – Dec. 2019
The Ohio State University Columbus, OH

- **Major:** *Computer Graphics and Data Visualization*
- **Minors:** *Artificial Intelligence*, *High Performance Computing*
- **Thesis:** *Statistical and Machine Learning Approaches for Visualizing and Analyzing Large-Scale Simulation Data*

B. Tech. | ***Computer Science and Engineering*** Aug. 2007 – May 2011
National Institute of Technology, Durgapur West Bengal, India

DISTINGUISHING ATTRIBUTES

- Experienced working on interdisciplinary data science projects across diverse scientific domains.
- Experienced combining statistical modeling, machine learning, visualization, and HPC technologies

WORK EXPERIENCE

Postdoctoral Researcher Jan. 2020 – present
Los Alamos National Laboratory | ***Data Science at Scale Team, Information Sciences (CCS-3)*** Los Alamos, NM

- ***In situ Statistical Inference***: Applying and developing bayesian statistical models *in situ*, (i.e, while the data still resides in the memory) for large-scale scientific simulations. This facilitates in performing various scientific data analyses and investigations without having to store the large-scale datasets.
- ***Adaptive Sampling based Data Reduction***: Implementing scalable feature-driven importance sampling algorithms to reduce the storage footprint of large-scale scientific data, while preserving importance data features. This is part of the Exascale Computing Project (ECP) to develop analysis algorithms for upcoming exascale (10^{18} FLOPS) capable supercomputers.

Graduate Research Associate May 2014 – Dec. 2019
Gravity Research Group | ***The Ohio State University*** Columbus, OH

- ***Explainable AI for Visual Analysis***: Utilized a trained neural network (NN) as an inexpensive surrogate model for a computationally-intensive scientific simulation. By applying various techniques for *interpreting*, *explaining*, and *quantifying uncertainty* of trained NNs, we designed a visual analytic system to facilitate interactive data analysis. **[Best Paper Honorable Mention at IEEE Vis (VAST) 2019]**
- ***Copula-based Multivariate Distribution Modeling***: Developed a flexible distribution-driven analysis framework for modeling multivariate distributions in an efficient manner using *Gaussian copula* functions. Utilized the copula functions to perform probabilistic feature detection and tracking.
- ***Information-theoretic Framework for Uncertainty Analysis***: Using information-theory measures like mutual-information, specific-information and conditional entropy, we proposed novel analysis techniques to quantify as well as visualize the uncertainty of important features in scientific datasets.

Graduate Research Intern May 2019 – Aug. 2019
Los Alamos National Laboratory | ***Data Science at Scale, Applied Comp. Science (CCS-7)*** Los Alamos, NM

- Multivariate relationship-aware adaptive sampling using principal component analysis.

Graduate Research Intern May 2017 – Aug. 2017
Los Alamos National Laboratory | ***Programming Models, Applied Comp. Science (CCS-7)*** Los Alamos, NM

- Implemented data analysis and visualization methods for ocean simulation data using *Regent*, a task-based programming parallel language.

Senior Software Engineer June 2011 – May 2013
Novell Software Development (India) Pvt. Ltd. Bengaluru, India

- Part of *ZENworks Configuration Management* design team.

Undergraduate Intern May 2010 – Aug. 2010
European Organization for Nuclear Research (CERN) Geneva, Switzerland

- Developed an end-to-end system to collect and visually track the sensor data from electromagnetic calorimeters in particle detectors of the LHC project.

Graduate Teaching Instructor Aug 2014 – April 2016
Dept. of Computer Science | ***The Ohio State University*** Columbus, OH

- *CSE 1222:* Introduction to Computer Programming in C++. In person classroom lectures and grading.

Technical Skills

Programming: C/C++, Python, Julia(elementary)
Visualization: VTK, VTK-m, OpenGL, D3.js
Stats/Machine Learning: Numpy, SciKit-Learn, Keras, PyTorch

Honors and Awards

- Best Paper Honorable Mention Award at IEEE Vis (VAST) 2019
- O'Donnell Graduate Fellowship for Ph.D, 2013
- Keynote Speaker at ICMLIP 2020 (Hyderabad, India)
- Summer Student at CERN, Geneva, 2010

Technical and Academic Services

- Organizing committee member (video chair) for IEEE Vis, 2021, New Orleans
- Student Volunteer for IEEE Vis (2017, 2018)
- Reviewer: IEEE TVCG, MDPI Entropy, IEEE PacficVis, EuroVis

Peer-Reviewed Publications

- **Subhashis Hazarika**, Haoyu Li, Ko-Chih Wang, Han-Wei Shen, Ching-Shan Chou: "*NNVA: Neural Network Assisted Visual Analysis of Yeast Cell Polarization Simulation* ", IEEE Transactions on Visualization and Computer Graphics, 26 (1), 34-44 (2020). **[Best Paper Honorable Mention at IEEE Vis (VAST) 2019]**.
- **Subhashis Hazarika**, Ayan Biswas, Phillip J. Wolfram, Earl Lawrence, Nathan Urban: "*Relationship-aware Multivariate Sampling Strategy for Scientific Simulation Data*", arXiv:2008.13306 (Accepted at IEEE Vis 2020 Shortpaper).
- Piyush Chawla, **Subhashis Hazarika**, Han-Wei Shen: "*Token-wise Sentiment Decomposition for ConvNet: Visualizing a Sentiment Classifier*", Visual Informatics, Elsevier 2468-502X (2020).
- **Subhashis Hazarika**, Soumya Dutta, Han-Wei Shen, Jen-Ping Chen: "*CoDDA: A Flexible Copula-based Distribution Driven Analysis Framework for Large-Scale Multivariate Datasets*", IEEE Transactions on Visualization and Computer Graphics, 25(1): 1214-1224 (2019).
- Junpeng Wang, **Subhashis Hazarika**, Cheng Li, Han-Wei Shen: "*Visualization and Visual Analysis of Ensemble Data: A Survey*", IEEE Transactions on Visualization and Computer Graphics, 25(9): 2853-2872 (2019).
- Qun Liu, **Subhashis Hazarika**, John M Patchett, James Paul Ahrens, Ayan Biswas: "*Deep Learning-Based Feature-Aware Data Modeling for Complex Physics Simulations*", International Conference for High Performance Computing, Networking, Storage, and Analysis (SC 2019).
- **Subhashis Hazarika**, Ayan Biswas, Han-Wei Shen: "*Uncertainty Visualization Using Copula-Based Analysis in Mixed Distribution Models*", IEEE Transactions on Visualization and Computer Graphics, 24(1): 934-943 (2018).

- **Subhashis Hazarika**, Ayan Biswas, Soumya Dutta, Han-Wei Shen: "*Information Guided Exploration of Scalar Values and Isocontours in Ensemble Datasets*", Entropy 2018, 20(7), 540. (Special Issue Information Theory Application in Visualization).
- **Subhashis Hazarika**, Soumya Dutta, Han-Wei Shen: "*Visualizing the Variations of Ensemble of Isosurfaces*", IEEE Pacific Visualization Symposium (PacificVis), 2016, 209-213.
- **Subhashis Hazarika**, Tzu-Hsuan Wei, Rajaditya Mukherjee, Alexandru Barbur: "*Visualizing the life and anatomy of dark matter*", IEEE Scientific Visualization Conference (SciVis), 2015, 101-106.
- Sanjib Sadhu, **Subhashis Hazarika**, Kapil Jain, Saurav Basu, Tanmay De: "*GRP-CH Heuristic for Generating Random Simple Polygon*", 23rd International Workshop on Combinatorial Algorithms 2012: Page 293-302, Springer LNCS Volume.

Book Chapters

- Soumya Dutta, **Subhashis Hazarika**, Han-Wei Shen: "*In Situ Statistical Distribution-based Data Summarization and Visual Analysis* ", In Situ Visualization for Computational Science, 2021. Publisher: Springer. (*Post-editing phase*)