
Sujata Sinha

PhD Candidate @ Virginia Tech, Visiting Researcher @ Argonne National Laboratory

sujatasinha@vt.edu, linkedin.com/in/sujatassinha

EDUCATION **PhD, Computer Engineering, Virginia Tech, USA** Dec 2026
Research Interests: Data Compression, AI for Science, Random Field Theory **Advisor:** Dr. Lingjia Liu
MS, Computer Science & Software Engineering, Auburn University, USA May 2021
Thesis: Development of Deep Learning Models for Biomedical Applications **Advisor:** Dr. Jingyi Zheng
B.Tech, Electronics & Communication Engineering May 2018
Uttar Pradesh Technical University, Lucknow Summa cum laude

PROFESSIONAL EXPERIENCE **Virginia Tech, VA** *Research Assistant* Aug 2021 – Present

- Developed the first theoretical compression limits for HPC scientific datasets under real-world parallelism constraints; framework applies directly to LLM checkpoint storage and model serving.
- Extended Shannon’s rate-distortion theory from classical 1D assumptions to finite-blocklength ND heterogeneous fields, closing a 50+ year gap between theory and practice in scientific compression.
- Built large-scale statistical validation pipeline ($> 10^6$ samples) using EM clustering, AIC/BIC, and likelihood-ratio testing to verify model fidelity across 33192 multi-dimensional scientific datasets.
- Established probabilistic frameworks for modeling token and scientific data distributions in large-scale AI systems, enabling tractable estimation of source complexity and scaling behavior.
- Developed information-theoretic optimization framework for compute-efficient allocation across model scale and token budgets, enabling analysis of scaling tradeoffs in foundation models.

Argonne National Laboratory, IL *Visiting Researcher* May 2025 – Present

- Developed covariance and correlation estimators for scientific data, improving spatial correlation modeling and covariance calibration in N-dimensional stochastic systems.
- Established information-theoretic bounds under metrics of interest, quantifying information efficiency, transport cost, and estimation risk in non-stationary environments.
- Built an modular Python tools for patchwise statistical analysis across 10,000+ slices, enabling adaptive compression strategies based on local signal complexity and autocorrelation structure.

Auburn University, AL *Research Assistant* Aug 2018 – May 2021

- Engineered deep neural architectures (U-Net, Residual U-Net) to segment 100,000+ MRI images, achieving over 99% accuracy and addressing large-scale data imbalance with AUC-optimization.
- Developed time-series analytics framework leveraging Hilbert-Huang Transform and deep learning to automate frequency band extraction, boosting classification performance to over 90%.

Pulse Prognostics, India *Co-Founder* Jan 2017 – June 2018

- Designed and led development of a real-time signal processing pipeline that transformed wrist pulse sensor data into personalized wellness and health analytics.
- Led cross-functional startup execution securing competitive early-stage funding through the VG Summit and BIRAC Seed Fund while establishing strategic partnerships and early user traction.

- PUBLICATIONS**
1. **Sujata Sinha**, Vishwas Rao, Robert Underwood, David Lenz, Sheng Di, Franck Cappello, & Lingjia Liu. *Rate-Distortion Bounds for Heterogeneous Random Fields on Finite Lattices*. In Review.
 2. **Sujata Sinha**, Sheng Di, Vishwas Rao, Robert Underwood, David Lenz, Lingjia Liu, & Franck Cappello. *Bridging Information Theory and Practice for Scientific Lossy Compression*. ACM HPDC (2026).
 3. Ramashish Gaurav, **Sujata Sinha**, Chunxiao Lin, Terrence C Stewart, Lingjia Liu, & Yang Yi. *Multivariate Legendre-SNN on Loihi-2 for Time Series Classification*. IEEE TCASAI (2025).
 4. **Sujata Sinha**, & Alkan Soysal. *Adversarial Attacks and Defenses for Wireless Signal Classifiers using CDI-aware Generative Adversarial Networks (GANs)*. IEEE ICC (2024).
 5. **Sujata Sinha**, & Alkan Soysal. *Channel Aware Adversarial Attacks are Not Robust*. IEEE MILCOM (2023).
 6. Jingyi Zheng, Mingli Liang, **Sujata Sinha**, Linqiang Ge, Wei Yu, Arne Ekstrom, & Fushing Hsieh. *Time-frequency analysis of scalp EEG with Hilbert-Huang transform and deep learning*. IEEE JBHI (2021).
 7. **Sujata Sinha**, Thomas Denney, Yang Zhou, & Jingyi Zheng. *Automated Semantic Segmentation of Cardiac Magnetic Resonance Images with Deep Learning*. IEEE ICML-Applications (2020).

TECHNICAL SKILLS

- Python, PyTorch, TensorFlow, Git.
- Data Compression, AI for Science, Multidimensional Random Fields, Theoretical ML