Rajashik Datta

+91 62941 32431 rajashikdatta215@gmail.com Google Scholar LinkedIn Github Portfolio

Research Interests

Computer Vision (Hyperspectral/Multimodal), Trustworthy ML & explainable AI, Representation/Transfer Learning, Online Learning under Distribution Shift

Education

Institute of Engineering & Management, Kolkata, India **August 2022 – July 2026**
B.Tech in Computer Science & Engineering (Artificial Intelligence) *CGPA: 9.19 / 10*
Ranked 6th in the top 10% of class in Year 3 (AY 2024–25).

D.A.V. Public School, Siliguri, India **April 2020 – March 2022**
Higher Secondary School *CBSE: 75.4 / 100*

Nirmala Convent School, Jalpaiguri, India **April 2008 – March 2020**
Primary School, High School *ICSE: 90 / 100*

Experience

University of Nebraska-Lincoln, USA **June 2025 – Present**
Research Intern *Remote (USA)*

Supervisor: *Dr. Sruti Das Choudhury* (*Offer Letter*)
- Spearheaded an explainable AI + data-storytelling clustering pipeline across precision agriculture and pediatric healthcare—grouping 22 Indian crop types using 7 agro-climatic/soil features and segmenting a 500-record hospital cohort—showing that z-score rescaling + removing binary gender prevents charge-dominated clusters and surfaces clinically meaningful cohorts (LOS up to 29 days; charges up to 34,644) for decision support.
- Developed a temporal-embedding visual analytics system for 42 plants from 9 genotypes over 25 days, engineering multi-scale phenotype descriptors (growth rates/accelerations, fourier spectra, wavelet energies, distributional stats) and achieving genotype-aligned DTW clustering (ARI 0.30; NMI 0.62) with cross-validated early-prediction curves and SHAP/LIME-linked causal graphs to explain when/why genotypes diverge.
- Implemented an interactive hyperspectral analysis tool, HyperProbe for calibrated datacubes spanning 517-1700 nm (B=243 bands), enabling rapid pixel/ROI annotation, band-difference + Otsu segmentation (IoU/F1 evaluation), and full-scene classification via 3 model families (MLP/logistic regression/random forest) with built-in ablations that log clicks/ROIs under fixed 5-min label budgets to quantify accuracy-per-effort.
- Featured in the university's news story for research contributions: snr.unl.edu (August, 2025)

University of Calcutta **January 2025 – Present**
Research Scholar *Kolkata, India*

Supervisors: *Dr. Arup Kumar Chattopadhyay*, *Prof. Amit Kumar Das*, *Prof. Amlan Chakrabarti*
- Engineered FHFAM (FH-FAM), a fuzzy-hypergraph feature selection algorithm, achieving the best mean accuracy (81.43%) and best mean feature reduction (89.28%) across **15** agriculture/remote-sensing datasets (5/15 wins) with 11.08s average runtime and statistically significant accuracy gains over key baselines (Wilcoxon $p < 0.05$).
- Proposed SIFHFAM, a stage-wise intuitionistic-fuzzy hypergraph selector with a monotone submodular coverage objective and greedy (1−1/e) guarantee, delivering the top average accuracy (≈84%) while pruning ≈99% features (typically retaining $< 2\%$) across 14 high-dimensional benchmarks in ∼0.1s/run under 10× repeated 75/25 train-test splits.

Generative AI Centre of Excellence, IEM **November 2024 – December 2025**
Student Research Lead at GenAI CoE *Kolkata, India*

Led GenAI CoE's end-to-end research execution and operations—recruited and onboarded members via interviews, mentored and staffed project teams, coordinated 10+ journal groups, maintained the CoE website, and launched *ReelBook* (Pearson collaboration) and Medium publishing to scale institute-wide research output and AI upskilling at IEM.

IEM Research Foundation **August 2024 – March 2025**
Project Intern at bair.ai (Certificate) *Kolkata, India*

Built *MemeMetric*, an end-to-end cluster-based cryptocurrency forecasting system by architecting the full data/ML pipeline with automated reporting, and integrated real-time Twitter/Telegram/Reddit sentiment signals via NLP to improve robustness and reduce forecast error/volatility.

Innovation & Entrepreneurship Development Cell (CSE), IEM **March 2024 – August 2024**
Research Assistant (Certificate) *Kolkata, India*

Co-authored an IEM-HEALS 2024 accepted study analyzing Jul 2019–Dec 2022 price dynamics of 20 pharma stocks using multivariate regression, volatility modeling, and event-study methods, and engineered *TraderBot*, a Flask+MongoDB real-time trading simulator wired to Yahoo Finance for live strategy backtesting and portfolio experiments.

National University of Singapore (NUS), Singapore **July 2023 (1 week)**
Study Abroad Program (Certificate) *Singapore*

Studied fundamentals of "Artificial Intelligence, Internet of Things, Machine Learning & Data Analytics", lectured by *Dr. Peter Leong*, *Dr. Eric Cambria*, *Dr. Matthew Chua*, *Dr. Yiliang Zhao*, *Dr. Gábor Benedek*, *Dr. Tan Kian Hua*, *Yong Heng Michael Tan*, *Marton Szel*, *Gillian Cheng*.

Publications

Published/Accepted

1. Sanjan Baitalik, **Rajashik Datta**, Utsho Banerjee, Rajarshi Karmakar, Vincent Stoerger, Himadri Nath Saha, Sruti Das Choudhury, "*ReproPheno and ReproPhenoNet: A Large-Scale Multimodal Benchmark Dataset and Deep Learning Framework for Reproductive-Stage Plant Phenotyping*", AAAI Workshop 2026, 2026. *(Published)*

2. **Rajashik Datta**, Sanjan Baitalik, Amit Kumar Das, Sruti Das Choudhury, "PlantPhenoLM: Phenotype-Genotype Mapping Inference with Multi-Turn LLM Reasoning and Selective Prediction", AAAI Bridge 2026, 2026.

3. Sanjan Baitalik, **Rajashik Datta**, Amit Kumar Das, Sruti Das Choudhury, "Conversation as Belief Revision: GreedySAT Revision for Global Logical Consistency in Multi-Turn LLM Dialogues", AAAI Bridge 2026, 2026.

4. **Rajashik Datta**, Sanjan Baitalik, Sruti Das Choudhury, Arup Kumar Chattopadhyay, Amit Kumar Das, "Fuzzy Hypergraph Feature Association Map for High-Dimensional Feature Selection in Agriculture and Remote Sensing", International Journal of Fuzzy Systems, 2026.

5. Sruti Das Choudhury, **Rajashik Datta**, Sanjan Baitalik, "Enhancing Interpretability Through Clustering, Explainable AI, and Narrative Visualization: Applications in Precision Agriculture and Healthcare Patient Segmentation", Information, 2025.

6. Sanjan Baitalik, **Rajashik Datta**, Sanket Ghosh, Darothi Sarkar, Ayan Chaudhuri, "Machine Learning-Driven Insights For Stock Market Analysis And Trading", International Conference on Interdisciplinary Research in Technology and Management (IRTM 2024).

7. Sanket Ghosh, Sanjan Baitalik, **Rajashik Datta**, Darothi Sarkar, "The COVID-19 Shock: An Analysis Of Impacts And Responses Of Indian Stock Market", International Conference on Interdisciplinary Research in Technology and Management (IRTM 2024).

8. **Rajashik Datta**, Sanjan Baitalik, Sanket Ghosh, Saugata Ghosh, Swarnendu Ghosh, "Is Indian Financial Market Ready for Pandemics?". In International Conference on Advancing Science and Technologies in Health Science (IEM-HEALS 2024) *Book of Abstracts*.

Submitted

1. Sanjan Baitaik, **Rajashik Datta**, Arup Kumar Chattopadhyay, Amit Kumar Das, Amlan Chakraborty, "Greedy Optimization with Provable Guarantees for Non-Uniform Intuitionistic Hypergraph-Based Feature Selection". Intended for submission to Pattern Recognition, 2026.
2. Sanjan Baitalik, **Rajashik Datta**, Darothi Sarkar, Ayan Chaudhuri, MiQ-MCP: Valid and Conditionally Robust Uncertainty Quantification for High-Frequency Financial Time Series via Mondrian Conformalized Quantile Regression (GitHub: MiQ-MCP), Computational Economics, 2025.

Skills & Activities

Programming: Python, C, C++, Java, MATLAB, LaTeX
ML/AI: PyTorch, TensorFlow, Scikit-learn, Transformers
XAI: SHAP, LIME
Data: Pandas, NumPy, SciPy
Databases: MySQL, PostgreSQL, MongoDB
Cloud: Google Cloud (Cloud Run/Compute), AWS (S3/EC2)
Visualization: Matplotlib, Seaborn, Plotly, Tableau
Tools: TensorBoard, MATLAB App Designer

Activities: GenSpark 1.0 Ideathon (Organizer; coordinated 50+ teams; shortlisted funded ideas), Jun–Aug 2025; IEM-ICDC 2025 (Conference volunteer; coordination & support), Apr 2025; Department of CSE, IEM (Assisted NBA accreditation documentation), Mar 2024

Projects

Quantization-Aware Momentum | *GitHub*

1-bit Momentum optimizer with error feedback; matches full-precision Momentum SGD. Logistic regression (n=4000, d=2000, 5000 steps): train loss 2.809×10^{-3} vs signSGD 3.7614×10^{-2} (13.39× higher); remains 6–13× better across weight decay sweeps.

Online Learning (VS-AdaGrad) | *GitHub*

Online learning for non-stationary time series via drift-aware, volatility-scaled AdaGrad; on piecewise AR(5) with 5 regimes (T≈5000, 10 seeds), improves regret proxy over AdaGrad by 18.4% (small drift) and 19.8% (medium), and beats tuned OGD by 23.7–63.8% across drift regimes.