Sanjan Baitalik

+91 89106 75837 rayan.baitalik@gmail.com Google Scholar LinkedIn GitHub Portfolio

Research Interests

Representation/Transfer Learning, Trustworthy ML (XAI + robustness), Online Learning under Distribution Shift, Computer Vision (Hyperspectral/Multimodal)

Education

Institute of Engineering & Management(IEM), Kolkata **August 2022 – Expected July 2026**

B.Tech in Computer Science & Engineering *(Transcripts) CGPA: 9.33/10*

Ranked in the top 15 % of class by CGPA.

Experience

University of Nebraska, Lincoln, USA **June 2025 – Present**

Research Intern *Remote*

Supervisor: *Dr. Sruti Das Choudhury* (*Offer Letter)*

- Co-authored a human-centered XAI study that combined k-means/hierarchical clustering, SHAP-driven interpretability, and narrative visualization to turn a 22-crop agronomic dataset and a 500-record/6-variable hospital-cost cohort into 4 actionable clusters (e.g., crop regimes sized 797/618/585/200), demonstrating how rescaling/categorical handling improves cluster trustworthiness for decision-making.
- Built an interactive visual-analytics pipeline on UNL greenhouse phenotyping data (42 plants, 9 genotypes, 25 days) that couples temporal embeddings, DTW-based clustering, and SHAP/LIME-linked causal views, achieving ARI=0.30 and NMI=0.62 genotype–cluster agreement while enabling "early prediction" analysis via accuracy-vs-day curves.
- Engineered HyperProbe, a lightweight Streamlit-based human-in-the-loop hyperspectral tool spanning 517-1700 nm (243 bands) that integrates pixel/ROI selection, derivative + spectral-angle analytics, Otsu band-difference segmentation, interactive MLP/LR/RF classification, clustering, and PCA/t-SNE embeddings into a single end-to-end workflow.
- Featured in the university's news story snr.unl.edu (August, 2025).

University of Calcutta **January 2025 – December 2025**

Research Scholar *Kolkata, India*

Supervisors: *Dr. Arup Kumar Chattopadhyay*, *Prof. Amit Kumar Das*, *Prof. Amlan Chakrabarti*

- Developed and benchmarked FH-FAM, a fuzzy-hypergraph feature-selection algorithm for high-dimensional agriculture & remote-sensing data, achieving 81.43% mean classification accuracy with 89.28% mean feature reduction across 15 public datasets, outperforming multiple established baselines.
- We propose SIF-HFAM, a strong intuitionistic fuzzy hypergraph framework equipped with a monotone submodular coverage objective. The method admits the classical greedy $(1 - 1/e)$-approximation guarantee and achieves an average accuracy of approximately 78% while removing ∼98% of the features (retaining $< 2\%$) across 14 high-dimensional benchmark datasets, all with favorable computational runtime.

Generative AI CoE, IEM **November 2024 – Present**

Student Research Lead at GenAI CoE *Kolkata, India*

Led GenAI CoE's end-to-end research execution and operations—mentored and staffed projects, onboarded members via interviews, drove 10+ journal teams, managed the CoE website, and spearheaded *ReelBook*(Pearson collaboration), Medium publishing, to scale research output and AI training across 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 and automated reporting, and integrated real-time Twitter/Telegram/Reddit sentiment signals via NLP to strengthen robustness and reduce forecast error/volatility.

Innovation & Entrepreneurship Development Cell (CSE) **March 2024 – August 2024**

Undergraduate Research Assistant (Certificate) *Kolkata, India*

Co-authored an IEM-HEALS 2024 accepted paper by modeling Jul 2019–Dec 2022 closing prices of 20 pharma stocks via multivariate regression, volatility, and event-study analysis, and built *TraderBot*, a Flask+MongoDB real-time trading simulator integrated with Yahoo Finance for live strategy and portfolio testing.

National University of Singapore (NUS) **July 2023 (1 week)**

Undergraduate 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 AgriAI Workshop, 2026.

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 on Logic & AI, 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 on Logic & AI, 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), 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), 2024.

8. Rajashik Datta, **Sanjan Baitalik**, Sanket Ghosh, Saugata Ghosh, Swarnendu Ghosh, "Is Indian Financial Market Ready for Pandemics?", International Conference on Advancing Science and Technologies in Health Science (IEM-HEALS 2024), 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", 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, Computational Economics, 2025.

Skills & Activities

Programming: Python, Java, C, MATLAB
ML/AI: PyTorch, TensorFlow, Scikit-learn; Transformers
XAI: SHAP, LIME
Data: Pandas, NumPy
Tools: LaTeX, Git, Docker, Jupyter, TensorBoard
Cloud: Google Cloud (Cloud Run/Compute), AWS (S3/EC2)
Activities: GenSpark 1.0 Ideathon (Organizer; coordinated 50+ teams; shortlisted 3 funded ideas), Jun–Aug 2025; ICDC 2025 (Conference volunteer), Apr 2025; Dept. of CSE, IEM (Supported NBA accreditation documentation), Mar 2024

Projects

Geodesic optimal transport (Transfer Geometry) | GitHub

Implemented sliced-Wasserstein OT diagnostics on frozen ResNet-18 features and benchmarked across 48 transfer settings (CIFAR-10/STL-10/SVHN), showing strong correlations with zero-shot transfer (Pearson $r \approx -0.71$) and low-data adaptation speed (Spearman $\rho \approx 0.60$ at 200-shot).

Grokking + LoRA (Low-Rank Tax) | GitHub

Built controlled modular-addition experiments (p=97) comparing full-parameter vs. LoRA-on-frozen-base training for 15k epochs, reproducing classic grokking in the full MLP (99% train @ 288 → 99% val @ 481) and quantifying rank/LR thresholds (e.g., r=8 fails to fit; r=16 reaches high acc without 99% val within budget).

Volatility-Scaled AdaGrad (Online Learning) | GitHub

Implemented VS-AdaGrad, a CPU-efficient drift-aware online optimizer that scales AdaGrad using discounted residual volatility; on piecewise AR(5) non-stationary series (5 regimes, $T \approx$5000, 10 seeds), reduced regret proxy vs. AdaGrad by 18.4% (small drift) and 19.8% (medium drift) and outperformed tuned OGD by 23.7–63.8% across drift regimes.