Abrar Zahin

Webpage: zahinabrar.github.io
PhD Student, Electrical Engineering, Arizona State University

Email : azahin@asu.edu
Mobile : +1-435-764-1745

Personal Summary

Machine Learning researcher with expertise in high-dimensional statistical ML models and multimodal foundation models for computational pathology, integrating image, text, and graph-structured data. Experienced in developing scalable, interpretable ML methods for large-scale networked systems, including probabilistic graphical models and generative approaches. Strong background in computational statistics, deep learning, and large-scale inference, with hands-on experience building end-to-end ML pipelines—from data acquisition and protocol design to modeling, evaluation, and deployment—for real-world scientific and healthcare applications.

Professional Experience

- **Machine Learning Research Fellow**, Mayo Clinic
  - *Multimodal Foundation Models – Computational Pathology*: As a Machine Learning Research Fellow, I develop multimodal foundation models for computational pathology, integrating whole-slide images, clinical text, and graph-structured representations. My work focuses on designing scalable, transferable, and interpretable ML systems using deep learning and statistical machine learning for networked systems, including probabilistic graphical models. I build and evaluate end-to-end research pipelines on large-scale data, contribute to model benchmarking and representation learning, and translate research ideas into practical, high-performance ML systems for real-world clinical and scientific applications.
- **Machine Learning Research Intern**, Mayo Clinic
  - *Time-Varying Graphical Models for Longitudinal Clinical Data* : During my internship, I developed and analyzed time-varying Ising graphical models to study longitudinal neuropsychiatric symptom interactions in dementia and non-dementia cohorts. I designed a structure learning pipeline for binary clinical data that enforces sparsity and temporal smoothness to recover evolving conditional dependency networks across multiple time bins. I conducted network-level and node-level analyses to quantify structural differences between disease trajectories and controls, and interpreted the results in the context of existing clinical and medical literature. The work resulted in publication-ready algorithms, figures, and quantitative analyses aimed at understanding symptom progression in dementia.
- **Graduate Research Associate**, Arizona State University
  - *Active and Sequential Model Selection of Graphical Models*: Interconnected systems are prevalent in various real-life applications, presenting a significant research challenge: unveiling the structure from generated measurements. My research focuses on developing query and runtime-effective machine learning algorithms, employing probabilistic graphical models, to understand the structure of such systems.
- **Research Assistant**, Utah State University
  - *Big Data Management for Secured Smart Healthcare System: A Machine Learning Framework*: I developed machine learning algorithms tailored to improve smart healthcare systems within an Internet of Things (IoT) framework. The developed algorithms exhibit the capacity to effectively manage substantial volumes of data generated within IoT ecosystems.
- **Instructor**, Utah State University
  - I served as an instructor for two wireless communication-based courses at Utah State University. In this role, I taught various communication-related concepts and assisted them in building different communication-related Matlab software.
- **Data Scientist**, Energypac Engineering Limited
  - I develop and implement statistical models to analyze complex datasets and solve business problems. I collaborate with cross-functional teams to identify challenges, design experiments, and build predictive models that drive

decision-making. My work involves processing and cleaning large-scale data, creating visualizations to communicate insights, and optimizing models for performance. I leverage my expertise in programming and data analysis to deliver actionable solutions, ensuring the alignment of data-driven strategies with organizational goals.

Education

- **Arizona State University** — Tempe, Arizona
  PhD, Electrical, Computer & Energy Engineering — *Spring 2020 -Fall 2025*
  Advisor: Prof. Gautam Dasarathy, CGPA: 3.81/4.00
- **Utah State University** — Logan, Utah
  Masters of Science, Electrical & Computer Engineering — *November 2019*
  Advisor: Prof. Rose Qingyang Hu, CGPA: 3.79/4.00

Technical Skills

- **Programming Language**: Python, MATLAB, SQL, SPSS, SAS, R, and C
- **Frameworks and Libraries:**PyTorch, TensorFlow, JAX, Keras, Pandas, NumPy, Scikit, NLTK, OpenCV, Matplotlib
- **Data Visualization and Analysis**: Tableau, Microsoft PowerBI, Seaborn, GraphPad
- **High-Performance Computing**: Batch Scripting, Python Multi-Processing, GPU Clusters, Cython
- **Cloud Development**: AWS, GCP, Docker, MySQL
- **DevOps**: CI/CD (Azure DevOps, GitHub Actions, Jenkins), Git, FastAPI, Flask, REST APIs

Research Interests

Recommender Systems, Probabilistic Modeling, Graph Machine Learning, Image Processing, Generative Models, Active Learning, Reinforcement Learning, Computer Vision, Multi-armed Bandits, and Deep Learning

Publications

- **Abrar Zahin**, Rajasekhar Anguluri, Oliver Kosut, Lalitha Sankar, Gautam Dasarathy,"*Robust Model Selection of Non Tree-Structured Gaussian Graphical Models*".Transaction of Machine Learning Research (TMLR), 2025.
- **Abrar Zahin**, Wezhi Li, Gautam Dasarathy, "*Rapid Change Localization in Dynamic Graphical Models*". IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2024.
- **Abrar Zahin**, Gautam Dasarathy, "*Computationally Efficient Active Learning of Gaussian Graphical Models*". Asilomar Conference on Signals, Systems, and Computers, 2024.
- **Abrar Zahin**, Le Thanh Tan, and Rose Qingyang Hu, "*Sensor-based human activity recognition for smart healthcare: A semi-supervised machine learning*", in international conference on artificial intelligence for communications and networks, pp. 450-472. Springer, Cham, 2019.
- **Abrar Zahin**, Le Thanh Tan, and Rose Qingyang Hu, "*A Machine Learning Based Framework for the Smart Healthcare System*", 2021 in Intermountain Engineering, Technology and Computing (IETC).

  In Preparation:

  - **Abrar Zahin**, Gautam Dasarathy, "*Hypothesis Testing between Two Gaussian Graphical Models using Covariance Queries.*".
  - **Abrar Zahin**, Gautam Dasarathy, "*Sample-Query Complexity Trade-Offs in Learning and Testing Tree-Structured Graphical Models*".

Relevant Courseworks

Random Signal Theory, Information Theory, Statistical Machine Learning, Mathematical Methods for Signals and Systems, Convex Optimization, Machine Learning for High Dimensions, Reinforcement Learning, Discrete Mathematics and Numerical Analysis, Microprocessor and Assembly Language Programming, Digital Signal Processing, Radio Frequency Engineering, Telecommunication Engineering, Advanced Communication Technique, Optical Communication, Wireless & Mobile Networking

Notable Project Implementations

- **Image Denoising Convoulutional Autonencoder:** Designed and trained a convolutional **autoencoder** for image denoising, achieving a **3dB PSNR improvement** over baseline models using Gaussian noise.[Code].
- **Summarization API**: Deployed a Hugging Face summarization model as a production-grade microservice using **FastAPI** (API development), **Docker** (containerization), and **Kubernetes** (orchestration) with autoscaling, load balancing, and rolling updates for high availability. [Code] [Report]
- **Diffusion Model API:** Deployed a text-to-image **diffusion model** as a full-stack ML application using **FastAPI** for serving, **Docker** for containerization, and **Kubernetes** for scalable orchestration, integrating a web frontend, REST API, and GPU inference backend for end-to-end generative image synthesis. [Code] [Report]
- Predicted electric vehicle charging demand using a **Long Short Term Memory**, achieving a **20% reduction in RMSE** over baseline time-series models, enabling smarter grid load balancing for EV infrastructure.[Code].
- **Built an end-to-end Retrieval-Augmented Generation (RAG) pipeline** using LLaMAIndex and local embedding models for question answering over custom PDFs, and **deployed a FastAPI interface** supporting real-time queries with **chunked document parsing**, **vector indexing**, and **LLM-based response generation**. [Code].
- **GAN** for speech generation and **Transformer** for speech anonymization [Code].
- Notable Technical Reports: Link 1, Link 2, Link 3, and Link 4.

Professional Service: Reviewer

International Conference on Machine Learning (ICML), 2023; Conference on Neural Information Processing Systems (NIPS), 2023; Artificial Intelligence and Statistics (AISTATS), 2022; Association for the Advancement of Artificial Intelligence (AAAI), 2022