

Md Abdur Rahaman

1087 Westshire PL NW Apt 12, Atlanta, GA 30318, USA

☎ 505-318-5391 | ✉ aabdur.rahaman007@gmail.com | 🌐 <https://a-rahaman.github.io> | 📄 md-abdur-rahaman-a43b2a153

Education

Ph.D. Computational Science and Engineering | Expected August 2024 | Advisor: [Dr. Vince Calhoun](#)

Georgia Institute of Technology, GA, USA

M.Sc. Computer Science, July 2019

University of New Mexico, NM, USA

Research Interests

Machine Learning, Big Data, Computer Vision, Generative AI, Reinforcement Learning, Multi-modal Fusion, Pattern Mining & summarization, Biclustering, NLP, Statistical Modeling, Computational Neuroscience

Experience

Georgia Institute of Technology

Atlanta, GA

GRADUATE RESEARCH ASSOCIATE

Jun 2019 - present

My research focuses on developing computational models to extract insights from extensive datasets. I design AI/ML frameworks for self-supervised, semi-supervised, and unsupervised learning with applications in medical informatics. I developed efficient algorithms for granular factoring into big data dynamics, encompassing pattern mining, summarization, and multi-dimensional clustering. My research extends to reinforcement learning from human feedback (RLHF) and neuromorphic computing for enhancing SOTA AI.

Selected Projects

Bi-clusformer: a Transformer based end-to-end biclustering framework.

- Leveraged transformer's self-attention across feature and sample dimensions to generate coherent submatrices
- A novel edge embedding enriched with graph semantics. Can capture submodules (biomarkers) of a dynamic graph.
- A cluster-guided attention for learning edge (token) inspired by ViT with computational complexity $O(kn)$

mBAM: deep multi-modal fusion with neuromorphic design

- A multi-modal latent space fusion using spatial and modality-wise attention inspired by the 'Bottleneck Attention Module'.
- Introduced a feedback loop from multi-modal neurons to unimodal. Fuse images, electronic health records, and genomics.
- Clinically deployed a large-scale distributed AI framework for initial mental disorder screening with an accuracy of 93%

Statelet: a summarization framework for time series data

- Discovers a set of 'k' most dominant and explanatory motifs from an extensive collection of time series.
- Novel implementation of Earth Mover Distance (EMD) for motifs comparison and Kernel Density Estimator (KDE) for smoothing their frequency subspace.
- Devised a probabilistic framework for selecting the summary shapes with maximum prevalence and diversity.

SpaDE: Semantic locality preserved clustering & segmentation

- Joint optimization of sample-feature distributions using an Auto-encoder architecture for instance-feature co-clustering
- Differentiable heuristic for sparsity & semantic locality to enhance 2D segmentation.
- RLHF fine-tuning with a reward model on neurologists subgrouping preference vs. clustering results from the base model.

Generative modeling with Concept activation vector (CAV) interpretability

- A latent diffusion model for 3D brain image generation & disease prediction.
- Introspect the trained model by finding active concepts - orthogonal vectors towards learned features.
- Allows testing against pre-defined concepts. Compute CAV (inclination) towards neuro-vision concepts: brain's region of interest (ROI), activation, and connectivity patterns.

A robust graph neural network (GNN) for modeling Biological Networks

- Model biological networks by instantiating their components as nodes and interaction/causal inferences as the edges.
- Multi-headed self-attention to learn node embedding and an orthonormal readout for graph-level representation.
- Improve performance on downstream tasks (e.g., segmentation, classification) with insights into the underlying system.

NOKIA BELL LABS

DATA SCIENCE RESEARCH INTERN

Murray Hill, NJ

Sep 2021 - Dec 2021

Project Title: Routing failure tickets to the corresponding service team

I collaborated with the log analytics group on developing a summarizer to convert piles (billions of lines) of machine logs into a manageable and learnable representation. Instantiated a BERT model to learn user-provided ticket descriptions and a multi-modal framework to combine machine logs. We deployed the framework for routing the IT tickets and it achieved 25.7 % improvement over human performance characterizing the tickets.

University of New Mexico

RESEARCH ASSISTANT

Albuquerque, NM

June 2017 - Apr 2019

I've worked on data mining models for multivariate time series & developed a time series motif detection and a summarization framework. Also, designed exhaustive-search-based solutions for biclustering and tri-clustering algorithms.

Selected Projects

N-BiC: greedy biclustering algorithm relaxing a specific model order (k)

- Constraint depth-first search (DFS) - based algorithm to semi-exhaustively explore all possible combinations of instances.
- Discover biclusters of co-expressed samples and features in the dataset without specifying 'k'.

IBRNN: Information-theoretic introspection for AI Interpretability

- Explore the theoretical upper/lower bound of information compression in RNN layers.
- CBOW for word2vec embedding of the text corpus and bi-LSTM for the downstream task.
- Inspired by information Bottleneck theory, approximate mutual information to quantify feature compression.

Department of Computer Science, University of New Mexico

TEACHING ASSISTANT

Albuquerque, NM

Aug 2016 - May 2017

Assisted with Linear algebra, Declarative Programming, and Computer Algorithms courses for CS undergrad students. I hold TA Office hours to help students with their assignments, grading, and tutorials on Haskell, Scheme, and GNU Emacs.

Skills

Programming Python, CUDA, C/C++, Scala, JAVA, C#, JavaScript, JQuery

Cloud Technologies & DB AWS, Google Cloud, Flask, Hadoop, Spark, GCP, Hive, MySQL, BigQuery, MongoDB, PostgreSQL

Libraries PyTorch, TensorFlow, OpenCV, Ray, Stanford CoreNLP, Pandas, NLTK, Scikit-Learn, PySpark, Hugging Face

Tools MATLAB, R, MLOps, FMOps, XGBoost, Docker, DeepSpeed, Slurm, SPM, Git, Heroku

Selected Publications

- Md Abdur Rahaman, Zening Fu, Armin Iraj and V. D. Calhoun, 2024, "A Deep Biclustering Framework for Brain Network Analysis". Accepted. In **CVPR Workshop on Domain adaptation, Explainability, Fairness in AI for Medical Image Analysis**.
- Rahaman, Md Abdur, Yash Garg, Armin Iraj, Zening Fu, Jiayu Chen, and Vince Calhoun. 2022. "Two-Dimensional Attentive Fusion for Multi-Modal Learning of Neuroimaging and Genomics Data." In **2022 IEEE 32nd International Workshop on Machine Learning for Signal Processing (MLSP)**.
- Baker, Bradley Thomas, Noah Lewis, Debratta Saha, Md Abdur Rahaman, Sergey Plis, and Vince Calhoun. "Information Bottleneck for Multi-Task LSTMs." In **NeurIPS 2022 Workshop on Information-Theoretic Principles in Cognitive Systems**.
- Dolci, G., Rahaman, M. A., Galazzo, I. B., Cruciani, F., Abrol, A., Chen, J., ... & Calhoun, V. D. (2023, June). "Deep Generative Transfer Learning Predicts Conversion To Alzheimer's Disease From Neuroimaging Genomics Data". In **2023 IEEE International Conference on Acoustics, Speech, and Signal Processing Workshops (ICASSPW)**
- M. A. Rahaman, E. Damaraju, D. K. Saha, V. D. Calhoun and S. M. Plis, "Statelets: A Novel Multi-Dimensional State-Shape Representation Of Brain Functional Connectivity Dynamics". **2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)**.
- Rahaman, Md Abdur, Jessica A Turner, Cota Navin Gupta, Srinivas Rachakonda, Jiayu Chen, Jingyu Liu, Theo GM Van Erp, Steven Potkin, Judith Ford, and Daniel Mathalon. 2019. "N-BiC: a novel biclustering algorithm for brain and behavioral features without specifying the model order". **IEEE Transactions on Biomedical Engineering (TBME)**.
- Rahaman, Md Abdur, Jiayu Chen, Zening Fu, Noah Lewis, Armin Iraj, Theo GM van Erp, and Vince D Calhoun. 2023. "Deep multi-modal predictome for studying mental disorders". **Human Brain Mapping**.
- M. A. Rahaman, J. Chen, Z. Fu, N. Lewis, A. Iraj and V. D. Calhoun, "Multi-modal deep learning of functional and structural neuroimaging and genomic data to predict mental illness," **2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)**.
- Dolci, Giorgio, Md Abdur Rahaman, Jiayu Chen, Kuaikuai Duan, Zening Fu, Anees Abrol, Gloria Menegaz, and Vince D Calhoun. 2022. "A deep generative multimodal imaging genomics framework for Alzheimer's disease prediction." In **2022 IEEE 22nd International Conference on Bioinformatics and Bioengineering (BIBE)**.
- Du, Yuhui, Zening Fu, Jing Sui, Shuang Gao, Ying Xing, Dongdong Lin, Mustafa Salman, Anees Abrol, Md Abdur Rahaman, and Jiayu Chen. 2020. "NeuroMark: An automated and adaptive ICA based pipeline to identify reproducible fMRI markers of brain disorders". **NeuroImage**.