DEBBRATA KUMAR SAHA

Atlanta, GA-30318; 505-903-3810; debbratakumar.saha@gmail.com; https://www.linkedin.com/in/debbratasaha/

CAREER SUMMARY

Ph.D. candidate with 7+ years hands-on Research & **Development experience** - skilled in machine learning, deep learning, natural language processing, computer vision, neuroimaging and software development.

- Developed a first-ever federated algorithm for identifying the bad samples/outliers from the distributed datasets
- Implemented two neuroimaging machine learning tools and deployed them into the existing production system
- Lead researcher in projects involving data collection, preprocessing, modeling, and method developments
- Strong publication record evidenced by thirteen published articles (4 journals, 9 conferences)

EDUCATION

Georgia Institute of Technology, Atlanta, Georgia May 2019 - Present
Ph.D. in Computational Science & Engineering, GPA : 3.83 / 4.00

University of New Mexico, NM, USA May 2019
MS in Computer Science, GPA : 3.61 / 4.00

TECHNICAL SKILLS

Programming	Python, Java, R, Matlab, C/C++, SQL, Oracle
Frameworks and tools	PyTorch, TensorFlow, Keras, HuggingFace, Lightning AI, Scikit-learn, Pandas, Plotly Seaborn, Tableau, OpenCV, NLTK, GenSim, CoreNLP, SpaCy, BERT, Transformer LangChain, Kubernetes, Docker, Ggplot, Matplotlib, PySpark, CUDA, AWS, Git

WORK EXPERIENCE

Graduate Research Assistant, **TReNDS center, Georgia Tech** May 2019 - Present
Privacy-preserving quality control tool

- Developed a federated algorithm to detect the bad scans (outliers) from the multi-site time series data
- Added differential privacy features for formal privacy guarantees and incorporated the automation features
- Incorporated into the current production system and available to users worldwide for utilization

Decentralized brain component mapping

- Introduced the first-ever approach for multivariate brain structure mapping from the distributed sMRI dataset
- Extracts significant group differences between healthy control and Schizophrenia patients from different data sources
- Developed and integrated a machine learning tool into the production system to identify significant brain components

Multi-model deep neural network

- Developed a CNN and LSTM-based multi-model classification framework
- Analyzed the predictability of the fMRI and sMRI data in the classification using large time series patients datasets
- Applied summarization and attention mechanism and improved the accuracy by 5% compared to conventional methods

Graduate Research Assistant, **The Mind Research Network, University of New Mexico** Aug. 2016-May 2019
Model order selection of the K-means algorithm

- Introduced a novel classification-based approach to identify the optimal model order for the K-means algorithm
- Obtained 6% higher accuracy compared with that of the traditional methods to select the optimal k in k-means
- Overcomes the complexity of the traditional model order selection methods (Elbow, Gap statistic, Silhouette, etc.)

Lecturer, **State University of Bangladesh, Bangladesh** April 2013 - July 2015

- Instructed undergrad courses (Data structure & Algorithm, Database Management, Programming Language etc.)
- Designed lab, exams, and advised students for different projects ; Organized programming workshops, and seminars

RELEVANT CLASS PROJECTS

- Created a vehicle and lane detection system by analyzing video frames. Used **OpenCV** and **Cascade** classifiers
- Developed a method for illustrating unique behavior during training, such as feature compression over sequences within recurrent layer in RNN; Used IMDB and Rotten Tomatoes datasets
- Developed a Software for inventory management to track and manage items through various stages along the supply chain; Language : **C#** ; Database : **Oracle**
- Built a spam classifier utilizing **Support Vector Machine (SVM)** and **Grading Boosting** techniques. Extracted features from real-time text data using the term frequency-inverse document frequency (tf-idf) approach
- Developed a content and collaborative filtering-based recommender system using **XGBoost**. Applied matrix factorization and feature engineering for extracting significant features

CONFERENCE PUBLICATIONS

- **Debbrata K. Saha**, Vince D. Calhoun, Sandeep R. Panta, Sergey M. Plis, *See without looking: joint visualization of sensitive multi-site datasets*, In *International Joint Conferences on Artificial Intelligence Organization (**IJCAI**)*, pages 2672–2678, 2017
- **Debbrata K. Saha**, Anees Abrol, Eswar Damaraju, Barnaly Rashid, Sergey M. Plis, Vince D. Calhoun, *Classification As a Criterion to Select Model Order For Dynamic Functional Connectivity States in Rest-fMRI Data*, *IEEE International Symposium on Biomedical Imaging(**ISBI**)*, pp. 1602-1605, 2019
- 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*, ***ISBI***, 2021, pp. 1822-1826
- **D. K. Saha,** R. F. Silva, B. T. Baker and V. D. Calhoun, "Decentralized Spatially Constrained Source-Based Morphometry," ***ISBI***, pp. 1-5, 2022
- R. Saha, **D. K. Saha**, M. Abdur Rahaman, Z. Fu and V. D. Calhoun, "Longitudinal Whole-Brain Functional Network Change Patterns Over A Two-Year Period In The ABCD Data," ***ISBI***, pp. 1-4, 2022
- Baker BT, Lewis N, **Saha Debbrata**, Rahaman MA, Plis S, Calhoun V., *Information Bottleneck for Multi-Task LSTMs*, in ***InfoCog, NeurIPS***, 2022
- **D. K. Saha**, Anastasia Boshali, Rekha Saha, Ihab Hajjar, Vince D. Calhoun, *A Multivariate Method for Estimating and comparing whole brain functional connectomes from fMRI and PET data*, in *the IEEE Engineering in Medicine and Biology Society (**EMBC**)*, 2023
- Rekha Saha, **Debbrata K. Saha**, Zening Fu, Rogers Silva, Vince D. Calhoun, *Functional and Structural Longitudinal Change Patterns in Adolescent Brain*, in ***EMBC***, 2023
- **Debbrata K. Saha**, Vince D. Calhoun, Soo Min Kwon, Anand D. Sarwate, Rekha Saha, Sergey M. Plis,, *Federated, Fast, and Private Visualization of Decentralized Data*, in ***FL, ICML***, 2023

JOURNAL PUBLICATIONS

- Harshvardhan Gazula, Ross Kelly, Javier Romero, Eric Verner, Bradley T. Baker, Rogers F. Silva, Hafiz Imtiaz, **Debbrata Kumar Saha**, Rajikha Raja, Jessica A. Turner, Anand D. Sarwate, Sergey M. Plis, Vince D. Calhoun, *COINSTAC: Collaborative Informatics and Neuroimaging Suite Toolkit for Anonymous Computation*, in proc. of *Journal of Open Source Software (**JOSS**)*, 2020
- **Debbrata K. Saha**, Anees Abrol, Eswar Damaraju, Barnaly Rashid, Sergey M. Plis, Vince D. Calhoun, *A Classification-Based Approach to Estimate the Number of Resting Functional Magnetic Resonance Imaging Dynamic Functional Connectivity States*, ***Brain Connectivity***, vol. 11, no. 2, pp. 132–145, 2021.
- **Debbrata K. Saha**,Vince D. Calhoun,Yuhui Du,Zening Fu,Soo Min Kwon,Anand D. Sarwate,Sandeep R. Panta,Sergey M. Plis, *Privacy-preserving quality control of neuroimaging datasets in federated environments*, ***Human Brain Mapping***, 43(7), 2289– 2310, 2022
- M. A. Rahaman, E. Damaraju, **D. K. Saha**, S. M. Plis and V. D. Calhoun and , *Statelets: Capturing recurrent transient variations in dynamic functional network connectivity*, ***Human Brain Mapping***, 43(8), 2503– 2518, 2022