

SAYAN GHOSAL

Baltimore * sghosal3@jhu.edu * <https://sayangsep.github.io> * LinkedIn * 4435315268

PROFESSIONAL SUMMARY

Data science researcher with 6+ years of experience in complex, high dimensional, and structured data analysis with an focus on computational biology and genomics.

Contributions include novel models and training strategies to adopt statistical and machine learning models for analyzing biological data.

Developed and built a collaborative network with scientists across the globe, and engaged in multiple collaborative projects which led to 7 publications in peer-reviewed conferences and journals.

Passionate about interdisciplinary research to merge the gap between ML and the scientific domain.

EDUCATION

Johns Hopkins University, Baltimore, USA

Ph.D., Electrical and Computer Engineering 2023

M.S., Applied Mathematics and Statistics 2021

Jadavpur University, Kolkata, INDIA

B.E., Electronics and Telecommunication Engineering 2017

SKILLS

ML for Genomics	Knowledge Graph Embedding, Pathway Analysis Using Geometric Deep Learning, Gene Interactions Using Graph Convolutions
Statistical Genetics	Mixed Linear Modeling, PRS Analysis, Finemapping
Deep Learning	Contrastive Learning on Graphs, Transfer Learning, Transformers, Autoencoders, Multiview Representation Learning
Applied Statistics	Deep Bayes Variational Models, Dictionary Learning, Linear and Non-linear Dimensionality Reduction Techniques
Model Interpretability	Bayesian Feature Selection, Graph Attention

RELEVANT EXPERIENCE

Johns Hopkins University, Baltimore

2017-Present

Research Assistant

BEATRICE: Bayesian Fine-mapping from Summary Data using Deep Variational Inference

- Developed a deep Bayes variational approach to find causal genetic variants from GWAS leading to robust credible sets with 2.2 fold increase in power and coverage.
- Utilized a deep neural network to optimize the parameters in < 50 seconds while handling multiple causal variants and increasing noise as determined by the polygenecity of the trait.

GUIDE: A Biologically Interpretable Graph Convolutional Network to Embed Genetic Architecture for Disease Phenotype Prediction and Pathway Discovery

- Developed a novel geometric deep learning tool to embed a genomic knowledge graph for disease prediction, leading to an 15% increase in AUC in schizophrenia risk prediction.
- Introduce explainability by using hierarchical graph convolution and attention operations and identified 152 significant pathways among 24K pathways associated with disorders.
- Fused multimodal imaging data with genetic data which further improved AUC by 10%.

G-MIND: An End-to-End Multimodal Imaging-Genetics Autoencoder Framework for Biomarker Identification and Disease Classification

- Developed a novel autoencoder to identify correlated brain and genetic networks from brain imaging and genetics study of schizophrenia.
- Introduced a novel model interpretation strategy leading to identification of implicated regions like dorsolateral prefrontal cortex and genes like GABRA, and RSG6.

Bridging Imaging, Genetics, and Diagnosis in a Coupled Low-Dimensional Framework

- Designed a novel multivariate linear generative model guided by graph Laplacian of biological interconnectedness and disease status.
- Performed in-depth study of hyper and hypo-active brain regions, and their correlated genetic components, leading to finding default mode brain network and correlated genes like LINC00599.

Johns Hopkins University, Baltimore
Supervisor

2021- Present

- Supervising a CS graduate student on deep learning projects aimed to learn the longitudinal effect of genetic variations on morphological changes in brain regions of 1K Alzheimer's patients.
- Currently, published a senior author paper at the International Conference of the IEEE Engineering in Medicine and Biology Society.

INTERNSHIP PROJECTS

X, The Moonshot Factory (Google X), Mountain View
AI Resident

2022

- Conceptualized, and successfully combined statistical models with language model to integrate time-series phenotype and genetics variation data to develop tools for sustainable agriculture.
- Parallely deployed models on high performance computing systems for quick prototyping and handling multiple terabytes of data.
- Communicated the biological findings and the ML models to a group of 30 people from multidisciplinary fields of data science, biology, and software engineering.

Siemens Healthineers, Princeton
Machine Learning Intern

2021

- Built a graph-based deep neural network to integrate multimodal data for multiclass risk prediction of cardiovascular disease.
- Implemented a graph based contrastive learning to utilize historical medical records of the population to perform prediction on new subjects.

HONORS AND AWARDS

Organization for Human Brain Mapping awarded \$700 for noteworthy abstracts.	2023
MINDS fellowship awarding \$30K for spring tuition.	2022
Best Paper Award, SPIE Medical Imaging (Image Processing Conference)	2021
MICCAI travel award of \$500.	2020
Dept. of Electrical and Computer Engineering, JHU, PhD fellowship	2017-2018
Mitacs Globalink Research Fellowship Award	2016

PUBLICATIONS

S. Ghosal, *et al.*, *BEATRICE: Bayesian Fine-mapping from Summary Data using Deep Variational Inference*.(Submitted in **PLOS Genetics**). biorXiv

S. Wu, A. Venkataraman, **S. Ghosal**. *GIRUS-net: A Multimodal Deep Learning Model Identifying Imaging and Genetic Biomarkers Linked to Alzheimer's Disease Severity*. Accepted in **EMBC**, 2023.

S. Ghosal, et al. *A Biologically Interpretable Graph Convolutional Network to Link Genetic Risk Pathways and Neuroimaging Markers of Disease.* **ICLR: International Conference on Learning Representations**, 2022 (**Accepted**). [biorXiv](#)

S. Ghosal, et al. *A Generative Discriminative Framework that Integrates Imaging, Genetic, and Diagnosis into Coupled Low Dimensional Space.* **NeuroImage**: 238:118200, 2021

S. Ghosal, et al. *G-MIND: An End-to-End Multimodal Imaging-Genetics Framework for Biomarker Identification and Disease Classification.* Proc. **SPIE**, Medical Imaging 2021: Image Processing. [arXiv:2101.11656](#)

Selected for Special Oral Presentation (<15% of Papers), and received best student paper award

S. Ghosal, et al. *Bridging Imaging, Genetics, and Diagnosis in a Coupled Low-dimensional Framework.* **MICCAI: Medical Image Computing and Computer Assisted Intervention**, 2019. **Selected for Early Acceptance (Top 18% of Submissions)**

S. Ghosal, et al. *A generative-predictive framework to capture altered brain activity in fMRI and its association with genetic risk: application to Schizophrenia.* Proc. **SPIE** 10949, Medical Imaging 2019: Image Processing.

S. Ghosal, Nilanjan Ray. *Deep deformable registration: Enhancing accuracy by fully convolutional neural net.* **Pattern Recognition Letters**.

S. Ghosal, et al. *A novel non-rigid registration algorithm for zebrafish larval images.* 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (**EMBC**), 2017.

INVITED SEMINARS AND TALKS

Title: Deep Imaging Genetics to Parse Neuropsychiatric Disorders

Regeneron (Host: Yu Bai)

2023

Google-Genomics Group (Host: Farhad Hormozdiari)

2022

Title: Biologically Inspired Regularization Models Integrating Multimodal Data to Parse Neuropsychiatric Disorders.

ECE Seminar Series (Host: Archana Venkataraman)

2022