Sayan Ghosal

Baltimore ✳ sghosal3@jhu.edu ✳ https://sayangsep.github.io ✳ LinkedIn ✳ 443-531-5268

Professional Summary

- ML researcher with 7+ years of experience in developing novel deep learning models to decode complex, high dimensional, and structured data with a focus in computational biology and genomics.
- Contributions include novel models and training strategies to adopt statistical and machine learning models for integrating multimodal biological data for disease prediction and biomarker discovery.
- 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.
- Enthusiastic about bridging the gap between machine learning and scientific domains through interdisciplinary research and collaborations.

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

Foundational Models	Transformer for Long Sequences, Attention Pooling, Handling Incosistent Sequence Dropout
Dimentionality Reduction	Latent Factor Models, Dictionary Learning, t-SNE, UMAP, Spectral Clustering, K-Means
ML for Genomics	Convolutional Networks for DNA-seq Data, Gene-Pathway Interaction Based GNN, Knowledge Graph Embedding
Deep Learning	Contrastive Learning on Graphs, Transfer Learning, Transformers, Autoencoders, Graph Neural Networks
Multimodal Learning	Multimodal CCA, Variational Autoencoders, Contrastive Learning
Model Iterpretibility	Bayesian Feature Selection, Attention Mechanisms, LASSO

Internship Projects

X, The Moonshot Factory (Google X), Mountain View — *2022*
AI Resident

- Conceptualized, designed, and successfully combined statistical models with LLMs to integrate time-series data and genomic data to develop tools for sustainable agriculture.
- Parallelly 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 — *2021*
Machine Learning Intern

- Developed a graph-based deep neural network to explain the cardiovascular disease status of a subject based on the disease status and observed features obtained from the population.

- Explained graph attention through the lenses of linear models to identify patterns leading to improved prediction on new subjects.
- Integrated patient-patient similarity in a graph convolutional framework to capture population level characteristics of the disease.
- Implemented a graph based contrastive learning to utilize historical medical records of the population to perform prediction on new subjects.

Relevant experience

Johns Hopkins University, Baltimore   *2017-Present*
Research Assistant, Electrical and Computer Engineering

BEATRICE: Bayesian Feature Selection from Summary Data using Deep Variational Inference

- Developed a first-ever 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 and infinitesimal effects from non-causal variants.

GUIDE: A Biologically Interpretable Graph Convolutional Network to Integrate Multimodal Genetic Data for Disease Prediction and Pathway Discovery

- Developed the first geometric deep learning tool capable of capturing complex relationships between genes and encoding whole genome genotype data for disease prediction, leading to an 15% increase in AUC in schizophrenia risk prediction.
- Introduced explainability by using hierarchical graph convolution and attention operations and identified 152 clinically relevant pathways associated with schizophrenia.
- Collaborated with cross-functional teams of biologists, data scientists, and clinicians, which led to a future million-dollar grant, scholarships[2], awards[1], and two publications.

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

- Designed a multimodal CCA based framework integrating multimodal data in a latent space while finding clinically relevant regulatory networks.
- Introduced graph Laplacian of biological interconnectedness and disease status to regularize the framework and find clinically relevant biomarkers.
- 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.

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.
- Handled missing data modalities via dropping out autoencoder branches.
- Introduced a novel model interpretation strategy leading to identification of implicated regions like dorsolateral prefrontal cortex and genes like GABRA, and RSG6.
- Received special mention in the Hopkins magazine and a best paper award[3] at SPIE.

Johns Hopkins University, Baltimore   *2021- Present*
Supervisor
- Advising a CS graduate student on building large language model that can handle inconsistent dropout events in time series data.
- Capturing longitudinal effect of genetic variations on morphological changes in brain regions of Alzheimer's patients.

- Currently, authored a senior author paper at the International Conference of the IEEE Engineering in Medicine and Biology Society.

Honors And Awards

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

Patents

Ghosal, S., Jacob, A. J., Sharma, P., & Gulsun, M. A. (2023). Subpopulation Based Patient Risk Prediction Using Graph Attention Networks. US Patent App. 17/647,613.

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: Benefits of Deep Learning to Parse Complex Genetic Architectures to Provide Mechanistic Insights

MIT (Host: Manolis Kellis) 2023

Title: Deep Imaging Genetics to Parse Neuropsychiatric Disorders

Regeneron (Host: Yu Bai) 2023

Google-Genomics, Google Health (Host: Farhad Hormozdiari) 2022

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

ECE Seminar Series (Host: Archana Venkataraman) 2022