JACOB LEIBY

jleiby@pennmedicine.upenn.edu · (484) 366-3060 · leibyj.github.io

SUMMARY

- AI researcher and biomedical data scientist with 5+ years of experience developing deep learning models for medical image analysis and multimodal data integration
- Expertise in self-supervised learning and large-scale deep learning architectures for pathology and radiology applications
- Strong publication record (7+ peer-reviewed papers, 4 first-author) and extensive experience collaborating with clinicians, engineers, and cross-functional teams

EDUCATION

University of Pennsylvania Philadelphia, PA
Ph.D., Genomics and Computational Biology, Biomedical Informatics track June 2024

Cornell University Ithaca, NY
B.S., Natural Resources May 2016
  Magna Cum Laude

RESEARCH EXPERIENCE

University of Pennsylvania Philadelphia, PA
Postdoctoral Researcher, Department of Pathology and Laboratory Medicine August 2024 – Present
Supervised by Zhi Huang, Ph.D. and Dokyoon Kim, Ph.D.

- Curate large-scale, heterogeneous datasets for multimodal histology imaging applications
- Develop a self-supervised foundation model framework to learn generalizable and transferable representations of multiplexed spatial proteomics data across diverse study designs
- Fine-tune a multimodal vision-language model for joint embedding alignment of multiplexed spatial proteomics and H&E imaging data to facilitate integrative analyses
- Trained a multi-task, multiple-instance learning vision encoder for automated immunohistochemistry (IHC) assessment on over 10 million images and integrated it into histology software for research and clinical use

University of Pennsylvania Philadelphia, PA
Ph.D., Department of Biostatistics, Epidemiology, and Informatics June 2019 – June 2024
Supervised by Dokyoon Kim, Ph.D.

- Developed multi-task deep learning models for organ segmentation and disease classification from 3D abdominal CT scans and predicted future cardiometabolic diseases
- Integrated supervised and self-supervised learning to predict genomic aberrations from histology whole-slide images
- Analyzed genomic and clinical data for prognostic analysis of cancer patients and biological age modeling of a cohort of over 30,000 individuals
- Initiated collaborations with a group of international physicians to develop study plans and formulate clinical hypotheses
- Trained and supervised undergraduate and graduate students in model development and predictive analyses

Genentech South San Francisco, CA
Machine Learning Engineer (Contractor, Part-time), Research and Early Clinical Development March – May 2024

- Investigated large language models and multimodal data integration techniques to combine radiology reports and clinical imaging data to enhance predictive modeling for biomarker quantification
- Collaborated with engineers to launch and maintain digital biomarker applications for internal research and external clinical applications
- Coordinated efforts with academic partners to align data labeling protocols with existing methods, supporting research planning and method development

Genentech South San Francisco, CA
Clinical AI Intern, Research and Early Clinical Development September 2023 – March 2024

- Developed computer vision models for biomarker quantification from chest radiographs for use in patient selection and clinical trial endpoints
- Analyzed clinical trial data to extract relevant features and performed time-to-event analysis
- Presented method development and analysis to interdisciplinary teams and stakeholders

University of Pennsylvania Philadelphia, PA

Research Specialist, Department of Microbiology February 2017 – May 2019

Supervised by Rick Bushman, Ph.D.

- Generated high-throughput sequencing libraries for microbiome and gene therapy experiments
- Led an independent project spanning DNA extraction, library prep and sequencing, data analysis, and manuscript preparation to investigate the absence of a detectable placental microbiome

SELECT PUBLICATIONS

Leiby, J. S.; Lee, M. E.; Shivakumar, M.; Choe, E. K.; Kim, D., Deep learning imaging phenotype can classify metabolic syndrome and is predictive of cardiometabolic disorders, *Journal of Translational Medicine*, 2024

Leiby, J. S.; Lee, M. E.; Choe, E. K.; Kim, D., Data-driven classification of fatty liver from 3D unenhanced abdominal CT scans, *44th International Workshop on Machine Learning in Medical Imaging (MLMI)*, 2023

Leiby, J. S.; Hao, J.; Kang, G. H.; Park, J. W.; Kim, D.; Attention-based multiple-instance learning with self-supervision to predict microsatellite instability in colorectal cancer from histology whole-slide images, *44th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2022

Leiby, J. S.; McCormick, K.; Sherrill-Mix, S.; Clarke, E. L.; Kessler, L. R.; Taylor, L. J.; Hofstaeder, C. E.; Roche, A. M.; Mattei, L. M.; Bittinger, K.; Elovitz, M. A.; Leite, R.; Parry, S.; Bushman, F. D., Lack of detection of a human placenta microbiome in samples from preterm and term deliveries, *Microbiome*, 2018

SELECT CONFERENCE PRESENTATIONS

Poster presentation, Medical Image Computing and Computer Assisted Intervention Society (MICCAI), "Data-driven classification of fatty liver from 3D unenhanced abdominal CT scans," 2023

Poster presentation, American Society for Human Genetics (ASHG), "A combined polygenic risk score and CT-derived biomarker for metabolic syndrome," 2022

Oral presentation, IEEE Engineering in Medicine and Biology Society (EMBC), "Attention-based multiple-instance learning with self-supervision to predict microsatellite instability in colorectal cancer from histology whole-slide images," 2022

Oral presentation, Mid-Atlantic Bioinformatics Conference, "A combined polygenic risk score for metabolic syndrome," 2022

TECHNICAL SKILLS

Programming Languages and Software: Python (PyTorch, Scikit-learn), R (Tidyverse, caret), Linux, git version control, Hydra, HPC and distributed computing, AWS, SageMaker, ClearML, SQL

Data Analysis: Machine learning, deep learning, computer vision, data manipulation and visualization, predictive modeling, time-to-event analysis