Daniel L. Zeiberg

(609) 410-5967 | zeiberg.d@northeastern.edu | linkedin.com/in/danielzeiberg

SUMMARY

Machine learning expert with 8+ years of experience in addressing complex and nuanced computational biology challenges. Specialized in training and deploying calibrated machine learning models excelled in managing complex and biased datasets to elucidate the functional and phenotypic effects of genetic variants. Passionate and eager to leverage expertise to tackle intricate problems in computational biology and medicine.

PROFESSIONAL EXPERIENCE

Northeastern University, Boston, MA

PhD Candidate, Computer Science 09/2018 - Present

- Published fast positive-unlabeled class prior estimation algorithm achieving 40% performance improvement with Python and Matlab open-source implementations, used in developing clinical variant classification guidelines
- Developed statistical bias models to better capture complex real-world bias, leading to 12% improvement in classification performance
- Trained protein language model to associate genetic variants with rare diseases, improving classification accuracy by 17%.
- Led a machine learning and computational biology bootcamp teaching 21 undergraduate and high-school students the principles in training machine learning models to classify genetic variants
- Implemented an end-to-end pipeline using sequence-based variant classification models and gene-phenotype association networks to identify the genetic variant causing rare diseases in 30 families
- Trained deep-learning-based sequence-to-sequence models using PyTorch and TensorFlow to forecast spatiotemporal data

University of Michigan, Ann Arbor, MI

Undergraduate Researcher 05/2017 - 07/2018

- Devised a state-of-the-art machine learning model to identify hospital patients at risk for developing acute respiratory distress syndrome

Comcast, Mount Laurel, NJ

Software Defined Core Network Engineering Intern 06/2016 - 08/2016

- Developed a dashboard using Python and JavaScript to track Comcast's network health and monitor outages used widely throughout the network engineering division

Comcast, Philadelphia, PA

Engineering Analysis Intern 05/2015 - 07/2015

- Created On-Demand network flow visualizations with D3.js, used by business operations personnel in network planning

TECHNICAL STRENGTHS

Languages:	Python, Matlab, C, C++
Tools & Frameworks:	TensorFlow, PyTorch, scikit-learn, high-throughput computing, MySQL
Computational Biology Resources:	ClinVar, gnomAD, dbNSFP, Ensembl VEP, BLAST

EDUCATION

Northeastern University, Boston, MA 09/2018 - 09/2024

PhD Candidate, Computer Science

Thesis: Learning Calibrated Classifiers from Nonrepresentative Data

University of Michigan, Ann Arbor, MI 08/2014 - 04/2018

Bachelor of Science, Computer Science | Minor, Statistics

- *Summa cum laude,* University of Michigan Engineering Honors College
- Engineering honors societies: *Tau Beta Pi, Eta Kappa Nu*
- Awards: *Most Likely to Have Transformative Scientific Impact* - Michigan Institute for Data Science

PUBLICATIONS

- IGVF Consortium. Deciphering the impact of genomic variation on function. *Nature* 633, 47–57 (2024).
- Jain, S., Trinidad, M., Nguyen, T. B., Jones, K., Diaz Neto, S., Ge, F., ... & Clark, W. T. (2024). Evaluation of enzyme activity predictions for variants of unknown significance in Arylsulfatase A. bioRxiv, 2024-05.

- Stenton, S. L., O'Leary, M., Lemire, G., VanNoy, G. E., DiTroia, S., Ganesh, V. S., ... & O'Donnell-Luria, A. (2023). Critical assessment of variant prioritization methods for rare disease diagnosis within the Rare Genomes Project. *medRxiv*.
- **Zeiberg, D.**, Jain, S., & Radivojac, P. (2023, June). Leveraging structure for improved classification of grouped biased data. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 37, No. 9, pp. 11113-11120).
- Chen, Y., Jain, S., **Zeiberg, D.**, Iakoucheva, L. M., Mooney, S. D., Radivojac, P., & Pejaver, V. (2022). Multi-objective prioritization of genes for high-throughput functional assays towards improved clinical variant classification. In *Pacific Symposium on Biocomputing 2023: Kohala Coast, Hawaii, USA, 3–7 January 2023* (pp. 323-334).
- Lugo-Martinez, J., **Zeiberg, D**., Gaudelet, T., Malod-Dognin, N., Przulj, N., & Radivojac, P. (2021). Classification in biological networks with hypergraphlet kernels. *Bioinformatics*, *37*(7), 1000-1007.
- **Zeiberg, D.**, Jain, S., & Radivojac, P. (2020, April). Fast nonparametric estimation of class proportions in the positive-unlabeled classification setting. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 34, No. 04, pp. 6729-6736).
- **Zeiberg, D.**, Prahlad, T., Nallamothu, B. K., Iwashyna, T. J., Wiens, J., & Sjoding, M. W. (2019). Machine learning for patient risk stratification for acute respiratory distress syndrome. *PloS one*, *14*(3), e0214465.