Balagopal (Balu) Unnikrishnan

Ph.D. Candidate, Computer Science
University of Toronto

balu@cs.toronto.edu
+1 (437) 991-1394
www.balagopal.me
GitHub

SUMMARY / RESEARCH INTERESTS

PhD Researcher focused on model generalization via mitigation of shortcut learning and spurious correlations in vision and language models. Skilled in developing robust AI systems for healthcare with prior industry experience in AI research, strong publication record and experience working with multi-disciplinary teams.

EDUCATION

University of Toronto, Toronto, Canada — 09/21 - 09/26 (Expected)
Ph.D. in Computer Science | *Advisors: Dr. Michael Brudno & Dr. Chris McIntosh* — CGPA: 4.0/4.0
Research supported by Vector Institute, Schwartz Reisman Institute, UHN & SickKids Hospital

National University of Singapore, Singapore — 2019
Masters in Intelligent Systems | *Advisors: Dr. Matthew Chua & Dr. Xulei Yang* — CGPA: 4.16/5.0
Research supported by Agency for Science, Technology & Research (A*STAR) and Ministry of Education, Singapore

RESEARCH EXPERIENCE

Ph.D. Candidate — 2021 - Present
University of Toronto
Schwartz Reisman Institute (SRI) & Vector Institute Graduate Fellow

- Developed novel shortcut learning detection framework analyzing 750K+ samples across 13 datasets, predicting out-of-distribution performance degradation with 96% accuracy (Published - Nature npj Digital Medicine)
- Created an attention-based mechanism for localizing and correcting for multiple shortcuts (spatial and spectral), surpasses SOTA by 7.5% in data with multiple spurious correlations occurring simultaneously (Under Review - CVPR 2025)
- Designed generative in-painting systems for confounder reduction in medical imaging tasks using diffusion models - identified and corrected for hidden stratification, improving diagnostic performance by 20%
- Built and deployed multiple clinical AI tools for radiology & ultrasound, implementing validation protocols and coordinating with interdisciplinary teams in hospital environments

AI Research Engineer — 2019 - 2021
*Institute for Infocomm Research (I2R), Agency for Science Technology & Research (A*STAR), Singapore*
Top 1% performer in the organization

- Designed NoTeacher, a novel semi-supervised learning framework reducing annotation requirements by 95% while maintaining 90% of fully-supervised performance across multiple imaging modalities (MICCAI Best Paper Award Runner-up)
- Co-developed multi-scale self-supervision technique for gigapixel pathology images, achieving SOTA performance (0.92 AUC) with only 10% labeled data
- Established cross-functional collaborations between research and clinical teams for successful translation of algorithmic advances
- Created IP "Semi-Supervised Process to Guide Annotation for Image Classification Tasks" – successfully licensed for commercial application

TECHNICAL SKILLS

- **AI & Deep Learning:** PyTorch, TensorFlow, algorithm development, generative models (GANs, diffusion), vision-language models for healthcare, domain shift, confounder mitigation, semi/self-supervised learning
- **Research Implementation:** Distributed training, experiment design, research codebases, ablation studies
- **Programming, Systems & Deployment:** Python, Git, Docker, AWS, Google Cloud, Flask, MongoDB
- **Data Science:** Pandas, NumPy, Scikit-learn, statistical modelling, data visualization
- **Medical Domain Expertise:** X-Ray Diagnostics, Ultrasound Imaging, Retinal Fundoscopy

PUBLICATIONS

Full list of publications available on Google Scholar | Citations: 454, h-index: 9

Journal Articles

- Ong Ly, C.*, **Unnikrishnan, B.***, Tadic, T., et al. (2024). "Shortcut learning in medical AI hinders generalization: method for estimating AI model generalization without external data." *Nature npj Digital Medicine* (*equal contribution)
- Nguyen, C., Raja, A., Zhang, L., Xu, X., **Unnikrishnan, B.**, et al. (2023). "Diverse and consistent multi-view networks for semi-supervised regression." *Machine Learning*
- **Unnikrishnan, B.**, Nguyen, C., Balaram, S., Li, C., et al. (2021). "Semi-supervised classification of radiology images with NoTeacher: A teacher that is not mean." *Medical Image Analysis*
- Koohbanani, N. A., **Unnikrishnan, B.**, Khurram, S. A., et al. (2021). "Self-path: Self-supervision for classification of pathology images with limited annotations." *IEEE Transactions on Medical Imaging*

Conference Proceedings

- **Unnikrishnan, B.**, Brudno, M., & McIntosh, C. (2025). "SilverLining: Data-First Debiasing of Spatial and Spectral Shortcuts through Attention." *CVPR 2025* (Under Review)
- **Unnikrishnan, B.**, Nguyen, C. M., Balaram, S., et al. (2020). "Semi-supervised classification of diagnostic radiographs with noteacher: A teacher that is not mean." *MICCAI 2020*
- Nguyen, Q. H., Nguyen, B. P., Dao, S. D., **Unnikrishnan, B.**, et al. (2019). "Deep learning models for tuberculosis detection from chest X-ray images." *ICT 2019*
- Dutta, R., Raju, S., James, A., Leo, C. J., Jeon, Y., **Unnikrishnan, B.**, et al. (2019). "Learning of multi-dimensional analog circuits through generative adversarial network (GAN)." *IEEE SOCC*

Workshop & Other Publications

- **Unnikrishnan, B.**, Singh, P. R., Yang, X., Chua, M. C. H. (2020). "Semi-supervised and unsupervised methods for heart sounds classification in restricted data environments." *arXiv preprint arXiv:2006.02610*.
- Yu, Y., Kumar, A. J. S., Guretno, F., Balaram, S., **Unnikrishnan, B.**, Krishnaswamy, P., Ho Mien, I. (2023). "Integrated Platform for Resource-efficient Medical Image Annotation." *International Conference on AI in Medicine (iAIM)*, Singapore.
- Ouardini, K., Yang, H., **Unnikrishnan, B.**, et al. (2019). "Towards practical unsupervised anomaly detection on retinal images." *DART/MICCAI Workshop 2019*, 225-234.
- Lecouat, B., Chang, K., Foo, C. S., **Unnikrishnan, B.**, et al. (2018). "Semi-supervised deep learning for abnormality classification in retinal images." *Machine Learning for Health (ML4H) Workshop at NeurIPS*.
- Jin, C., Badawi, A. A., **Unnikrishnan, B.**, et al. (2019). "CareNets: Efficient Homomorphic CNN for High Resolution Images." *Privacy in Machine Learning workshop at NeurIPS*.

PROJECTS

Correction for Spurious Correlations on Vision-Language Models 2024 - Present

- Studied the effect of spurious correlations in vision-language models, identifying critical issues in diagnostic accuracy and retrieval.
- Leading the research on unsupervised identification of bias and lightweight correction techniques for various medical modalities.

Pneumothorax Detection & Triaging 2021 - 2024

- Curated and processed 200K X-ray dataset for comprehensive model training
- Developed detection pipeline deployed at University Health Network (UHN) for improving scan-to-intervention response times
- Identified and corrected critical data bias and confounder issues impacting model generalization

Red Teaming Multi-Modal Vision-Language Models for Healthcare 2023 - 2024

- Conducted analysis of 4 leading vision-language models (both open and closed-source) for healthcare applications
- Identified critical confounder dependencies and systematic biases in model reasoning across demographic factors
- Discovered hallucination effects for rare diseases, demonstrating potential impacts on clinical workflow

Resource-Efficient Diffusion Models for Healthcare 2023 - 2024

- Developed lightweight diffusion model training strategy for GPU-constrained settings
- Created cached latent augmentation technique improving generation quality by 40% while maintaining throughput
- Reduced GPU memory usage by 37% and compute by 6x through these optimizations

AI Robustness & Generalization Framework for Healthcare 2021 - 2024

- Developed novel multi-view loss function improving semi-supervised radiology model performance from 0.91 to 0.96 AUC while achieving 97% of fully supervised performance with 100x fewer labels
- Created framework for quantifying and mitigating AI model deterioration during healthcare deployment

ACHIEVEMENTS / AWARDS / VOLUNTEER POSITIONS

- Schwartz Reisman Institute (SRI) Graduate Fellowship 2024
- IEEE Transactions on Medical Imaging (TMI) Distinguished Reviewer 2024
- Reviewer: Nature Scientific Reports 2024
- University of Toronto Fellowship, Faculty of Arts and Science 2023
- Vector Institute Research Grant 2022 - 2024
- Mentor: Toronto Graduate Application Assistance Program (GAAP) 2022 - 2024
- AI Product Manager Nanodegree - Udacity 2020
- Richard E Merwin Scholar - IEEE Computer Society 2017

RESEARCH IN MEDIA

- Featured as invited guest on ATGO-AI (Accountability, Trust, Governance, and Oversight of AI) podcast discussing data biases and generalization issues in AI for Healthcare – Episodes available on Spotify