Balagopal (Balu) Unnikrishnan

Ph.D. Candidate, Computer Science
University of Toronto

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

PROFESSIONAL SUMMARY

PhD Researcher specializing in model generalization and safety for healthcare AI, with expertise in detecting and mitigating shortcut learning, data quality issues and spurious correlations. Proven track record building end-to-end AI pipelines: from identifying high-impact use cases, LLM-centered data curation, fixing critical data/shortcut issues, deploying and monitoring models for drift and compliance. Prior industry experience at A*STAR, Singapore with successful technology commercialization. Strong publication record with 14+ publications (including Nature Digital Medicine) and interdisciplinary collaborations across clinical, engineering, and regulatory teams.

EDUCATION

University of Toronto, Canada — 2021 - 2026 (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 (NUS), 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

EXPERIENCE

Ph.D. Candidate — 2021 - Present
University of Toronto | Vector Institute | University Health Network
Schwartz Reisman Institute & Vector Institute Graduate Fellow

- **Algorithmic Innovation:** Developed SilverLining framework for data-centric mitigation of spatial/spectral confounders (WACV 2024); Co-created framework detecting shortcut learning across 13 medical datasets, revealing 20% average performance overestimation and enabling external accuracy prediction within 4% without validation data (npj Digital Medicine 2024)
- **Clinical Translation:** Led study evaluating shortcut mitigation algorithms on 140K+ chest X-rays and reports for pneumothorax triage, achieving a 27-minute median turnaround time reduction for critical cases (UHN); developed semi-supervised GUI and models for detecting papilledema with 0.98 AUC from ultrasounds (SickKids)
- **AI Safety:** Studied AI usage and issues in 1000+ FDA-cleared AI medical devices; red teamed vision-language models identifying failure modes; evaluated 13 shortcut mitigation algorithms; investigated economic advantages of inclusive AI development in market expansion and product development

AI Research Engineer — 2019 - 2021
*Institute for Infocomm Research (I2R), A*STAR Singapore*
Top 1% performer

- **Algorithm Development:** Developed NoTeacher achieving 90-95% fully-supervised performance with 5-15% labels across X-ray, CT, MRI modalities (Medical Image Analysis 2021, MICCAI 2021); co-designed self-supervision framework for pathology images achieving state-of-the-art semi-supervised performance with limited annotations (IEEE TMI 2021)
- **Industry Translation/Clinical Collaboration**: Generated semi-supervised algorithm IP which was successfully licensed for commercial applications; Partnered with 2 leading hospitals for algorithm development and validation; established cross-functional research-to-deployment pipeline

Research Intern — 2018 - 2019
*Institute for Infocomm Research (I2R), A*STAR Singapore*

- Investigated semi-supervised GANs for disease detection in retinal imaging (ML4H Workshop/NeurIPS 2019), anomaly detection (DART Workshop/MICCAI)
- Developed novel patch-based architectures and transfer learning approaches for limited-annotation medical image analysis

KEY PUBLICATIONS

Full list: Google Scholar | Citations: 552, h-index: 9

- **Unnikrishnan, B.**, Brudno, M., McIntosh, C. "SilverLining: Data-First Mitigation of Spatial and Spectral Shortcuts Without Introducing New Confounders." *IEEE/CVF WACV 2026*
- Ong Ly, C.*, **Unnikrishnan, B.***, et al. "Shortcut learning in medical AI hinders generalization: method for estimating AI model generalization without external data." *Nature Digital Medicine 2024* (*equal contribution)
- **Unnikrishnan, B.***, Adames, A. G.*, Adibi, A.*, Peesapati, S., et al. "Beyond Ethics: How Inclusive Innovation Drives Economic Returns in Medical AI." arXiv:2510.10338. *Under review at Nature Digital Medicine 2025*
- **Unnikrishnan, B.**, et al. "Semi-supervised classification of radiology images with NoTeacher: A teacher that is not mean." *Medical Image Analysis*, 73. **MICCAI Best Paper Runner-up 2021**
- Koohbanani, N. A., **Unnikrishnan, B.**, et al. "Self-path: Self-supervision for classification of pathology images with limited annotations." *IEEE Transactions on Medical Imaging 2021*

SELECTED PROJECTS

Mitigating Shortcut Learning in Medical AI   2021 - Present

- Developed SilverLining algorithm for identifying/mitigating spatial/spectral shortcuts (42-73% reduction), achieving 0.94 AUC on counter-shortcut evaluation; data-centric method enabled extension to classification and object detection tasks (WACV 2026)
- Co-developed method to predict model generalization without external data, revealing 21% average performance overestimation across 13 datasets and predicting external accuracies with <4% error (Nature Digital Medicine 2024)
- Utilized generative AI debiasing algorithm to achieve SOTA performance with 7% margin over competing methods, matching unbiased training performance while using 95% biased data

Pneumothorax Detection & Clinical Triage System   2021 - Present

- Built scalable pipeline processing 140K+ X-ray image-report dataset at UHN for LLM-based structured label extraction
- Assessed triage equity demonstrating reduction in performance disparities across demographic/technical subgroups
- Led evaluation of 13 shortcut mitigation algorithms on 140K+ chest X-rays for pneumothorax triage, achieving 27-minute median time-to-treatment reduction

Clinical Papilledema Detection System   2023 - Present

- Led end-to-end development from data collection to deployment: designed annotation workflow/tool, developed keypoint detection model, and implemented clinical evaluation pipeline
- Achieved state-of-the-art performance (0.98 AUC) in disease detection and precise localization (<1% normalized MSE)
- Ensured robust cross-device generalization by validating across 6 types of ultrasound devices and implementing systematic bias mitigation strategies

Resource-Efficient Diffusion Models for Healthcare   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

Semi-Supervised Learning Framework (NoTeacher)   2020 - 2021

- Developed novel algorithm achieving 90% of fully supervised performance with just 5% labels across X-ray and CT modalities (Medical Image Analysis 2021, MICCAI 2020 Runner-up Best Paper)
- Generated intellectual property "Semi-Supervised Process to Guide Annotation for Image Classification Tasks" - successfully licensed for commercial revenue
- Improved performance on radiology data from 0.91 to 0.96 AUC through multi-view loss function promoting learning of non-spurious features

Self-Supervised Learning for Pathology Images   2020

- Co-developed self-supervision framework using multi-resolution and semantic features for gigapixel histopathology images
- Achieved state-of-the-art performance on whole slide images with limited annotations (IEEE TMI 2021)
- Extended framework to handle domain transfer scenarios with minimal target domain data

IP & RESEARCH IN MEDIA

- **Technical Disclosure:** "Semi-Supervised Process to Guide Annotation for Image Classification Tasks" – Filed at A*STAR, Singapore. Successfully licensed for commercial revenue generation
- **Invited Guest:** ATGO-AI Podcast (Accountability, Trust, Governance, and Oversight of AI) – Invited guest discussing shortcut learning and generalization challenges in healthcare AI. Episodes available on Spotify (2024)

TECHNICAL SKILLS

- **ML/AI Frameworks:** PyTorch, TensorFlow, distributed training, model deployment, generative models (GANs, diffusion), vision-language models (CLIP variants), LLMs (offline/cloud variants, vLLM)
- **Cloud & Production:** Google Vertex AI, Google AutoML, AWS, Docker, REST APIs, unit testing, model monitoring
- **Medical Imaging:** DICOM processing, SimpleITK/ITK, OpenCV, X-ray/ultrasound/pathology pipelines, annotation platforms
- **MLOps & Research:** Experiment tracking, ablation studies, SHAP/GradCAM (interpretability), technical writing/documentation, drift detection, semi/self-supervised learning, shortcut learning
- **Data & Development:** Python, Git, pandas, NumPy, scikit-learn, SQL, Flask, MongoDB, Jupyter, LaTeX

HONORS, LEADERSHIP, & VOLUNTEERING

- Schwartz Reisman Institute (SRI) Graduate Fellowship   2024
- IEEE Transactions on Medical Imaging Distinguished Reviewer - Silver Level   2024
- Vector Institute Research Grant   2022 - 2025
- Runner-up: MICCAI Best Paper Award   2020
- University of Toronto Graduate Fellowship   2023
- Graduate Mentor: Toronto Graduate Application Assistance Program   2022 - Present
- Reviewer: Nature Digital Medicine, Medical Image Analysis   2024 - Present