Balagopal (Balu) Unnikrishnan

Ph.D. Candidate, Department of Computer Science
University of Toronto

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

PROFESSIONAL SUMMARY

PhD Researcher working on robustness and generalization in AI. I study why models break (shortcut learning, spurious correlations, distribution shift) and build methods to detect and fix these failures. Industry and academic experience across the full pipeline: data curation, model development, clinical deployment, drift monitoring, and commercialization. 14+ publications including Nature Digital Medicine, with interdisciplinary experience across clinical, regulatory, engineering and research teams.

EDUCATION

University of 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 (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); 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)
- **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 1200+ 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, MICCAI); co-designed self-supervision framework for pathology images achieving state-of-the-art semi-supervised performance with limited annotations (IEEE TMI)
- **Industry Translation/Clinical Collaboration**: Generated semi-supervised algorithm IP, 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*

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

KEY PUBLICATIONS

Full list: Google Scholar — Citations: 564, 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

RAG-Based Document Analysis for Medical Device Research 2024 - 2025

- Designed multi-stage RAG pipeline: Gemini 2.5 Flash-Lite for multimodal extraction and structured JSON parsing (24,000+ pages), local Mistral-7B via Ollama for cost-efficient QnA, achieving 94% retrieval precision with 60% reduction in API costs
- Reduced document analysis from 250+ manual hours to under 4 hours; identified systematic gaps including homogeneous validation datasets and sparse change control plans in 510(k) AI device filings
- Characterized clinical integration patterns across 1,200+ FDA-cleared AI devices: 96% require human oversight, 53% operate parallel to standard workflow; informing research on real-world deployment constraints and shortcut learning risks

Pneumothorax Detection & Clinical Triage System 2021 - Present

- Built scalable pipeline for processing 140K+ X-ray image-report dataset for LLM-based structured label extraction
- Led evaluation of 13 shortcut mitigation algorithms for pneumothorax triage, reducing performance disparities by 22 to 42% across demographic/technical confounders and achieving 27-minute median report turnaround time reduction

Resource-Efficient Generative Models for Confounder Mitigation 2021 - Present

- Developed lightweight diffusion training with cached latent augmentation: 40% improved generation quality, 37% reduced GPU memory, 6x compute reduction, enabling deployment in GPU-constrained clinical settings
- Applied confounder mitigation using generative methods: 7% subgroup AUC improvement over baselines, matching clean data performance while using only 5% clean data

Clinical Papilledema Detection System 2023

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

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, MICCAI Best Paper Runner-up)
- Generated IP/technical disclosure "Semi-Supervised Process to Guide Annotation for Image Classification Tasks" - successfully licensed for commercial application

TECHNICAL SKILLS

- **Deep Learning:** PyTorch, distributed training, generative models (GANs, diffusion), semi/self-supervised learning, domain adaptation, data-efficient learning, model interpretability
- **Multi-Modal AI & LLMs:** Vision-language models, fine-tuning, Ollama/vLLM, prompt engineering, RAG pipelines
- **Medical Imaging:** Radiology (X-ray, CT, MRI, ultrasound), digital pathology, DICOM processing, image reconstruction, clinical workflow integration, OpenCV
- **Robustness & Generalization:** Shortcut mitigation, spurious correlation detection, distribution shift, out-of-distribution detection, responsible AI evaluation
- **Research & MLOps:** Experiment design, ablation studies, reproducibility, Azure/AWS, Docker, model monitoring, drift detection, cross-functional collaboration
- **Data & Development:** Python, Git, pandas, NumPy, SQL, large-scale data pipelines, scientific writing

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

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)

CERTIFICATIONS

- Agentic AI - Deeplearning.ai 2025
- AI Agents in LangGraph - Deeplearning.ai 2025
- Fine-tuning & RL for LLMs - Deeplearning.ai 2025
- Retrieval Augmented Generation (RAG) - Deeplearning.ai 2025
- Finetuning Large Language Models - Deeplearning.ai 2024
- AI Product Manager Nanodegree - Udacity 2020
- Deep Learning Specialization - Deeplearning.ai / Coursera 2017