Shantanu Ghosh

shawn24@bu.edu   linkedin.com/in/shantanuai   Boston, MA, USA

Dear Recruitment Team,

I am writing to express my interest in the Research Scientist Intern position at your company for Summer, 2026. My research focuses on demystifying deep learning models and addressing their biases, particularly in chest X-rays (CXR) and mammography, through the following 2 projects:

Project 1: Error Slice Discovery and Mitigation in Vision Models with Language: Error slices are coherent data subsets where models systematically fail. Current slice discovery and mitigation methods face key limitations: 1. These methods lack complex reasoning capabilities, preventing them from fully explaining model biases. 2. They fail to integrate domain knowledge, limiting their usage in specialized fields like radiology. 3. They only identify visual biases and are unable to locate biases arising from metadata, data collection, or preprocessing pipelines. To overcome these challenges, I propose LADDER (**ACL 2025**). LADDER leverages the alignment of vision language models (VLMs), reasoning ability and domain knowledge of LLMs to identify biases for slice discovery. For mitigation, LADDER generates pseudo attributes from the discovered slices to mitigate errors across all biases without explicit attribute annotations or prior knowledge of biases. **Application.** We leverage LADDER in the popular risk prediction model MIRAI for predicting breast cancer risk. As LADDER requires a VLM to probe MIRAI, we develop Mammo-CLIP (**MICCAI 2024, top 11%**), a robust VLM that demonstrates state-of-the-art performance in classifying and localizing key mammographic attributes critical to breast cancer detection, offering robustness and data efficiency.

Project 2: Making Deep Models Interpretable Without Compromising Performance: Deep models are termed as black-boxes.Post-hoc explanation methods maintain the performance of the black-box models, but their explanations are not faithful to the predictions and do not support recourse. Conversely, concept-based interpretable-by-design methods allow for concept-level intervention but often underperform compared to black-box models. To address this challenge, I develop MoIE (**ICML 2023**) to carve out concept-based models from a black box, enhancing interpretability and maintaining performance. **Application.** The concepts learned by our method are domain-invariant, similar to clinical rules in radiology. Building on this hypothesis, I introduce MoIE-CXR (**MICCAI 2023, top 14%**), applying the framework to CXRs for efficient transfer learning with limited data and computational resources.

Additional projects: During my recent internship with the Amazon AWS SAAR team, I address challenges in learning robust representations for tabular data in self-supervised models (**Table Representation Workshop@NeurIPS 2024**). Earlier, I utilize anatomical landmarks from the Stanford RadGraph to develop an attention-driven algorithm to localize disease in the MIMIC-CXR dataset (**MICCAI 2022** and **Radiology: AI journal**).

Regards,

Shantanu Ghosh
Ph.D. Candidate
Electrical Engineering, Boston University