

Shaibal Saha

+1-248-710-7404 | shaibalsaha@oakland.edu | [My portfolio](#)

 [LinkedIn](#) |  [GitHub](#) |  [Google Scholar](#)

Auburn Hills, Michigan - 48326, USA

EDUCATION

■ Oakland University

PhD Candidate, Department of Computer Science and Informatics

September 2022 - August 2027 (Expected)

Rochester Hills, USA

■ Oakland University

Masters of Science, Department of Computer Science and Informatics

September 2022 - December 2024

Rochester Hills, USA

■ North South University

Bachelor of Science in Computer Science and Engineering

January 2014-August 2018

Dhaka, Bangladesh

WORK EXPERIENCE

■ Oakland University

Graduate Research Assistant

September 2022 - Present

Rochester Hills, USA

- Conduct research on optimization techniques for deep neural networks in Edge environments to enhance performance and efficiency.
- Implement and analyze new model compression techniques, including pruning and quantization, for transformers and hybrid models for generalized application/domain-specific applications.
- Implement an efficient RAG-based system utilizing client-side caching to reduce the LLM cost.
- Study multi-modal detection models for Autonomous Driving Systems, focusing on hardware/software co-design for optimized deployment.
- **Frequently used tools:** PyTorch, TensorRT, ONNX Runtime, Hugging Face, CUDA & NVIDIA Nsight, Docker, Triton Inference Server, LangChain, FAISS/Redisstack, MLflow.

Graduate Teaching Assistant

- CSI3620: Data Structures and Algorithms
- CSI3480: Security and Privacy in Computing

- CSI3660: System Administration
- CSI2310: Data Structures

■ Code Aristos

Software Developer

January 2019 - June 2022

Dhaka, Bangladesh

- Started as an intern (Jan 2019 - Apr 2019) and transitioned to a full-time role due to strong performance.
- Worked on Hospital Management ERP for large-scale data handling for both web and mobile apps.
- Provided real-time client support for bug fixes for live systems all across the world.
- Designed and implemented robust backend functions for multi-level user access and customization based on client needs.
- **Frequently used tools:** Django, Laravel, MySQL, MongoDB, Azure, REST APIs, Docker, Redis, Nginx, GitHub Actions.

RESEARCH INTEREST

Efficient AI, Edge Computing, Large Language Model Systems, Autonomous Vehicles, Connected Health.

PUBLICATIONS

B= BOOK CHAPTER, C= CONFERENCE, J= JOURNAL

[B] **Shaibal Saha**, Qiren Wang, Lichen Xia, Yongtao Yao. Chapter 4: Towards edge intelligence. In *Edge Computing: Systems and Applications*, edited by Lanyu Xu and Weisong Shi. Wiley Press, 2025.

[J] **Shaibal Saha**, Lanyu Xu. Vision transformers on the edge: A comprehensive survey of model compression and acceleration strategies. *Neurocomputing*, 643 (2025), 130417. [DOI](#)

[C] **Shaibal Saha**, Lanyu Xu. EfficientQuant: An Efficient Post-Training Quantization for CNN-Transformer Hybrid Models on Edge Devices. *T4V-CVPR Workshop*, 2025. [arXiv/DOI](#)

[C] Yunge Li, **Shaibal Saha**, Lanyu Xu. The architectural implications of multi-modal detection models for autonomous driving systems. *IEEE MOST*, 2024. [DOI](#)

[C] Ronald Tudu, **Shaibal Saha**, Prasun Nandy Pritam, Rajesh Palit. Performance analysis of supervised machine learning approaches for Bengali text categorization. *IEEE APWC on CSE*, 2018. [DOI](#)

SKILLS

- **Programming Languages:** **Expert:** Python, PHP, VHDL | **Proficient:** C, C++, JavaScript.
- **Libraries:** PyTorch, TensorFlow, Matplotlib, Seaborn, Hugging Face, MONAI, Scikit-learn, NumPy, Pandas.
- **Model Optimization & Deployment:** TensorRT, ONNX Runtime, NVIDIA Triton, Vitis-AI, XFormers, VLLM, TVM.
- **Profiling & Evaluation Tools:** PyTorch Profiler, MLflow, NVIDIA Nsight, LangSmith, Vaitrace.
- **RAG & Vector Databases:** LangChain, LlamaIndex, FAISS, RedisStack.
- **HPC:** Oakland University (OU) central HPC Linux-based cluster (Matilda)
- **Hardware Platforms:** NVIDIA GPUs, Jetson Nano, Jetson AGX Xavier, Xilinx ZCU104 MPSoC.
- **Cloud & MLOps:** Docker, CI/CD (GitHub Actions), Microsoft Azure, AWS.
- **Web & Backend Development:** Django, Laravel, REST APIs, Nginx.
- **Databases & Collaboration:** MySQL, PostgreSQL, MongoDB, Git, GitHub, GitLab, Jira.
- **Completed Courses:** Embedded System Design Using FPGAs, Introduction to EdgeAI, Computer Architecture, Advanced Data Structure and Algorithms, Cloud Computing, Database System, Discrete methods I, Component-based software engineering.

HONORS AND AWARDS

- Participated in the 32nd Intelligent Ground Vehicle Competition (IGVC) 2025 in the self-driving category, *Position: 3rd (Grand award and overall)*.
- NSF Student Travel Award for attending IEEE/ACM SEC 2023.
- Bangladesh-Sweden Trust Fund (BSTF)- Travel Grant 2023.
- NSF Student Travel Award for attending IEEE/ACM CHASE 2023.
- Oakland University Provost Graduate Student Research Award grant - Spring 2023.
- Selected for university special scholarship as an undergraduate for maintaining good results.

PROFESSIONAL SERVICES

- Reviewer of Elsevier Neurocomputing. [2026-Present]
- Reviewer of Elsevier Expert Systems with Applications. [2025-Present]
- Reviewer of Elsevier Smart Health. [2025-Present]
- Reviewer of IEEE transactions on emerging topics in computing. [2025-Present]
- Reviewer of Journal of Ambient Intelligence and Smart Environment. [2023-Present]
- Reviewer of Workshop on Multi-lingual Representation Learning (MRL). [2024,2025]

LEADERSHIP EXPERIENCE & PROFESSIONAL MEMBERSHIPS

- **ACM Student Chapter**,President *September 2024 - Present*
- **ACM Student Chapter**,Treasurer *September 2023 - August 2024*

REFERENCES

1. **Lanyu Xu, Ph.D**
Assistant Professor, Computer Science and Engineering
Oakland University
Email: lxu@oakland.edu
Phone: (248)370-4079
Relationship: PhD Advisor
2. **Arun Iyengar**
Computer Scientist
Intelligent Data Management and Analytics, LLC
Email: aruniyen3@gmail.com
Relationship: [Mentor]
3. **Debatosh Debnath, Ph.D**
Associate Professor, Computer Science and Engineering
Oakland University
Email: debnath@oakland.edu
Phone: (248)370-2701
Relationship: Course Instructor and Organization Advisor