

Shaibal Saha

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

[!\[\]\(919a2cb85b99741a73c0c31a427236a8_img.jpg\) LinkedIn](#) | [!\[\]\(c9cd5a1c35167a83f09a35036fe5dcbd_img.jpg\) GitHub](#) | [!\[\]\(ae1936640fabdea8c18f922ca69733fe_img.jpg\) Google Scholar](#)

Auburn Hills, Michigan - 48326, USA

EDUCATION

- **Oakland University** *September 2022 - August 2027 (Expected)*
PhD Candidate, Department of Computer Science and Informatics Rochester Hills, USA
CGPA: 4.00 / 4.00

- **Oakland University** *September 2022 - December 2024*
Masters of Science, Department of Computer Science and Informatics Rochester Hills, USA
CGPA: 4.00 / 4.00

- **North South University** *January 2014-August 2018*
Bachelor of Science in Computer Science and Engineering Dhaka, Bangladesh
CGPA: 3.44 / 4.00

RESEARCH INTEREST

Deep learning, Edge Computing, Model Compression, Cache in LLM, Connected Health, Autonomous Driving

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, Weisong Shi

- [C] **Shaibal Saha**, Lanyu Xu. EfficientQuant: An Efficient Post-Training Quantization for CNN-Transformer Hybrid Models on Edge Devices (Accepted in T4V-CVPRW 2025).

- [J] **Shaibal Saha**, Lanyu Xu. Vision Transformers on the Edge: A Comprehensive Survey of Model Compression and Acceleration Strategies. arXiv, DOI: <https://doi.org/10.48550/arXiv.2503.02891> (Accepted in NeuroComputing).

- [C] Yunge Li, **Shaibal Saha**, Lanyu Xu (2024). The Architectural Implications of Multi-modal Detection Models for Autonomous Driving Systems. In *IEEE International Conference on Mobility: Operations, Services, and Technologies 2024 (MOST 2024)*, DOI: 10.1109/MOST60774.2024.00030o

- [C] Ronald Tudu, **Shaibal Saha**, Prasun Nandy Pritam and Rajesh Palit (2018). Performance Analysis of Supervised Machine Learning Approaches for Bengali Text Categorization. In *2018 5th IEEE Asia-Pacific World Congress on Computer Science and Engineering (APWC on CSE)*, DOI: 10.1109/APWCConCSE.2018.00043

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.
- Developing an efficient cache management for LLM-based system.
- Study multi-modal detection models for Autonomous Driving Systems, focusing on hardware/software co-design for optimized deployment.
- **Frequently used tools:** PyTorch, TensorRT, Vitis-AI, Vaitrace, CUDA, Langchain, VLLM

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 to 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:** Codeigniter, Django, Laravel, MongoDB, MySQL, Azure, Apache

SKILLS

- **Programming Languages:** [Expert] Python, PHP, VHDL
[Moderate] C, C++, Javascript
- **Tools:** PyTorch, TensorFlow, TensorRT, Pandas, Langchain, VLLM, XFormers, Monai, NumPy, Scikit-learn, Vaitrace
- **Hardware used:** GPU, Xilinx ZCU104 MPSoC, Jetson Nano, NVIDIA Jetson AGX Xavier
- **Other Technologies:** Docker, Vitis-AI, Intel Quartus, Vivado Design suite, Microsoft Azure, Apache, AWS, Huggingface
- **Web Technologies:** Django, Codeigniter, Laravel
- **Database Systems:** MySQL, NoSQL(MongoDB), Oracle, RedisStack, Redis
- **Logging & monitoring:** Drawio, Jira, Github, Git, Gitlab, ELK Stack, Huggingface
- **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.
- **Languages:** English (Fluent), Bangla (Native), Hindi (Professional)

PROFESSIONAL SERVICES

- 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 [2024,2025]

HONORS AND AWARDS

- Participated in the 32nd Intelligent Ground Vehicle Competition (IGVC) 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 at the undergraduate for maintaining good result

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 Department
Oakland University
Email: lxu@oakland.edu
Phone: (248)370-4079
Relationship: PhD Advisor

2. Debatosh Debnath, Ph.D

Associate Professor, Computer Science and Engineering Department
Oakland University
Email: debnath@oakland.edu
Phone: (248)370-2701
Relationship: Course Instructor and Organization Advisor