
Siddhartha Shankar Das

Email: siddhartha.das@pnnl.gov / siddhartha047@gmail.com

[Google Scholar](#) | [LinkedIn](#) | [Website](#) | [PNNL Staff Page](#)

INTERESTS

I am a Post-Doctorate Researcher at Pacific Northwest National Laboratory (PNNL) in the **Language Intelligence** group within the **Data Science and Machine Intelligence** department. My research focuses on artificial intelligence and machine learning for scientific and applied domains, with interests spanning graph neural networks, graph sparsification, natural language processing, cybersecurity, and scalable AI systems. I develop learning-based methods for system design and optimization, including hardware/software partitioning and efficient learning on large graphs. I also build LLM-based, multi-modal, and agentic AI systems for document intelligence, retrieval-augmented generation, georeferencing, and interactive decision support. My work has been published in top-tier venues and has received best paper awards.

EDUCATION

Purdue University, West Lafayette, IN, USA

Ph.D. in Computer Science

Aug 2018 – May 2025

Advisor: [Alex Pothén](#)

Thesis: Scaling Graph Neural Networks for Large Graphs and Analyzing Common Vulnerabilities via Large Language Models [\[pdf\]](#).

Topics: Machine Learning on Graphs, Graph Neural Networks, Cybersecurity, Large Language Models

Bangladesh University of Engineering and Technology (BUET), Dhaka, Bangladesh

M.Sc. in Computer Science and Engineering

Sep 2017

GPA: 4.00/4.00

Thesis: Evolutionary Algorithm Using Adaptive Fuzzy Dominance and Reference Point for Many-Objective Optimization (*F*-DEA) [\[pdf\]](#).

Advisor: [Md. Monirul Islam](#)

Topics: Multi-objective Optimization, Evolutionary Algorithms

Bangladesh University of Engineering and Technology (BUET), Dhaka, Bangladesh

B.Sc. in Computer Science and Engineering

Jul 2014

GPA: 3.88/4.00, Major: 3.99/4.00

Thesis: A Many-objective Evolutionary Approach Using Fuzzy Dominance with Bidirectional Bias [\[pdf\]](#).

Advisor: [Md. Monirul Islam](#)

Topics: Multi-objective Optimization, Evolutionary Algorithms

WORK EXPERIENCE

Postdoctoral Researcher, Data Science & Machine Intelligence

May 2025 – Present

[Pacific Northwest National Laboratory \(PNNL\)](#), Richland, WA, USA.

- Conduct research in artificial intelligence, machine learning, graph algorithms, and system design, with a focus on scalable Graph Neural Network (GNN)-based methods for optimization problems such as hardware/software partitioning, scheduling, graph partitioning, and graph sparsification.
- Develop LLM-based, multimodal, and agentic AI systems for scientific and applied applications, including large-scale processing of environmental and regulatory documents. Current work includes

metadata extraction, georeferencing, project summarization, retrieval-augmented generation, and interactive decision-support systems.

- Contribute to interdisciplinary projects spanning cybersecurity, document intelligence, and scientific machine learning, translating foundational methods into scalable tools, datasets, deployable systems, and research publications.

Graduate Research Assistant

Aug 2018 – May 2025

Department of Computer Science, Purdue University, West Lafayette, IN, USA.

- Conducted research on scalable graph learning, graph sparsification, and machine learning on large graphs, along with LLM-based methods for software vulnerability analysis and cyber threat understanding.
- Developed new sampling and sparsification methods for Graph Neural Networks, leading to publications in venues including KDD, DSAA, Big Data, and HST.

Ph.D. Intern, Data Analytics

Summers 2019–2021

[Pacific Northwest National Laboratory \(PNNL\)](#), Richland, WA, USA

Mentored by [Mahantesh Halappanavar](#).

- Developed learning-based methods for classification and analysis of software vulnerabilities, including mapping CVEs to weaknesses (CWEs) and attack patterns (CAPECs).
- Used Large Language Models (LLMs) and Natural Language Processing (NLP) techniques, and scaled these methods for modern AI accelerators.
- Published four papers during the internship period, including work recognized with two best paper awards.

Assistant Professor, Department of Computer Science and Engineering

Dec 2017 – Jul 2018

[Bangladesh University of Engineering and Technology \(BUET\)](#), Dhaka, Bangladesh

Roles: University teaching, project supervision, and office administration.

Study Leave for Ph.D. Jul 2018 – Jul 2024

Lecturer, Department of Computer Science and Engineering

Aug 2014 – Dec 2017

[Bangladesh University of Engineering and Technology \(BUET\)](#), Dhaka, Bangladesh

Roles: University teaching.

SELECTED PUBLICATIONS (FULL LIST AT [GOOGLE SCHOLAR](#))

Conference and Workshop Publications

- **Siddhartha Shankar Das**, N. Arafat, M. Rahman, S. M. Ferdous, A. Pothen, M. Halappanavar, and D. Rawat. “SGS-GNN: A Supervised Graph Sparsifier for Graph Neural Networks.” *In Proceedings of the 32nd ACM SIGKDD Conference on Knowledge Discovery and Data Mining V.2 (KDD 2026)*, August 9–13, 2026, 12 pages. [\[doi\]](#)
- **Siddhartha Shankar Das**, Zakaria Mehrab, James Kotary, Rounak Meyur, S. M. Ferdous, Erdal Mutlu, and Mahantesh Halappanavar. “Diff-GNN: Differentiable Graph Neural Networks for Joint Hardware–Software Partitioning and Scheduling.” Accepted at HPEC 2026.
- James Kotary, Zakaria Mehrab, **Siddhartha Shankar Das**, Rounak Meyur, S. M. Ferdous, Erdal Mutlu, and Mahantesh Halappanavar. “Joint Optimization of HW/SW Partitioning and Scheduling with Deep Reinforcement Learning.” Accepted at HPEC 2026.

-
- Moqsadur Rahman, Monika Akbar, Joseph T. Aguayo, Mahmud Shahriar Hossain, Tina M. Ellis, **Siddhartha Shankar Das** et al. “ATT&CK to CVE: A Large-Scale Automated Knowledge Graph for Threat Intelligence.” *IEEE International Conference on Data Mining (ICDM) GTA’3 Workshop*, 2025. [\[paper\]](#)
 - **Siddhartha Shankar Das**, S. M. Ferdous, M. Halappanavar, E. Serra, and A. Pothén. “AGS-GNN: Attribute-guided Sampling for Graph Neural Networks.” In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD)*, 2024. [\[paper\]](#)
 - **Siddhartha Shankar Das**, E. Serra, M. Halappanavar, A. Pothén, and E. Al-Shaer. “V2W-BERT: A Framework for Effective Hierarchical Multiclass Classification of Software Vulnerabilities.” In *2021 IEEE 8th International Conference on Data Science and Advanced Analytics (DSAA)*, pp. 1–12, 2021. [\[doi\]](#) [\[Best Application Paper\]](#)
 - **Siddhartha Shankar Das**, A. Dutta, S. Purohit, E. Serra, M. Halappanavar, and A. Pothén. “Towards Automatic Mapping of Vulnerabilities to Attack Patterns Using Large Language Models.” In *2022 IEEE International Symposium on Technologies for Homeland Security (HST)*, 2022. [\[doi\]](#) [\[Best Paper\]](#)
 - **Siddhartha Shankar Das**, M. Halappanavar, A. Tumeo, E. Serra, A. Pothén, and E. Al-Shaer. “VWC-BERT: Scaling Vulnerability–Weakness–Exploit Mapping on Modern AI Accelerators.” In *2022 IEEE International Conference on Big Data*, pp. 1224–1229, 2022. [\[doi\]](#)
 - K. Panchal, **Siddhartha Shankar Das**, L. Torre, J. Miller, R. Rallo, and M. Halappanavar. “Efficient Clustering of Software Vulnerabilities Using Self-Organizing Maps (SOM).” In *2022 IEEE International Symposium on Technologies for Homeland Security (HST)*, pp. 1–7, 2022. [\[doi\]](#)
 - A. Rizvi, T. R. Toha, **Siddhartha Shankar Das**, Sriram Chellappan, A. B. M. Alim Al Islam. “Many-objective Performance Enhancement in Computing Clusters.” In *2017 IEEE 36th International Performance Computing and Communications Conference (IPCCC)*, 2017. [\[doi\]](#)
 - A. Rizvi, T. R. Toha, **Siddhartha Shankar Das**, Sriram Chellappan, A. B. M. Alim Al Islam. “Exploiting a Synergy between Greedy Approach and NSGA for Scheduling in Computing Clusters.” In *2018 5th International Conference on Networking, Systems and Security*, 2018. [\[doi\]](#)

Journal Articles, Book Chapters, and Reports

- Moqsadur Rahman, Aaron Sanchez, Krish Piryani, **Siddhartha Shankar Das**, Sai Munikoti, Luis de la Torre Quintana, Monowar Hasan, Joseph Aguayo, Monika Akbar, Shahriar Hossain, and Mahantesh Halappanavar. “Generative AI for Advanced Cyber Defense.” *AI for Cybersecurity: Research and Practice*, John Wiley & Sons, Inc., pp. 109–146, 2025. [\[book chapter\]](#)
- Sai Munikoti, Dan Nally, Sai Dileep Koneru, **Siddhartha Shankar Das**, et al. “NEPATEC v2.0: Standardized Metadata and Text Corpus of National Environmental Policy Act Documents.” Pacific Northwest National Laboratory, 2025. [\[report\]](#)
- **Siddhartha Shankar Das**, Md. Monirul Islam, Naheed Anjum Arafat. “Evolutionary Algorithm Using Adaptive Fuzzy Dominance and Reference Point for Many-Objective Optimization.” *Swarm and Evolutionary Computation*, 44:1092–1107, 2019. [\[doi\]](#)

Preprints

- **Siddhartha Shankar Das**, S. K. Navuluru, S. M. Ferdous, R. A. Rossi, B. Coskunuzer, L. Tamil, E. Serra, A. Pothén, R. Rallo, and M. M. Halappanavar. “Scaffold: Support Graph Theory Based Sparsification for Graph Neural Networks.” *arXiv:2609.31466*, 2026. [\[arxiv\]](#)
- S. K. Navuluru, **Siddhartha Shankar Das**, F. Dernoncourt, S. M. Ferdous, R. A. Rossi, N. K. Ahmed, B. Coskunuzer, A. Pothén, L. Tamil, and M. M. Halappanavar. “EDiS: Edge Disjoint

Subgraph Sparsification Framework for Graph Neural Networks.” *arXiv:2610.09059*, 2026. [\[arxiv\]](#)

- S. K. Navuluru, **Siddhartha Shankar Das**, B. Ni, H. Chen, Y. Wang, B. Coskunuzer, N. K. Ahmed, F. Dernoncourt, M. Halappanavar, T. Derr, R. A. Rossi, and L. Tamil. “Unifying Graph Neural Networks Through a Common Layer Equation.” *arXiv:2608.16097*, 2026. [\[arxiv\]](#)

ONGOING AND PAST PROJECTS

Graph Learning, System Design, and Optimization

- **Graph-Based Circuit Equivalence Checking:** Developing graph learning methods for circuit equivalence verification using GNN-based subcircuit identification, gate embeddings, and graph matching.
- **Benchmarking Graph Sparsification Methods and Applications:** Developing a comprehensive textbook and benchmarking framework covering graph sparsification, rewiring, augmentation, theoretical guarantees, and applications.
- **Graph Neural Network Architecture Generation:** Exploring automated GNN architecture generation through a unified framework for message passing, aggregation, and transformation.

NEPA, Multimodal AI, and Agentic Systems

- **NEPATEC 2.0: Standardized NEPA Document Corpus:** Contributed to a publicly available NEPA dataset for document intelligence, released on [Hugging Face](#).
- **NEPATEC 2.5: Multimodal Metadata Extraction:** Developing multimodal LLM methods for NEPA metadata extraction, map-based geolocation, and project summarization.
- **PermitCE: AI-Assisted Chat for NEPA Documents:** Developing a retrieval-augmented generation system for semantic search, recommendations, summarization, and question answering.

Cybersecurity and Applied Machine Learning

- **Composite Metric for Cyber Risk Prioritization:** Developing a vulnerability prioritization metric combining CVSS, CWSS, EPSS, and LEV.
- **Predictive Analytics for Russet Potato Selection:** Applying machine learning and feature selection to phenotypic data for potato breeding decisions.
- **Machine Learning for USEPA SPECIATE Database:** Applying clustering methods to identify similarities among chemical source profiles.
- **Miscellaneous Collaborative Research:** Collaborating on interdisciplinary research in graph learning, cybersecurity, environmental science, and applied AI.

More Projects: A complete list of my other projects is available [here](#).

TEACHING AND MENTORING

Pacific Northwest National Laboratory (PNNL)

Aug 2025 – May 2026

Research Mentoring: Mentored five interns over the past year on graph neural networks, circuit representation learning, and cybersecurity, resulting in three accepted publications.

Purdue University, West Lafayette, IN, USA

Aug 2018 – May 2019

Department of Computer Science

Position: Graduate Research Assistant.

Courses: Algorithms, Data Structures, etc.

Bangladesh University of Engineering and Technology (BUET), Dhaka, Bangladesh Aug 2014 – Jul 2018
Department of Computer Science and Engineering
Position: Assistant Professor (2017-2018), Lecturer (2014-2017).
Courses: Machine Learning, Artificial Intelligence, Computer Graphics, Data Communication, and others ([List of Courses](#))

SKILLS

Research Areas: AI/ML, Graph neural networks, Graph Algorithms, Graph Sparsification, Natural Language Processing, Cybersecurity, Scientific Machine Learning.
Languages: Python, C/C++, Java, MATLAB, Julia, C#, SQL, JavaScript.
Frameworks/Tools: PyTorch, PyTorch Geometric, DGL, TensorFlow, Keras, MPI, Git, Linux, MOEA Framework.
LLM and AI Systems: Large language models, multimodal AI, agentic AI, retrieval-augmented generation (RAG), document intelligence, metadata extraction, georeferencing, interactive AI assistants.
Systems and Applications: System design, hardware/software partitioning, scalable AI systems, large-scale document processing, vulnerability analysis, CVE–CWE–CAPEC mapping, cyber risk prioritization.
Databases: Oracle, Microsoft SQL Server, MySQL, SQLite.

ACADEMIC SERVICES

Reviewer: [ICML'25](#), [KDD'24](#), [DSAA'23](#), [The Web Conference \(formerly WWW\)](#), [IEEE Transactions on Artificial Intelligence](#) and other venues.
Organizing Member: NSysS [2015](#), [2016](#), [2017](#).

AWARDS AND ACHIEVEMENTS

[Best Application Paper Award, DSAA 2021](#)
[Best Paper Award, HST 2022](#)
Dean's Honor List, BUET, 2011–2014
University Merit Scholarship, BUET

REFERENCES

[Alex Pothén](#)

Professor,
Department of Computer Science,
Purdue University,
West Lafayette, IN, USA.
Email: apothén@purdue.edu

[Edoardo Serra](#)

Associate Professor,
Department of Computer Science,
Boise State University,
Boise, ID, USA.
Email: edoardoserra@boisestate.edu

[Mahantesh Halappanavar](#)

Scientist and Acting Group Leader,
Advanced Computing, Mathematics and Data,
Pacific Northwest National Laboratory,
Richland, WA, USA.
Email: hala@pnnl.gov

[Md. Monirul Islam](#)

Professor,
Department of Computer Science,
Bangladesh University of Engineering and Technology,
Dhaka, Bangladesh.
Email: mmislam@cse.buet.ac.bd