

MANISH MUNIKAR

+1 (817) 917-1024 | munikarmanish@gmail.com | linkedin.com/in/munikarmanish | github.com/munikarmanish

PROFESSIONAL SUMMARY

Computer Science PhD candidate with 6+ years of experience in high-performance systems, ML infrastructure, distributed computing, and network stacks, with multiple research publications at top-tier systems conferences. Expertise spans Linux kernel development (C/C++, eBPF/XDP, Bash), applied machine learning (Python, PyTorch), and LLM inference (vLLM, LMCache). Proven track record applying ML at scale through three Amazon Applied Scientist internships. Currently exploring KV cache optimizations using CXL shared memory. Seeking an engineering role at the intersection of systems and machine learning.

EDUCATION

University of Texas - Arlington

PhD in Computer Science

January 2020 - December 2026

GPA: 4.0

- Researching in operating systems design and LLM inference infrastructure, supervised by Dr. Jia Rao.

Tribhuvan University - Pulchowk Engineering Campus

Bachelor's in Computer Engineering

October 2013 - August 2017

Grade: 80%

- Admitted with merit-based full scholarship, with a rank of 102 out of 14,000+ applications.

PROFESSIONAL EXPERIENCE

Amazon

Applied Scientist Intern

Bellevue, WA, USA

May 2024 - August 2024

- Replaced a rule-based return grading system (20–30% precision, 30–40% recall) by building an AutoGluon neural network classifier on M5 multi-modal embeddings, achieving up to **89% F1 score** using only **2-3K** labeled training samples.
- Benchmarked **5+** semi-supervised learning methods (BNLP, Self-learning, Correct & Smooth, UniMP, VAT) across performance, scalability, and training cost dimensions, identifying CleanLab + Self-learning as the most practical approach, avoiding the **45× training overhead** of VAT for large-scale deployment.

Amazon

Applied Scientist Intern

Sunnyvale, CA, USA

May 2022 - August 2022

- Designed and implemented a human label anomaly detection system using approximate neighbor algorithms (e.g., LSH) on multi-modal embeddings, achieving up to **94% FP yield** and **100% FN yield**, enabling generation of a cleaner training dataset by removing ~5% noisy labels across millions of human-audited product labels.

Amazon

Applied Scientist Intern

Remote

May 2021 - August 2021

- Developed multi-modal ML models to identify millions of inconsistent products (~2% of all active ASINs), achieving up to **95% precision** by ensembling CLIP, custom similarity model, and in-house collaborative filtering-based embedding models.
- Built and evaluated a scalable multi-signal product inconsistency detection pipeline (class, image, and text modalities) using Spark on AWS EMR and SageMaker, targeting millions of active ASINs across Amazon's global marketplaces.

Docsumo

Data Scientist

Kathmandu, Nepal

July 2018 - December 2019

- Engineered computer vision pipelines using Faster R-CNN, YOLO, and SSD, alongside custom template-based algorithms, to automatically identify and extract key fields from documents, achieving **95%+ accuracy** in production.
- Containerized and deployed end-to-end ML applications to production via Docker, Kubernetes, and CI/CD pipelines, delivering a scalable, high-availability architecture capable of handling real-world document processing loads.
- Drove all phases of the ML lifecycle — from data collection and annotation through model design, implementation, evaluation, and production deployment — ensuring tight feedback loops between experimentation and live performance.
- Joined as one of 3 founding data scientists, taking on full-stack ownership of the ML platform at an early-stage startup and helping establish the technical foundation from the ground up.

LIS Nepal Pvt. Ltd.

Software Developer

Lalitpur, Nepal

October 2017 - June 2018

- Engineered and optimized complex SQL queries within Oracle enterprise databases to deliver scalable Business Intelligence (BI) reports for global retail clients, driving data-driven decision-making across international markets.
- Designed and implemented robust ETL data integration pipelines leveraging big data technologies — including Hadoop, Hive, Sqoop, and Flume — to streamline large-scale data ingestion, transformation, and loading workflows.

RESEARCH PUBLICATIONS

- J. Lei, M. Munikar, H. Lu and J. Rao, “**Accelerating Packet Processing in Container Overlay Networks via Packet-level Parallelism**,” in *37th IEEE International Parallel and Distributed Processing Conference (IPDPS 2023)*.
- M. Munikar, J. Lei, H. Lu and J. Rao, “**PRISM: Streamlined Packet Processing for Containers with Flow Prioritization**,” in *Proceedings of the 42nd IEEE International Conference on Distributed Computing Systems (ICDCS 2022)*.
- J. Lei, M. Munikar, K. Suo, H. Lu and J. Rao, “**Parallelizing Packet Processing in Container Overlay Networks**,” in *Proceedings of the Sixteenth European Conference on Computer Systems (EuroSys 2021)*.

- M. Munikar, S. Shakya and A. Shrestha, “**Fine-grained Sentiment Classification using BERT**,” in *IEEE International Conference on Artificial Intelligence for Transforming Business and Society (AITB 2019)*.
- P. Dhakal, M. Munikar and B. Dahal, “**One-Shot Template Matching for Automatic Document Data Capture**,” in *IEEE International Conference on Artificial Intelligence for Transforming Business and Society (AITB 2019)*.

PROJECTS

Container Network Fastpath in Linux

2024 - 2025

- Designed and implementation a per-flow cache-based fastpath in the Linux network stack to accelerate the processing of container network packets, offering up to **121% higher throughput**, **61% lower latency**, and **43% lower CPU usage**, all without requiring any application or hardware modifications.
- Implemented in the Linux kernel with ~800 lines of C code.
- [Link to Project](#)

Automatic Data Extraction from Scanned Documents using One-shot Template Matching

2019

- Developed a one-shot template-matching algorithm to automatically extract data from scanned documents by combining visual (SVD-based) and textual (OCR + edit distance) similarity metrics, achieving **~90% accuracy**.
- Eliminated the need for large, labeled training datasets in invoice data capture, requiring only one annotated example per document type, by designing a region-proposal and keyword-projection pipeline that combines OCR, image correlation, and relative positioning to localize field values.

Movie Review Mining and Recommendation System

2017

- Designed and trained a deep learning recursive neural network model (RNTN) for sentiment analysis with **95% accuracy**.
- Built a web application using Python/Django where users can submit movie reviews and get personalized movie recommendations generated using a collaborative-filtering-based recommendation algorithm.
- [Link to Project](#)

Photocrypt: Image Steganography

2015

- Engineered a cross-platform text-to-image steganography application in C++ with both CLI and GUI (gtk) interfaces.
- Implemented a modified LSB (Least Significant Bit) substitution algorithm that embeds text data into image pixel values with password-based encryption, making hidden content visually imperceptible.
- Leveraged OpenCV for image I/O and pixel-level manipulation, and GTK+/gtkmm for building a responsive GUI with file dialogs, text editors, clipboard support, and dynamic image scaling.
- Packaged the project with a GNU Makefile supporting build, install, uninstall, and Doxygen documentation generation targets; published to the Arch Linux AUR for public distribution.
- [Link to Project](#)

SKILLS

Languages & Frameworks: Python, C/C++, Bash, SQL, eBPF/XDP, Spark, HTML/CSS

Systems & Infrastructure: Linux/Unix, Kernel programming, Docker, Kubernetes, AWS, Google Cloud Platform, DPDK, RDMA, Virtualization, NVIDIA DOCA, Distributed Systems

Machine Learning: PyTorch, Pandas, NumPy, Scikit-learn, OpenCV, vLLM, LMCache, NIXL