

Md Kamruzzaman (Methun)

mhmethun@gmail.com | (509) 336-9938 | Folsom, CA, USA | linkedin.com/in/mhmethun

Summary

Accomplished Postdoctoral Researcher with over 5 years of experience in applied machine learning, specializing in bioinformatics, healthcare, and physics. Proficient in developing advanced ML models, including SFT, PEFT, LoRA, and generative AI (VAE, WGAN, Diffusion), achieving up to 4x performance improvements and 95% efficiency gains. Skilled in Python, C++, and HPC environments (A100, Hopper GPUs). Experienced in leading teams, collaborating with industry partners like Google, and delivering actionable insights through data-driven solutions. Seeking to apply expertise in machine learning and data science to drive innovation.

Skills

- Machine Learning: SFT, PEFT, LoRA, VAE, WGAN, Diffusion Models, Ensemble Models, Hierarchical Models
- Data Science: ETL, Topological Data Analysis (TDA), Statistical Modeling, Graph-Based Algorithms
- Programming: Python, C++, NodeJS, D3.js, .NET, HTML/CSS, JavaScript, SQL
- Tools & Frameworks: Huggingface (dnaBERT2, molgpt), MongoDB, SLURM, PyTorch, TensorFlow, HPC Clusters, GCP
- Additional: Privacy-Preserving ML, Data Visualization, Full-Stack Development
- Soft Skills: Collaborative, Team Player, Team Leadership, Bias for Action, Results-Driven

Professional Experience

Postdoctoral Researcher (Applied Machine Learning), Sandia National Laboratories, Livermore, CA May 2023–Present

- Accelerated 5'UTR DNA sequence analysis by 4x, reducing training time, using SFT and PEFT with LoRA on Huggingface/dnaBERT2 across 10M samples on a 10-node A100 GPU cluster.
- Enhanced lab efficiency by 95% in molecular design throughput by training Huggingface/molgpt on 300k drug-like SMILES strings, fine-tuning with 1,000 candidate molecules, and generating 10k novel molecules.
- Discovered two industry-manufacturable novel molecules through an ensemble-based predictive model and rule-based recommendation system from 10k generated molecules.
- Developed a recommendation system for FDA-approved drug candidates using human-virus PPI, transcriptomic, and viral gene-drug interaction data for drug repurposing.
- Reduced anomaly detection execution time by 5x on wearable health data using multilayered anomaly detection on NVIDIA Hopper GPU.
- Improved predictive accuracy for a 45,000-population digital twin by advancing generative AI models (VAE, WGAN, Diffusion), minimizing model uncertainty.
- Enhanced lattice QCD matrix element modeling accuracy at short distances and large momentum using generative AI techniques.

Postdoctoral Researcher (Applied Machine Learning), University of Virginia, Charlottesville, VA Oct 2020–May 2023

- Streamlined data preparation for model training by developing an ETL process to integrate daily data from two hospitals into a common data model.
- Improved model AUC by 15% by constructing a 10GB temporal network from 10 years of patient-provider contact data, extracting latent features stored in MongoDB for optimized querying.
- Enhanced model accuracy by 7% on unbalanced EHR datasets by designing a novel hierarchical model algorithm for HIPAA-compliant data.
- Strengthened healthcare application performance by implementing privacy-preserving ML solutions in collaboration with the Google research team.
- Identified infection sources and missing cases using an agent-based model on an HPC cluster with SLURM, leveraging claim and prescription data.

Research Assistant (Data Science), Washington State University, Pullman, WA Aug 2015–Sep 2020

- Developed Hyppo-X, a C++-based data analysis framework with integrated Topological Data Analysis (TDA) algorithms and interactive visualization tools, uncovering complex patterns in high-dimensional datasets.
- Enabled breakthroughs in maize phenomics research by engineering a graph-based data analysis algorithm to extract scientific hypotheses from structural properties.
- Delivered actionable insights for agricultural research by implementing statistical modeling for genotypic analysis under environmental stress.
- Enhanced clinical insights by integrating NodeJS and D3.js for advanced data visualization of hospital patient trajectories.

Software Development Engineer, KBGroupUK & NRG, Dhaka, Bangladesh May 2009–Jul 2015

- Built full-stack e-commerce websites using .NET, HTML/CSS, and JavaScript, increasing ticket sales by 13%.
- Led a 6-person team from 2013 to 2015, driving 5% year-over-year revenue growth.

Internship Experience

Applied Scientist Intern, Amazon, Seattle, WA Jun 2019–Sep 2019

- Optimized ML models (ARIMA, LSTM) for sales predictions, improving accuracy by 3%.

Research Intern, Pacific Northwest National Laboratories, Richland, WA Jul 2018–Aug 2018

- Engineered C++ solutions with SPDLOG, resolving Python asynchronous bottlenecks.

Education

Ph.D., Computer Science, Washington State University (WSU), Pullman, WA Aug 2015–Aug 2020

B.Sc., Computer Science and Engineering, Bangladesh Univ. of Eng. and Tech. (BUET), Dhaka, Bangladesh Mar 2004–Mar 2009

Additional Information

- Proficient in collaborating across interdisciplinary teams and presenting complex technical results to diverse stakeholders.
- Available for relocation and authorized to work in the United States without sponsorship.