Jillur Rahman Saurav

Machine Learning Engineer / Applied Scientist
Email: jrsaurav.contact@gmail.com| Website: facevoid.github.io
LinkedIn: linkedin.com/in/jillur-rahman-saurav | GitHub: github.com/facevoid
Google Scholar: scholar.google.com

Summary

ML engineer specializing in generative models for medical imaging and production NLP systems. Built virtual staining pipelines achieving SOTA on clinical benchmarks and scaled search infrastructure processing thousands of articles daily from 5,500+ sources. Published in NEJM AI, IEEE, and Springer venues with 9+ papers.

Technical Skills

Languages: Python, Java, SQL
Machine Learning: Foundation models, fine-tuning, representation learning, contrastive learning, model compression, multi-task learning
Deep Learning: PyTorch, TensorFlow, Keras, GANs, VAEs, Diffusion Models, Transformers
Computer Vision: Image synthesis, image-to-image translation, virtual staining, contrastive learning, image retrieval
NLP: Transformers, BERT, GPT, speech recognition, NER, POS tagging, low-resource NLP
ML Tooling: scikit-learn, XGBoost, Weights & Biases, OpenCV, scikit-image
Systems & Infrastructure: Docker, Linux/Unix, Git, HPC clusters, embedded inference (Jetson, ARM)
Data Engineering: Elasticsearch, Apache Solr, Django, MongoDB, PostgreSQL

Industry Experience

- **Software Engineer — ML & Search Infrastructure** Nov 2017 – Dec 2020
  Pipilika *Dhaka, Bangladesh*
  - **Production NLP Systems**: Built and deployed end-to-end Bangla speech recognition, POS tagging, and NER pipelines, achieving 3.96% WER, 93.86% tagging accuracy, and 86% F1 across 22K+ annotated sentences. Served as core NLP backend for search and voice applications.
  - **Scalable Search Infrastructure**: Engineered distributed ingestion, indexing, and retrieval pipelines aggregating news from 5,500+ sources using Elasticsearch and Solr. Improved search relevance via query expansion, spelling correction, and n-gram language models.
  - **Applied ML at Scale**: Led ML analytics on public health datasets with 535,000+ participants (COVID-19 screening, dengue surveillance), delivering actionable insights for national health monitoring initiatives.

Research Experience

- **Graduate Research Assistant** Jun 2022 – Present
  University of Texas at Arlington, Luber Lab *Arlington, TX*
  - **Virtual Staining Pipeline**: Built foundation-model-guided virtual staining system (UNIStainNet) translating H&E to multiple IHC markers (HER2, Ki67, ER, PR) in a single model, achieving SOTA on MIST and BCI benchmarks.
  - **Generative Models for Medical Imaging**: Developed GAN- and VQGAN-based pipelines for cross-domain image synthesis (mIF to H&E), reducing experimental staining cost while preserving nuclei segmentation and tissue classification performance. Published at MICAD 2025 and IEEE CIBCB 2023.
  - **Biomedical Signal Compression**: Contributed to VAE-based compression pipeline achieving 1:293 ratio on EEG spectrograms while retaining 91% seizure detection accuracy; focused on model architecture design and spectrogram representation.
  - **Clinical-Scale AI Evaluation**: Co-authored NEJM AI 2024 study benchmarking histopathology search engines against real clinical workflows, establishing quantitative readiness criteria for deployment.

Education

- **University of Texas at Arlington** Arlington, Texas
  Ph.D. in Computer Science *Aug 2021 – May 2026 (Expected)*
  Advisor: Dr. Jacob M. Luber
  Dissertation: *Generative Imaging for Computational Pathology*

- **Shahjalal University of Science & Technology** Sylhet, Bangladesh
  B.Sc. (Engg.) in Computer Science *Jan 2013 – Jun 2018*

Selected Publications

- **Saurav, J.R.**, et al. UNIStainNet: Foundation-Model-Guided Virtual Staining of H&E to IHC. **arXiv preprint**, 2026.
- **Saurav, J.R.**, et al. Histopathology Slide Indexing and Search — Are We There Yet?. **NEJM AI**, 2024.
- **Saurav, J.R.**, et al. Cross-Domain Image Synthesis: Generating H&E from Multiplexed Biomarker Imaging. **MICAD**, Springer LNCS, 2025.

Additional publications available at: Google Scholar

Honors & Leadership

- **Google Research Computer Science Research Mentorship Program (CSRMP)** Class A Recipient
  Google Research *Jan 2023 – Apr 2023*
  - Selected for competitive research mentorship program pairing PhD students with Google researchers

- **Teaching Assistant — Artificial Intelligence (CSE 5360)** Aug 2021 – May 2022
  University of Texas at Arlington *Arlington, TX*
  - Instructed 120+ graduate students on machine learning fundamentals and neural networks