

Nafiz Ahmed

+88-01767457527 | nafizahmed273273@gmail.com | nafishsy.github.io |  LinkedIn |  GitHub |  Scholar

Basundhara Residential Area, Dhaka, Bangladesh

PROFESSIONAL SUMMARY

Machine Learning Engineer with 3+ years of combined experience in Software Engineering, Computer Vision & Research. Currently working on Agentic Solutions for various private sectors like Banking, Travel Agencies.

EDUCATION

American International University - Bangladesh (AIUB)

July 2019 - August 2024

Dhaka, Bangladesh

Bachelor of Science in Computer Science and Engineering

- CGPA: 3.93/4.00 (Top 2% of graduating class)

- Medals: Magna Cum Laude (22nd Convocation [22nd February, 2025])

- Awards: Recipient of 4 consecutive Dean's Awards for academic excellence

PROFESSIONAL EXPERIENCES

Periscopelabs

October 2025 – Present

Banani, Bangladesh (Onsite)

- Machine Learning Engineer (Mid-Level)
- Architected and deployed an enhanced defaulter warning system for DBH Bank, focusing on individual clients. This system utilized a traditional **time-series analysis model (Prophet model)** and was successfully integrated into the bank's existing Customer Relationship Management (CRM) project, leading to improved proactive risk mitigation.
 - Engineered a comprehensive RAG pipeline featuring OCR preprocessing, **YOLOv5n** for pattern annotation and column detection (**90% accuracy**), advanced Hybrid Chunking (semantic + LLM-guided), **ChromaDB** vector retrieval, and **GPT-4** generation (temperature 0.1) for deterministic legal outputs.
 - Leveraged cutting-edge tools including **n8n with self-hosted Python nodes**, **LangChain/LangGraph**, **CrewAI**, **FAISS**, and **OpenAI APIs** to build scalable, production-ready multi-agentic solutions.

MetroSofts

January 2024 – October 2025

Gothenburg, Sweden (Remote)

Machine Learning Engineer

- Engineered multi-modal AI pipeline combining dual-YOLOv5n nano detectors (87% accuracy, 120 FPS on GPU) with LLaVaNext vision-language model for automated UI element annotation, generating hierarchical JSON structures which eliminates 80% of manual annotation workload
- Integrated open-source Rasa with custom API endpoints enabling digital twin method calls and intelligent test failure orchestration with optional human-in-the-loop

Shapla Infosys

Junior Software Engineer

January 2023 – December 2023

Dhaka, Bangladesh

- Built and deployed custom **web applications** using **Laravel**, **MySQL**, and **React**, increasing client satisfaction by 40%.
- Designed 200+ **UI/UX prototypes** in **Figma**, improving user engagement by 35%.
- Developed responsive front-end interfaces achieving 98% on Google **Lighthouse** performance metrics.
- Implemented a scalable **Laravel/MySQL backend**, improving query response time by 45%.

RESEARCH EXPERIENCES

ELITE Lab

September 2025 – Present

New York, USA (Remote)

Researcher

- Conducting research on **Computer Vision** and **Large Language Models (LLMs)**; contributing to three academic papers under preparation.
- Developed a modified **U-Net** model inspired by MedT architecture for **retinal vessel segmentation** and a custom **YOLOv12** architecture for **brain tumor localization**.
- Designed **Guardrails** for **child-safe LLMs**, focusing on prompt safety and ethical AI alignment.

Advanced Machine Intelligence Research Lab (AMIR Lab)

September 2023 – March 2024

Dhaka, Bangladesh

Research Intern (Hybrid)

- Published a peer-reviewed literature review on **Natural Language Processing (NLP)**, **Deep Learning**, and **Human-Agent Interaction**.
- Applied the **PRISMA** methodology for a systematic and reproducible literature survey.
- Led a 3-member research team, overseeing literature analysis, experimentation, and manuscript preparation.
- Acquired initial research skills in **systematic review**, **critical evaluation**, and **academic writing**.

PROJECTS

• Early Warning System (EWS) for DBH Bank

April 2025 - Current

XGBoost, Linear Regression, Prophet, Vue.js, Streamlit, VPS

- Inherited and refactored a financial early warning system from a former employee, transitioning the frontend from Streamlit to Vue.js for improved scalability and user experience.
- Implemented predictive models using XGBoost, Linear Regression, and Prophet for time-series forecasting to identify potential financial risks and anomalies.
- Successfully deployed the refactored application to VPS infrastructure, ensuring robust production environment and continuous availability for bank operations.

• Retinal Vessel Segmentation Model

April 2025 - Present

TensorFlow, Keras, SegNet, MedT, Transformer Block, Multi-Scale Feature Extraction

[Inference Demo (Few Shot Model)]

- Developed a novel few-shot learning architecture requiring only 5% of available training data while achieving exceptional generalization.
- On Fullshot Outperformed existing solutions with metrics of 92.72% Dice coefficient (F1 score for segmentation task), 86.58 Jaccard Index (IoU).
- Innovated an optimized Medical Transformer (MedT) architecture with hybrid attention mechanisms, multi-scale feature extraction, transformer blocks.

• PassPorted: Identity-Consistent Generative ID System

Nov 2025 - Present

Diffusion Models, PyTorch, ControlNet, HuggingFace Spaces

[MVP Demo]

- Developing an AI-powered MVP utilizing Latent Diffusion Models to transform unconstrained input images into standardized, full-frontal pose passport photos.
- Implementing advanced pose-control mechanisms to ensure strict alignment compliance while maintaining high facial identity fidelity.
- Deployed the inference pipeline on HuggingFace Spaces, streamlining the workflow for automated professional ID photo generation.

• Enterprise-Scale Bengali Legal RAG System

Oct 2025 - Present

RAG, Fine-tuned YOLOv8, FastAPI, ChromaDB, React, Redis, Ragas

Inference

- Architected high-throughput ingestion pipeline for **50,000+ legal precedents**; achieved **98%** layout fidelity in complex dual-column Bengali PDFs using **fine-tuned YOLOv8** and recursive semantic chunking.
- Engineered retrieval system evaluated on the **Ragas framework**, attaining **0.95 Faithfulness** and **0.92 Context Recall**; reduced hallucinations by 40% via hybrid search (Dense + BM25) and metadata filtering.
- Deployed fault-tolerant architecture with **Redis semantic caching** to minimize latency; implemented continuous evaluation loops to dynamically optimize GPT-4 temperature and prompts based on live feedback.

• Custom YOLOv12 for Multi-Class Brain Tumor Detection

Sep 2025 - Oct 2025

Ultralytics YOLOv12, Attention Mechanisms (SE, SimAM, CoordAtt, ECA), DenseBlock, Grad-CAM Explainability

- Engineered a custom YOLOv12 backbone with DenseBlock integration and hybrid attention modules, reducing parameters by 15% while boosting inference speed by 20% on GPU.
- Achieved state-of-the-art performance on a 4-class brain tumor dataset (Glioma, Meningioma, No Tumor, Pituitary) with 94.35% mAP@50, surpassing YOLOv12 baselines by 95.14% in mAP@50.
- Incorporated multi-class Grad-CAM for interpretable AI, yielding 0.72 IoU alignment with ground-truth regions and enabling clinical trust through tumor-specific attention visualization.

• AI System for Early Detection & Staging of Pediatric Nephroblastoma

2025

PyTorch, U-Net, DeepLabV3 (ResNet-50), ResNet18/34

- Developed a dual-task deep learning system for nephroblastoma (Wilms' tumor) detection and semantic segmentation using **35,000+ 2D grayscale CT images**.
- Implemented advanced segmentation models (**DeepLabV3** with ResNet-50 backbone) and optimized training using a combined **Dice Loss** and **Binary Cross-Entropy** function.
- Achieved superior tumor localization with the DeepLabV3 model, securing a Dice score of 90.81 and Sensitivity of 96.64, demonstrating high reliability for clinical staging.

PUBLICATIONS & ONGOING RESEARCH

- J.1 Ahmed, N., et al. (2024). **Deep Learning-Based Natural Language Processing in Human-Agent Interaction: Applications, Advancements and Challenges**. *Natural Language Processing*, 9, 100112. DOI: [10.1016/j.nlp.2024.100112](https://doi.org/10.1016/j.nlp.2024.100112)
- J.2 Ahmed, N., et al. (2023). **Two Proposed Models for Securing Data Management for Enterprise Resource Planning Systems Using Blockchain Technology**. *International Journal of Information Engineering and Electronic Business*, 15(6), 52-66. DOI: [10.5815/ijieeb.2023.06.02](https://doi.org/10.5815/ijieeb.2023.06.02)

- W.1** **Submitted:** *Hybrid MedT: A Data-Efficient Transformer-Unet Framework for Retinal Vessel Segmentation.* Developing a lightweight architecture combining HMSDM encoders and bottleneck transformers to optimize segmentation on small retinal datasets (FIVES dataset).
- W.2** **Submitted:** *Explainable Multi-Modal Brain Tumor Localization using Attention-Guided YOLOv12.* Extending custom object detection pipelines with Grad-CAM and attention mechanisms to enhance interpretability in clinical tumor diagnostics.
- W.3** **Ongoing:** *AI-Driven Staging of Pediatric Nephroblastoma in Low-Resource Settings.* (Based on your Pediatric Oncology Project). Engineering a dual-stream ResNet34 and DeepLabV3 system for the detection and semantic segmentation of Wilms' tumor on CT scans to reduce diagnostic delays.

EXTRACURRICULAR ACTIVITIES

• Researcher & Administration — SPEAR Foundation

Supportive Palliative, Emotional Assistance & Research Foundation

Jan 2024 – Present

Dhaka, Bangladesh

- Organized nutritional support, caregiver assistance, and play therapy for pediatric patients.
- Conducted parent education workshops to improve treatment adherence and family support.

TECHNICAL SKILLS

Machine Learning & Deep Learning: PyTorch, TensorFlow, Keras, Scikit-learn, Transformers, CNNs, Representation Learning

Computer Vision: Image Classification, Object Detection, Semantic Segmentation, Medical Image Analysis, Grad-CAM

Generative AI & LLMs: Retrieval-Augmented Generation (RAG), Prompt Engineering, Multi-Modal Models

Data & Systems: Vector Databases (FAISS, ChromaDB), Redis, ETL Pipelines

MLOps & Deployment: Docker, Cloud Platforms (AWS, Azure), CI/CD

Programming: Python, PHP (Laravel), C++, JavaScript, SQL

CERTIFICATIONS & ADDITIONAL INFORMATION

- **Competitive Programming - Beginner** - Devskill (2021)
- **Languages:** English (C1, IELTS Band 7.5), Bengali (Native)

REFERENCES

Md Kishor Morol

Adjunct Professor, Department of Computer Science
St. Thomas University, Miami, Florida, USA
PhD Student, Cornell University
Email: mmorol@stu.edu, kishor@elitelab.ai
Research Advisor @ Elite Lab

Tariqul Islam

Senior Infotainment Developer, Volvo Group
Email: tariqul@metrossofts.se
Phone: +46 70 025 83 96
Team Lead @ Metrossofts

Prof. Dr. M. Firoz Mridha

Professor, Faculty of Science (FST), AIUB
Email: firoz.mridha@aiub.edu
Phone: +880-1674791594
Research Advisor @ AMIR Lab

Md Mohsin Kabir

Lecturer (On leave), CSE, BUBT
PhD Student, Mälardalen University, Sweden
Email: *Email: md.mohsin.kabir@mdu.se*
Phone: +46 73 662 13 77
Research Advisor @ AMIR Lab