Rafi Majid

Efficient & Reliable AI · LLM Compression · Multimodal Learning · Medical AI

rafi.majid93@gmail.com | | Dhaka, Bangladesh

Portfolio | Google Scholar | LinkedIn | GitHub | LeetCode

Research Profile

Artificial Intelligence researcher focused on efficient, reliable, and scalable machine learning systems. Research interests include deep learning, multimodal AI, knowledge distillation, model compression, medical AI, and representation learning. Experience spans vision-language learning, climate forecasting, graph learning, and transformer-based code intelligence.

Research Interests

- Efficient Deep Learning and Model Compression
- Knowledge Distillation and Representation Learning
- Large Language Models and Multimodal Learning
- Medical Image Analysis and Healthcare AI
- Code Intelligence and Software Vulnerability Detection
- Graph Neural Networks and Higher-Order Learning

Education

Bachelor of Science in Computer Science and Engineering 2023

North South University, Dhaka, Bangladesh

CGPA: 3.49/4.00

Publications

Bangla-Bayanno: A 52K-Pair Bengali Visual Question Answering Dataset with LLM-Assisted Translation Refinement , *arXiv Preprint, 2025* [Paper]

Mohammed Rakibul Hasan, ***Rafi Majid****, Ahanf Tahmid*

- Developed a 52K-pair Bengali visual question answering dataset with LLM-assisted translation refinement for low-resource vision-language research.

A Deep Learning-Based Study for Cyclone Track Forecasting: Comparative Analysis Using Historical Data from the Bay of Bengal , *Springer Nature, 2024* [Paper]

Rafi Majid*, Akmam Hasan, Shayrey Mostarin, Kazi Rabiul Alam, Rashedur M Rahman*

- Developed and evaluated LSTM, ConvLSTM, and GRU architectures for cyclone trajectory forecasting using Bay of Bengal historical data.

When Does Higher-Order Representation Improve Protein-Complex Recovery? A Leakage-Controlled, Matched-k Comparison of Graph and Penalized Hypergraph Spectral Clustering , *Preprint, 2026* [Paper]

Md Asad Kazi Hafiz, ***Rafi Majid****, Md Tanjeelur Rahman Labib, Ahsanur Rahman*

- Investigated higher-order representations for protein-complex recovery through leakage-controlled graph and hypergraph spectral clustering.

Research Experience

Efficient AI Systems and Knowledge Distillation

- Developing **CoGS-KD (Confidence-Gated Structural-Contrastive Knowledge Distillation)**, a framework for compressing transformer-based models while preserving teacher-level representation capability.
- Investigating confidence-aware knowledge transfer, structural supervision, and contrastive objectives for improving the efficiency of compact neural architectures.

Transformer-Based Code Intelligence and Software Security

- Developing CodeBERT-based vulnerability detection models for automated source-code security analysis.
- Designing knowledge distillation strategies to transfer contextual code representations from large transformer

teachers into lightweight student models.
- Studying whether compressed models can maintain vulnerability detection capability while reducing computational complexity.

Medical Vision-Language Intelligence and Reliability Monitoring

- Developing **TRUST-VL-CXR**, a decoupled reliability monitoring framework for chest X-ray-report alignment under hard cross-patient mismatch conditions.
- Designing compatibility monitors trained independently from frozen vision-language task predictors to evaluate image-report consistency.
- Analyzing multimodal robustness through representation drift, controlled perturbations, and uncertainty behavior in medical AI systems.

Teaching Experience

Teaching Assistant

North South University

2022 - 2025

Technical Skills

Machine Learning:

Deep Learning, Transformers, Large Language Models, Knowledge Distillation, Representation Learning, Graph Neural Networks

Computer Vision and NLP:

Medical Image Analysis, Multimodal Learning, Natural Language Processing, Vision-Language Models, Code Intelligence

Programming:

Python, C++, Java, JavaScript, C#

Frameworks and Tools:

PyTorch, TensorFlow, Hugging Face Transformers, CUDA, Git, Linux, LaTeX

Industry Experience

Junior Software Engineer

TECH NEX SOLVE

2024 - Present

- Developed scalable backend applications using Spring Boot, React, and MySQL.
- Designed enterprise systems including authentication, inventory, and transaction management modules.
- Built Unity-based simulation environments using C#.

Selected Engineering Projects

- Developed full-stack software systems integrating backend services, databases, and modern frontend frameworks.
- Built interactive simulation environments using Unity and C#.
- Designed software solutions emphasizing scalability, maintainability, and performance.

Languages

- **Bengali:** Native proficiency
- **English:** Professional proficiency (CEFR C1)