Shah Nawaz Haider

nawazwithai@gmail.com | Github | Google Scholar | LinkedIn | Website

RESEARCH INTERESTS

Foundation Models, World Models, Large Language Models, State Space Models, Reasoning, Multimodal Learning

EDUCATION

University of Science and Technology Chittagong (USTC) — **Chittagong, Bangladesh**
Bachelor of Science in Computer Science & Engineering — *Graduation Date – May 2026*

CGPA: 3.72/4.0

Relevant Coursework: Discrete Mathematics & Probability, Linear Algebra & Matrices, Data Structures & Algorithms, Database Management Systems, Artificial Intelligence & Machine Learning, Operating Systems & Computer Networks, Software Engineering & Web Technologies, Digital Logic Design.

Thesis: QuantFlow - Federated Mamba-Based Post-Transformer Foundation Model for Time-Series Forecasting
Supervisor: Dr. Hadaate Ullah

SELECTED PUBLICATIONS

[1] **Haider, S. N.**, Austin, S., Barua, A., Shawon, S. M., and Ullah, H. *QuantFlow: A Federated Mamba-Based Post-Transformer Foundation Model for Time-Series Forecasting.* (arXiv:2607.02632)

[2] Shawon, S. M., **Haider, S. N.**, Barua, A., Austin, S., Adan, I. A., Hossain, M. S., and Zubair, H. T. *Hybrid CNN-LSTM Model for Urban Energy Load Forecasting with IGA-XAI for Smart Grids.* Results in Engineering (Elsevier, Q1), 2025. DOI: 10.1016/j.rineng.2025.107245

[3] Barua, A., **Haider, S. N.**, Austin, S., and Shawon, S. M. *Urban Energy Load Dataset of Chattogram: Peak and Off-Peak Variability Insights.* Data in Brief (Elsevier), 2025. DOI: 10.1016/j.dib.2025.112245

[4] **Haider, S. N.**, Barua, A., A., Austin, and Shawon, S. M. *Lightweight Federated and Explainable AI Enhanced Crop Yield Forecasting in Heterogeneous Agricultural Systems.* IEEE ICCIT, 2025.
DOI: 10.1109/ICCIT68739.2025.11491605

[5] Shawon, S. M., Dutta, P., **Haider, S. N.**, Barua, A., Austin, S., and Adan, I. A. *Agri-Light-Fed: Lightweight Federated Learning with Explainable AI for Efficient and Privacy-Preserving Crop Yield Prediction.* IEEE ICCIT, 2025. DOI: 10.1109/ICCIT68739.2025.11491295

[6] Shawon, S. M., **Haider, S. N.**, Barua, A., Austin, S., and Noor, K. R. *Predictive Modelling for Multi-Location Deep Learning-Based Load Forecasting: An Integrated Approach.* IEEE ICECE, 2024.
DOI: 10.1109/ICECE64886.2024.11024879

[7] Austin, S., **Haider, S. N.**, Shil, R., Lija, N. S., and Noor, K. R. *Bi-LSTM-RF Based Hybrid Deep Learning Approach to Identify Level of Depression from Bengali Facebook Status.* IEEE ICCIT, 2024.
DOI: 10.1109/ICCIT64611.2024.11022340

[8] Shawon, S. M., **Haider, S. N.**, Chakma, A., Alam, M. W., Islam, M. T., and Rana, M. F. *DeepFlowNet: Deep Learning-Based Daily Water Flow Forecasting of Test River.* IEEE PEEIACON, 2024.
DOI: 10.1109/PEEIACON63629.2024.10800657

[9] Shawon, S. M., **Haider, S. N.**, Austin, S., and Barua, A. *HybridNet: ResNet50-GRU Integration for Enhanced Rice Leaf Disease Classification in Precision Agriculture.* Springer LNNS (MIET), 2024.
DOI: 10.1007/978-981-96-2721-9_39

Complete publication list and citation metrics available on Google Scholar.

RESEARCH & INDUSTRY EXPERIENCE

Quantropy **Chittagong, Bangladesh**
AI/ML Instructor *June 2026 – Present*

- Designed and taught an advanced AI curriculum spanning machine learning, deep learning, large language models (LLMs), foundation models, and modern generative AI.
- Delivered lectures and hands-on workshops covering supervised learning, representation learning, transformers, retrieval-augmented generation (RAG), AI agents, and efficient model development.
- Mentored students on research methodology, experimental design, model evaluation, and best practices for reproducible AI research and real-world ML applications.

Synnax AI **Dubai, United Arab Emirates (Remote)**
Data Scientist *Aug 2024 – May 2026*

- Engineered and deployed end-to-end ML pipelines for credit risk and default prediction, processing SEC filings and market data for 16,000+ public companies.
- Built LLM-powered AI agents to extract critical risk signals from 10-K, 10-Q, and 8-K filings, automating identification of high-risk and defaulted entities.
- Transformed large-scale unstructured regulatory data into production-ready features for downstream modeling and analytics.
- Worked with vector databases, PostgreSQL, Redis, and MongoDB in a production MLOps environment.

Omdena **Milan, Italy (Remote)**
Data Scientist (Part-time) *May 2024 – July 2024*

- Collected, cleaned, and preprocessed urban agriculture datasets, including environmental, spatial, and crop-related variables for machine learning workflows.
- Conducted exploratory data analysis (EDA) to uncover trends, correlations, and actionable insights supporting sustainable urban farming initiatives.
- Developed and evaluated machine learning models for crop optimization, resource allocation, and data-driven agricultural decision-making.
- Contributed to research documentation, model evaluation, and presentation of insights to project stakeholders.

University of Science and Technology Chittagong (USTC) **Chittagong, Bangladesh**
Peer Tutor (Part-time) *September 2023 – November 2023*

- Mentored undergraduate students in programming, algorithms, and problem-solving through individualized tutoring and hands-on coding sessions.
- Strengthened students' computational thinking, debugging, and analytical skills in introductory computer science courses.

NOTABLE PROJECTS

Bengali BPE Tokenizer [GitHub]

- Designed and trained a Byte Pair Encoding (BPE) tokenizer from scratch for Bengali and other non-Latin scripts and achieved 7.33× fewer tokens compared to GPT-2.
- Processed 3.4 billion characters from the CC-100 Bengali corpus to build a fixed-size vocabulary of 8,000 subword units.
- Implemented custom handling for multi-byte UTF-8 characters and complex conjuncts, significantly reducing token inflation and inference costs for under-resourced languages.

Lunar Seismic Data Visualization - NASA Space Apps Challenge [GitHub]

- Built an interactive web application to visualize lunar seismic data directly sourced from NASA's Planetary Data System (PDS).
- Integrated spatial data overlays to map mineral composition and Bouguer gravity disturbances against seismic event locations.

TECHNICAL SKILLS

Programming Languages: Python, Java, JavaScript, C/C++, MATLAB, SQL
ML / DL Frameworks: PyTorch, TensorFlow, Keras, Scikit-learn, XGBoost, Hugging Face Transformers
LLM & Agentic AI: LangChain, Retrieval-Augmented Generation (RAG), Vector Databases, LoRA/QLoRA
Data & Databases: Pandas, NumPy, Matplotlib, Plotly, PostgreSQL, MongoDB, Redis