Sanju Basak

Dhaka, Bangladesh
sanjubasakndc18@gmail.com Sanju Basak Sanju-Basak

RESEARCH INTERESTS

- Natural Language Processing in Low-Resource Settings
- Multilingual and Cross-Lingual Natural Language Processing
- Human AI and Human Computer Interaction

EDUCATION

Bangladesh University of Engineering and Technology **Apr 2019 – July 2024**
B.Sc in Computer Science and Engineering

- ***CGPA:*** *3.92 on a scale of 4.00*
- ***Position:*** *Ranked* 18^{th} *in a class of 123 students*

RESEARCH EXPERIENCES

Effective Retrieval-Augmented Generation for Open Domain Question Answering in Bengali [Under Review on COLING 2025]
May 2023 – Sept 2024

- Developed the first-ever Bengali **Retrieval-Augmented Generation (RAG)** pipeline and benchmarked two Bengali open-domain question-answering datasets (**SQuAD_BN** and **BanglaRQA**) using six state-of-the-art embedding models and three retrieval methods, encompassing both sparse and dense approaches.
- Evaluated the performance of three LLMs' response with and without our RAG pipeline.
- Analyzed LLMs' factual response capabilities on Bengali-region specific data in comparison to global data.
- **Tools and Technology:** Python (LlamaIndex, Ragatuille), OpenAI GPT API, Gemini API, Hugginface
- **Supervisor:** Dr. Rifat Shahriyar, CSE, BUET , Abhik Bhattacharjee, GRA, BUET

Multi-ToM: Evaluating Multilingual Theory of Mind Capabilities in Large Language Models [Under Review on COLING 2025]
Aug 2024 – Sept 2024

- Presented a multilingual ToM dataset translated from a bilingual version, covering seven major languages, and develop an additional culturally nuanced dataset.
- Evaluated six state-of-the-art LLMs on both datasets to assess the impact of cultural relevance on their social reasoning abilities.
- Our analysis explored how LLM performance varies across different languages and cultural contexts.
- **Tools and Technology:** Python (Ollama, RAG), OpenAI GPT API, Kaggle
- **Supervisor:** Dr. Rifat Shahriyar, CSE, BUET, Abhik Bhattacharjee, GRA, BUET

Exploring Post-Mortem Neural Signal Processing: Uncovering Computational Potentials in Deceased Animal Brains [Under Review on CHI 2025]
Mar 2021 – Present

- Performed microcontroller interfacing, visualization, and regression analysis
- **Tools and Technology:** Arduino, Python
- **Supervisor:** Dr. A. B. M. Alim Al Islam, Professor, CSE, BUET

WORK EXPERIENCES

Bangladesh University of Engineering & Technology **Aug 2024 – Present**
Adjunct Lecturer, Department of CSE

Course Instructor on:

• Signals and Linear Systems • Operating System • Structured Programming Language

Bangladesh University of Engineering & Technology **Sept 2024 – Present**
Research Assistant

- **Supervisor:** Prof. Dr. Md. Mostofa Akbar

ResTech-AI (USA-Based Software Company) **July 2024 – Sept 2024**
Software Engineer - Generative AI

TECHNICAL SKILLS

- **Programming Languages:** Python, C/C++, C#, Java, JavaScript, SQL, HTML/CSS, Bash, LaTeX
- **Libraries/Frameworks:** PyTorch, Keras, TensorFlow, LangChain, Scikit-learn, LlamaIndex, Node.js
- **Tools/Platforms:** Docker, Git, Google Colab, kaggle, Visual Studio Code, Linux

ACADEMIC PROJECTS

Bangla Handwritten Character Recognition **Feb 2024**

Python , Pytorch, Ultralytics YOLO, EfficientNet, Kaggle

- A machine learning project to detect bangla handwritten characters from a form fields images.
- Finetuned **YOLOv8** with **WTW dataset**
- Finetuned EfficientNet with **BanglaLekhaIsolated dataset** with **93.5%** accuracy.

Machine Learning Algorithms and Neural Network from Scratch **Dec 2022**

Python, Numpy, Scikit-learn, Pandas, Seaborn

- Implemented matrice transformation and image reconstruction using singular value decomposition
- Implemented adaboost algorithm with logistic regression, exploratory data analysis and preprocessing techniques
- Trained and evaluated a feed-forward neural network from scratch using only numpy
- Implemented PCA and clustering with expectation-maximization algorithm on gaussian mixture models

OCR Based CRVS Form Digitalization **Sep 2023**

Tesseract, ReactJS, NodeJS, PostgreSQL

- A web platform to process handwritten CRVS(Civil Registration and Vital Statistics) forms through OCR, facilitating accurate extraction, digitalization, and correction of information.
- Participated in box detection of various form fields and designing the front end.

Honors & AWARDS

- **Dean's List Award:** Bangladesh University of Engineering and Technology **2019, 2021**
- **Student Research Poster Champion:** 8th **NSysS 2021** Conference [Poster Link] **2021**
- **University Merit Scholarship:** Bangladesh University of Engineering and Technology **2019, 2021**
- **Board Merit Scholarship:** Government of Bangladesh **2016 - 2023**

LEADERSHIP EXPERIENCES & VOLUNTARY WORKS

- **BUET Cyber Security Club** **2023-2024**
  Director (Logistics)
  Organized national-level contests, hackathon and capture-the-flag competitions [Event Link]

- **BADHAN, BUET Zone** **2019-2024**
  Vice President of Panel 2024
  Badhan is a voluntary blood donor organization. I managed and developed an app consisting of 5,000 donors database, coordinated numerous blood donations, and have personally donated 11 times. [App Link]

- **Satyen Bose Science Club, BUET** **2022-2024**
  Assistant General Secretary

NOTABLE COURSES

- Artificial Intelligence
- Machine Learning
- Computer Networks
- Numerical Methods
- Probability & Statistics
- Computer Architecture
- Operating Systems
- Discrete Mathematics

REFERENCES

- **Dr. Rifat Shahriyar** **Email: rifat@cse.buet.ac.bd**
  Professor, CSE, BUET [Thesis Advisor]

- **Dr. A. B. M. Alim Al Islam** **Email: alim_razi@cse.buet.ac.bd**
  Professor, CSE, BUET [Research Advisor]