

MEHERAJ HOSSAIN

418, Makki Mashjid Goli, West Rampura, Dhaka-1219

+88 01853594333 meherajj@gmail.com [meheraj-hossain](https://meheraj-hossain.github.io) github.com/meheraj2325 meheraj2325.github.io

RESEARCH INTEREST

Natural Language Processing, Computer Vision, Large Language Models, Vision Language Models

EDUCATION

University of Dhaka

Bachelor of Science (B.Sc.) in Computer Science and Engineering

CGPA: **3.74** / **4.00** (**7th** out of **65** students)

Dhaka, Bangladesh

January 2017 – August 2021

PROFESSIONAL EXPERIENCE

Therap (BD) Ltd.

Therap Services LLC, Connecticut, USA

Dhaka, Bangladesh

Machine Learning Engineer

October 2022 – Present

- Working on a project to develop a **computer vision**-based product for remote patient monitoring within medical care facilities.
- Led the development of a **video-redaction** system that enables precise redaction of target individuals in a video using **Person Detection**, **Instance Segmentation** models, and **Multi-object Trackers**.
- Designed efficient stream-processing pipelines with optimized computer vision models using **NVIDIA Deepstream SDK**, enabling real-time video analytics on NVIDIA Jetson devices.
- Contributed to a **U.S. patent** for a non-invasive, real-time identification and redaction system in a monitored environment.

Associate Machine Learning Engineer

September 2021 – September 2022

- Explored recent research regarding various deep learning-based Computer Vision tasks such as Object Detection, Face Recognition, Pose Estimation, Segmentation, Activity Recognition, etc. for different scenarios.
- Experimented with cutting-edge vision-based Activity Recognition models on established benchmark datasets and assessed the real-time performance of these models on live camera feeds.
- Worked with a variety of depth sensors to utilize depth data in evaluating the effectiveness of depth-map based Human Activity Recognition models.

RESEARCH EXPERIENCE

Center for Computational & Data Sciences (CCDS)

Research Assistant (Part-Time)

Dhaka, Bangladesh

September 2023 – Present

- Evaluated the performance of Large Language Models (LLMs) across various downstream tasks in low-resource languages like Bangla using techniques including prompting, fine-tuning, etc.
- Pretrained a new Bangla Language Model **Bangla-Llama-2-7B** based on meta's Llama-2 7B model with a large Bangla corpus of around 12 GB using LORA methodology. [\[Code\]](#) [\[HF HUB\]](#)
- Currently working on training **instruction-following LLaMA-based models** for the Bangla language, focusing on enhancing natural language understanding and generation in a low-resource setting.

PUBLICATIONS

- Faria Sultana*, Md Tahmid Hasan Fuad*, Md Fahim, Rahat Rizvi Rahman, **Meheraj Hossain**, M Ashraful Amin, A K M Mahbubur Rahman, Amin Ahsan Ali, “**How Good are LM and LLMs in Bangla Newspaper Article Summarization**”, in Proceedings of the *27th International Conference on Pattern Recognition, ICPR 2024*, To Appear. [\[Paper\]](#)
- Md Fahim, **Meheraj Hossain**, Sadman Rohan, Md Ashraful Amin, AKM Mahabubur Rahman, Amin Ahsan Ali, “**L-Context: Layer-wise Context Vectors for Better Text Classification Using Pre-trained Language Models**”, In Review. [\[Paper\]](#)

PATENTS

- David Lawrence Turock, Justin Mark Brockie, James Michael Kelly, Richard Allen Robbins, **Meheraj Hossain**, et al., “**Automated, Non-Invasive Artificial Intelligence Machine Learning Method and System for Identifying and Redacting Personally Identifiable Information in a Monitored Environment using Real-Time Sensor Data**”, US Patent Publication No. **US 2024-0212804 A1**, published June 27, 2024. (Status: Pending) [\[Patent\]](#)

PROJECTS

Bangla Llama | *LLM, Llama, Huggingface, LoRA, PEFT, SFT* **April 2024 - Present**

- Pretrained **Bangla-Llama-2-7B** based on meta's Llama-2 7B model with around 12 GB Bangla corpus using LORA methodology. [\[Code\]](#) [\[HF HUB\]](#)
- Currently conducting pre-training of **Bangla-Llama-3.1-8B** and **Bangla-Llama-3.2-3B**, based on Meta's Llama-3.1 8B and Llama-3.2 3B models.
- Preparing an instruction dataset for training instruction-following Llama-based models for the Bangla language.

Undergrad Thesis | *Machine Learning, Feature Selection, Bioinformatics, Data Mining* **April 2021**

Title: mMultiSURF- A relief based feature selection method considering class overlapping area among neighbouring instances and prior information. [\[Thesis Book\]](#) [\[Code\]](#)

- Enhanced the state-of-the-art MultiSURF algorithm to improve its robustness and accuracy in selecting relevant feature subsets within very high-dimensional datasets.
- Incorporated an individual instance weighting method within the neighborhood, reflecting the likelihood of non-overlapping regions in calculating feature importance.

Amar Health | *HTML, CSS, Node.js, Express.js, MongoDB* **January 2020**

- Developed a web-based application for electronic health record management and patient monitoring at "Telepsychiatry Research and Innovation Network Ltd" in Dhaka, facilitating psychiatric care and research. [\[Code\]](#)

ML Algo Simulator | *HTML, CSS, Python, Flask, Machine Learning* **April 2019**

- Developed a web application for providing a platform to simulate basic machine learning algorithms on sample datasets. [\[Code\]](#)

MedAdvisor | *Java, Android, SQLite* **December 2018**

- Built an Android app that provides users with disease information, medication details, and includes a medication reminder calendar for tracking doses. [\[Code\]](#)

TECHNICAL SKILLS

Programming Languages: Python, C, C++, Java, JavaScript

Libraries: Pytorch, PyTorch-Lightning, Tensorflow, Keras, OpenCV, Scikit-Learn, Numpy, Pandas, Matplotlib, Seaborn

Web Development: FrontEnd - HTML, CSS, Bootstrap, jQuery, Ajax | BackEnd - Node.js, Express.js

Database: MongoDB, SQL, SQLite

Hardware Tools: Nvidia Jetson Xavier NX, Jetson AGX Orin, Jetson Orin Nano, Raspberry Pi

Miscellaneous: Git, Docker, MATLAB, LaTeX, TensorRT

AWARDS & ACHIEVEMENTS

- Secured 5th Position in Apurba Presents **Bhashabhrom: Bangla Grammatical Error Detection** Challenge Datathon 2023. (**Team:** *Team Aambella*) [\[Link\]](#)
- Selected as Finalist in **Robi Datathon 2.0** (**Team:** *The Anomalies*) [\[Link\]](#)
- **University merit scholarship** awarded by the Government of Bangladesh for outstanding academic performance at the undergraduate level.

EXTRACURRICULAR ACTIVITIES

Competitive Programming

- Solved 1000+ problems in different Online Judges including Codeforces(Max Rating: **1527**), LightOJ, and UVA.
- Participated in several national and international programming contests during my undergraduate studies.

Kaggle Competitions

- Participated in several Kaggle competitions, including [Google Brain Ventilator Pressure Prediction](#) (Time Series Analysis) and [Global Wheat Detection](#) (Computer Vision Challenge) etc.

REFERENCES

Amin Ahsan Ali, Ph.D.

Associate professor, Department of Computer Science and Engineering, Independent University, Bangladesh
Email: aminali@iub.edu.bd, Website: <http://iub.ac.bd/academics/departments/cse/faculty-and-staff/aminali>

Hasnain Heickal

Assistant Professor, Department of Computer Science and Engineering, University of Dhaka
Email: hasnain@cse.du.ac.bd, Website: <https://hheickal.github.io/>