

Sharmin Sultana

Lowell, MA | ✉ Sharmin_Sultana@student.uml.edu | ☎ +1267-301-2190

🐙 [Github](#) | 🎓 [Google Scholar](#) | 🔗 [LinkedIn](#) | 👤 [Portfolio](#)

SUMMARY

- Ph.D. student experience in the development and optimization of natural language processing (NLP) systems, with a focus on leveraging electronic health record (EHR) notes for advanced healthcare applications.
- Experience in working with LLM on large-scale data focusing on improving the efficiency, scalability, and performance for downstream applications.

EDUCATION

UNIVERSITY OF MASSACHUSETTS LOWELL, MA, US | PH.D. IN COMPUTER SCIENCE | SPRING'22 - PRESENT

Advisor: Dr. Hong Yu

Research Focus: Natural Language Processing (NLP), Medical-LLM, Patient Education

GPA: 3.74 / 4.00

CHITTAGONG UNIVERSITY OF ENGINEERING AND TECHNOLOGY, BANGLADESH | B.SC. IN COMPUTER SCIENCE AND ENGINEERING | MARCH 2013 - NOVEMBER 2017

Advisor: Dr. Assaduzzaman

Thesis: Minimum Spanning Tree Based Routing Protocol For Multi-hop And Multi-channel Cognitive Radio

GPA: 3.69 / 4.00, Merit Position: 5th among 120

RESEARCH EXPERIENCE

RESEARCH ASSISTANT, UMASS BIONLP LAB | FALL'23 - PRESENT

- **Extraction of Key Medical Jargon from EHR**
 - Conducted a comprehensive benchmark of both close-source and open-source LLMs to evaluate their effectiveness in identifying the most important medical jargon from EHR notes for patients.
 - Developed a novel data augmentation technique by leveraging MIMIC discharge summaries to address the challenges of training in low-resource settings, resulting in significant performance improvements. Notably, our method enabled smaller open-source LLMs (<10B parameters) to outperform much larger models from the GPT and Claude3 families.
- **NoteAid - ChatBot**
 - Developing a Clinical QA Benchmark dataset that includes medical notes and contextually relevant questions to assess and improve patients' understanding of their electronic health record (EHR) notes.
 - Aiming to create a mobile chatbot to enhance patient education and understanding of their medical notes.

RESEARCH ASSOCIATE, VHA, BEFORD, MA | FALL'24 - PRESENT

- Developing predictive models to identify the risk of Opioid Use Disorder (OUD), as well as the initiation and discontinuation of Medication for Opioid Use Disorder (MOUD), for a cohort of veterans enrolled in the Veterans Justice Outreach (VJO) program.
- Analyzing the factors associated with social determinants of health that contribute to the risk of developing fatal and non-fatal OUD.

RESEARCH INTERN, CUBICS LAB, UMASS LOWELL | SUMMER'22 - SUMMER'23

- Worked on the development of a Multi-Attribute FairLoss function for a CNN model to predict pain status, which effectively addresses bias introduced by sensitive attributes in the dataset and promotes fair classification with 84% accuracy, utilizing wearable device data.

[\[Link\]](#)

PUBLICATIONS

IN PROGRESS

[1] Jang, W. S. *, **Sultana, Sharmin** *, Yao, Z., Tran, H., Yang, Z., & Yu, H., Enhancing the Extraction of Key Medical Jargon from EHR Notes for Patients : Using Data Augmentation for finetuning Large Language Models. Submitted in JAMIA (* These authors contributed equally)

PUBLISHED

[1] **Sultana, Sharmin** & Yidong Zhu, M. A., Shao-Hsien Liu. Wearable-based fair and accurate pain assessment using multi-attribute fairness loss in convolutional neural networks. **Accepted in International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services (MobiQuitous 2024)** (2024).

[2] Bhowmik, S., **Sultana, Sharmin**, Sajid, A. A., Reno, S. & Manjrekar, A. Robust multi-domain descriptive text classification leveraging conventional and hybrid deep learning models. *International Journal of Information Technology* **16**, 3219–3231 (2024).

- [3] Bhowmik, S., Reno, S., **Sultana, Sharmin** & Ahmed, M. Clusterization of different vulnerable countries for immigrants due to covid-19 using mean probabilistic likelihood score and unsupervised mining algorithms. In *2021 International Conference on Information and Communication Technology for Sustainable Development (ICICT4SD)*, 285–290 (IEEE, 2021).
- [4] **Sultana, Sharmin** & Asaduzzaman, A. A minimum spanning tree based routing protocol for multi-hop and multi-channel cognitive radio. In *2018 Joint 7th International Conference on Informatics, Electronics & Vision (ICIEV) and 2018 2nd International Conference on Imaging, Vision & Pattern Recognition (icIVPR)*, 206–211 (IEEE, 2018).

SELECTIVE COURSE PROJECTS

All projects can be found in [GitHub](#)

1. Sentiment Analysis Using Traditional DeepLearning and BERT [GitLink](#) | Pytorch, Tensorflow
 - The project explores the performance of different neural network architectures (LSTM, BiLSTM, & CNN) and compares them with BERT's approach to understand how each model processes and predicts sentiment.
 - BERT outperformed all traditional models with an accuracy of 87%.
2. Comparing ML Algorithms for Cyberbullying Sentiment Analysis on Twitter [GitLink](#) | Python, Sklearn, TfidfVectorizer, TextBlob, NLTK
 - Employed diverse methodologies for detecting and categorizing cyberbullying tweets, involving data pre-processing using Pandas, NumPy as well as feature extraction using TF-IDF techniques to extract tweet text features.
 - Implement, compare, and validate common machine learning (ML) models for tweet classification
3. Twitter Data Retrieval and Hashtag Analysis for Trending Topics [GitLink](#) | Python, Tweepy
 - Data extraction from Twitter and subsequent basic statistical analysis on named attributes (tags) within the retrieved tweets, focusing on hashtags representing specific topics of interest.
 - Implemented a hashtag-based filtering system, ensuring selected hashtags met specific criteria, including clear topic representation, common usage, and an adequate volume of tweets (exceeding 500) for analysis.
4. News Article Extraction and Text Analysis [GitLink](#) | Python, BeautifulSoup, Textacy, Spacy
 - Designed a Python program that automates the process of downloading news articles from the Reuters World News feed using web scraping techniques and the BeautifulSoup library.
 - Used advanced NLP techniques, generated lemmas, nouns, noun phrases, and entity pairs for each article.
5. Infant Cognitive Capacity Prediction through Causal Inference and ML [GitLink](#) | Python, Sklearn, TensorFlow, DoWhy
 - Implemented state-of-the-art machine learning models and causal inference methods to predict the impact of specialized therapeutic treatments on premature infants' cognitive capacity, quantifying the Average Treatment Effect (ATE) and comparing outcomes.

PROFESSIONAL AND ACADEMIC EXPERIENCE

UNIVERSITY OF MASSACHUSETTS LOWELL | GRADUATE TEACHING ASSISTANT

- Introduction to Computing I - Spring'22
- Human-Computer Interaction - Fall'23
- Data Science - Fall'24

BANGLADESH ARMY INTERNATIONAL UNIVERSITY OF SCIENCE & TECHNOLOGY | TEACHING LECTURER

- Object Oriented Programming
- Machine Learning

TECHNICAL SKILLS

NLP & ML Tools: TensorFlow, Keras, PyTorch, Scikit Learn, Hugging Face, TextBlob, SpaCy, Gensim

Data Analytics & others: Pandas, Numpy, SQL, Matplotlib, Seaborn

Programming Language: Python, C/C++, HTML, CSS