Jillur Rahman Saurav Email : sauravsust71@gmail.com
https://facevoid.github.io/

Education

- **University of Texas at Arlington** Arlington, Texas
  PhD in Computer Science *2021 – present*
- **Shahjalal University of Science & Technology** Sylhet, Bangladesh
  B.Sc. (Engg.) in Computer Science *2013 – 2018*

Publications

Preprints

- **Saurav, J.R.**, Nasr, M.S., Koomey, P., Robben, M., Huber, M., Weidanz, J., Ryan, B., Ruppin, E., Jiang, P. and Luber, J.M., 2022. **A SSIM Guided cGAN Architecture For Clinically Driven Generative Image Synthesis of Multiplexed Spatial Proteomics Channels**. https://arxiv.org/abs/2205.10373
- Nasr, M.S., Hajighasemi, A., Koomey, P., Malidarreh, P.B., Robben, M., **Saurav, J.R.**, Shang, H.H., Huber, M. and Luber, J.M., 2023. **Clinically Relevant Latent Space Embedding of Cancer Histopathology Slides through Variational Autoencoder Based Image Compression** https://arxiv.org/abs/2303.13332

Conference Papers

- **Saurav, J.R.**, Xiang, K., Deb, N., Amin, M.R. (2021). **A Comparative Study of Language Dependent Gender Bias in the Online Newspapers of Conservative, Semi-Conservative and Western Countries**. 23rd International Conference on Human-Computer Interaction (HCII 2021)
- Tasnim, N., Shihab, M.I.H., Rahman, M., **Saurav, J.R.**, Islam, S.R., Amin, M.R. (2021). **Observing the Unobserved: A Newspaper Based Dengue Surveillance System for the Low-Income Regions of Bangladesh.** The 34th International FLAIRS Conference (FLAIR-34)
- Sarker, S., Islam, M. E., **Saurav, J.R.**, & Nahid, M. M. H. (2020, November). **Word Completion and Sequence Prediction in Bangla Language Using Trie and a Hybrid Approach of Sequential LSTM and N-gram**. 2020 2nd International Conference on Advanced Information and Communication Technology (ICAICT). https://doi.org/10.1109/icaict51780.2020.9333518
- Islam, M. R., **Saurav, J.R.**, Talha, M. R., & Chowdhury, F. (2020). **Query Expansion for Bangla Search Engine Pipilika**. 2020 IEEE Region 10 Symposium (TENSYMP). https://doi.org/10.1109/tensymp50017.2020.9231043
- **Saurav, J.R.**, Haque, S., & Chowdhury, F. (2019, September). **End to End Parts of Speech Tagging and Named Entity Recognition in Bangla Language**. 2019 International Conference on Bangla Speech and Language Processing (ICBSLP). https://doi.org/10.1109/icbslp47725.2019.201541
- **Saurav, J.R.**, Amin, S., Kibria, S., & Shahidur Rahman, M. (2018, September). **Bangla Speech Recognition for Voice Search**. 2018 International Conference on Bangla Speech and Language Processing (ICBSLP). https://doi.org/10.1109/icbslp.2018.8554944

Experience

- **Pipilika** Bangladesh
  Software Engineer *Nov 2017 - Dec 2020*
  - **News aggregator Service based on Bangla Newspapers**: Performed tasks that include designing architecture, developing generic parser, clustering news, categorizing news, summary extraction.
    Technology: Django, Scrapy, Elasticsearch, Keras, Redis, Docker
  - **Knowledge graph based on Bangladesh's national portals data**: Built a knowledge graph using the data from Bangladesh's national portals by analyzing the text of the entities on the websites (5552 web portals) using K-means clustering and the Nearest Neighbour method.
    Technology: Python, Elasticsearch, sklearn, Docker
  - **Context-aware spell checker for Bangla language**: Worked as a team member for developing Bk-tree, n-gram based spell checker for Bangla language.
    Technology: Spring boot, Apache Solr

- **Query Analysis**: Developed a deep-learning-based query classifier to understand search queries, implemented autocomplete and related search features.
  Technology: Keras, Elasticsearch
- **Sentiment Analysis Dataset for Bangla language**: Worked as a team member for developing the largest sentiment analysis dataset for Bangla language. The performed tasks included scraping data from various sources, cleaning data, and selecting data for annotation.
  Technology: Scrapy, Selenium, Keras, Pandas
- **Stemmer for Bangla Language**: A suffix-stripping-based stemmer for Bangla language that can perform light-weight stemming and heavy-weight stemming.
  Technology: Java
- **Ngram Generation and Search API**: Built Largest Bangla n-gram corpus from Newspaper data. Used thread pool executors to minimize corpus development time.
  Technology: Java, Thread pool, Mongo
- **Location Parser**: Developed a module for parsing both clean and inflected administrative location names [i.e. District, Upazila, Union] from raw text.
  Technology: Java
- **Perceptive Scheduler**: Developed a regression-based module to determine the schedule of crawlers. This module analyzes the newspaper's article publish times and determines how frequently that should be crawled.
  Technology: Python

- **Projects**
  Side Projects
  - **Data Analytics for COVID-19 self-screening tool**: Performed various statistical analyses on a Covid-19 self-screening tool's data (535,291 participants) comprising association analysis among symptoms, symptoms clustering, identifying danger zones, correlation with Covid cases.
    Technology: Pandas, Sklearn
  - **Computer Vision Projects**: Worked on several computer vision projects. Tasks included real-time object detection, reverse image search, Image captioning in the Bangla language.
    Technology: Keras

Certification

- **Deep Learning Specialization:**
  https://www.coursera.org/account/accomplishments/specialization/QNV3G4LH6W9Q

Skills

- **Languages**: Python, Java **Technologies**: Pytorch, Keras, Elasticsearch, Django, Docker, Scrapy