Kowshika Sarker

Computer Science Ph.D. student, University of Illinois at Urbana-Champaign

ksarker2@illinois.edu | kowshikasarker.github.io | kowshika-sarker | kowshikasarker

Research interest

Computational biology, Machine learning, Big data analytics

Personal statement

I design data-driven solutions for real-world problems, especially for the biomedical domain. I develop machine learning and algorithmic methods to analyze omics- and other high-dimensional data. My current focus is utilizing prior domain knowledge via knowledge networks for improved empirical performance and interpretable insights.

Education

- **University of Illinois at Urbana-Champaign** Aug 2021 – Present
  PhD in Computer Science. Advisor: Prof. ChengXiang Zhai
- **University of Illinois at Urbana-Champaign** Aug 2021 – Dec 2024
  MS in Computer Science. Advisor: Prof. ChengXiang Zhai
  Thesis: Imputing metabolomics with graph denoising autoencoders
- **Bangladesh University of Engineering & Technology** Jul 2014 – Oct 2018
  BSc in Computer Science & Engineering. CGPA: 3.84/4.00. Rank: 18/126
  Thesis supervisor: Prof. Md. Shamsuzzoha Bayzid
  Thesis: (1) STELAR: a statistically consistent coalescent-based species tree estimation method by maximizing triplet consistency. (2) eMED-DNA: An in silico operating system for clinical medical data storage within the human genome.

Publication

- **Kowshika Sarker**, ChengXiang Zhai. Graph-based prior-guided synthetic metabolomic data generation. Manuscript in preparation.
- **Kowshika Sarker**, Ruoqing Zhu, Hannah D Holscher, ChengXiang Zhai. Integrating prior knowledge from genome-scale metabolic model with metabolomics for diet assessment. Submitted to *IEEE Transactions on Computational Biology and Bioinformatics (TCBB)*.
- **Kowshika Sarker**, Ruoqing Zhu, Hannah D Holscher, ChengXiang Zhai. Prior-guided longitudinal metabolomic analysis. *IEEE International Conference on Bioinformatics and Biomedicine (BIBM), 2024*. [DOI]
- **Kowshika Sarker**, Ruoqing Zhu, Hannah D Holscher, ChengXiang Zhai. Augmenting nutritional metabolomics with a genome-scale metabolic model for assessment of diet intake. *ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics (ACM-BCB), 2023*. Accepted among top 10% of manuscripts. [DOI]
- Yasamin Tabatabaee, **Kowshika Sarker**, Tandy Warnow. Quintet Rooting: rooting species trees under the multi-species coalescent model. *Bioinformatics. 2022;38 (Suppl 1):i109-i117*. [DOI]
- Mazharul Islam, **Kowshika Sarker**, Trisha Das, Rezwana Reaz, Md. Shamsuzzoha Bayzid. STELAR: A statistically consistent coalescent-based species tree estimation method by maximizing triplet consistency. *BMC Genomics. 2020; 21(1):1–13*. [DOI]

Project

- **Multi-omic analysis of rare disease.** Analyzing multi-omic data - genomic variants, transcriptomic reads, and disease phenotypes - to prioritize pathogenic genes based on aberrant signals across multiple omic modalities in a cohort of rare disease patients.
- **Prior-knowledge guided metabolomics imputation.** Representing metabolomic samples with graphs based on different correlation measures and prior knowledge of metabolite pairs participating in common reaction(s), imputing missing metabolomes with graph denoising autoencoders.

- **Transcriptional binding prediction.** Predicting transcription factor binding affinity to genome segments of fixed lengths with one-dimensional convolutional neural networks.
- **Medical record archival.** In-silico demonstration of co-storing genome with nucleotide-encoded electronic health records (EHRs) to facilitate data archival and transfer, by proposing novel encoding techniques to transform different EHR components such as medical images, and clinical notes into nucleotide sequences. [Preprint]
- **Illegal fishing monitoring.** Detecting illegal fishing of prohibited species by localization of fishes in CCTV images and multi-class classification on the localized image segments.
- **Disease based diet recommender.** Scraping webpages for texual diet recommendations of different disease, food items are classified into 3 categories - recommended, detrimental, and neutral - using sentiment analysis.

Experience

- **Lecturer**, Dept. of Computer Science & Engineering, East West University, Bangladesh — Jan 2019 – Aug 2021
- **Software engineer**, Samsung R&D Institute, Bangladesh — Nov 2018 – Dec 2018

Recognition

- Mayo Clinic and Illinois Alliance fellowship for technology-based healthcare research — Aug 2021 – Aug 2023
- Best student poster, International Conference on Networking, Systems and Security — 2017
  Title: Archiving Medical Records in DNA Sequences
- Merit scholarship, BUET.
- Dean's list, BUET.

Activity

- Judge, Regional science fair, Region-4, Illinois Junior Academy of Science (IJAS) — 2024
- Organizing member, International Conference on Advances in Science, Engineering & Robotics Technology — 2019
- Reviewer, Digital textbooks by Bangladesh Technical Education Board — 2017

Teaching

- **Graduate teaching assistant, University of Illinois at Urbana-Champaign**
  CS412: Introduction to Data Mining, CS410: Text Information Systems
- **Instructor, East West University**
  Computer Graphics, Operating Systems, Database Systems, Numerical Methods, Software Engineering & Information System Design, Discrete Mathematics, Computer-aided Engineering Drawing

Skill

- **Language:** Python, Java, C, C++, R, Javascript, Assembly, LaTeX, Matlab, HTML, CSS, Shell script
- **Frameworks:** Django, PyTorch, Keras, Tensorflow, OpenGL
- **Database:** Oracle, SQLite
- **Hardware:** ATmega32, Arduino, RomeoV2
- **Simulator:** NS2, Cisco packet tracer, Nachos, Proteus