

# Kowshika Sarker

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

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

## Research interest

---

Computational biology, Machine learning, Big data analytics

## Personal statement

---

I focus on designing data-driven solutions for real-world problems, especially for the biomedical domain. I develop machine learning and algorithmic approaches to analyze omics- and other high-dimensional modalities.

## Education

---

- **University of Illinois at Urbana-Champaign** Aug 2021 – Present  
Ph.D. in Computer Science. Advisor: ChengXiang Zhai
- **Bangladesh University of Engineering & Technology** Jul 2014 – Oct 2018  
B.Sc. in Computer Science & Engineering. Thesis supervisor: Md. Shamsuzzoha Bayzid  
CGPA: 3.84/4.00. Rank: 18/126

## Publication

---

- **Kowshika Sarker**, ChengXiang Zhai. Graph-based prior-guided synthetic metabolomic data generation. Submitted to *RECOMB 2025*.
- **Kowshika Sarker**, Ruoqing Zhu, Hannah D Holscher, ChengXiang Zhai. Prior-guided longitudinal metabolomic analysis. Accepted at *BIBM 2024*.
- **Kowshika Sarker**, Ruoqing Zhu, Hannah D Holscher, ChengXiang Zhai. Augmenting nutritional metabolomics with a genome-scale metabolic model for assessment of diet intake. Accepted among top 10% of manuscripts at *ACM-BCB 2023*. [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 genomic variants, transcriptomic reads, and phenotypes to prioritize pathogenic genes in a cohort of rare disease patients.
- **Prior-knowledge guided metabolomics imputation**. Representing metabolites with graphs based on correlations and prior knowledge and imputing missing metabolomes with denoising autoencoders.
- **Transcriptional binding prediction**. Predicting transcription factor binding affinity to genome segments with convolutional neural networks.
- **Medical record archival**. In-silico demonstration of co-storing genome sequence with nucleotide-encoded electronic health records to facilitate data archival and transfer. [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 textual diet recommendations of different disease, food items are classified into 3 categories - recommended, detrimental, and neutral - using sentiment analysis.

## Experience

---

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

## Recognition

---

- **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**  
Introduction to Data Mining, 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, ~~TeX~~  $\text{\LaTeX}$ , 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