Kanchan Chowdhury

Email : kchowdh1@asu.edu
Webpage: kanchanchy.github.io
Mobile : +1-480-410-8677
Linkedin: linkedin.com/in/kanchan-chowdhury-5729699a

Research Interests

Machine Learning, Database Systems, and Geospatial Data Analytics

Education

- **Arizona State University** — Tempe, Arizona
  PhD & MS in Computer Science — *Aug. 2018 – June 2024*
  Advisor: Prof. Mohamed Sarwat & Prof. Jia Zou, CGPA: 4.00

- **Chittagong University of Engineering and Technology** — Chittagong, Bangladesh
  Bachelor of Science in Computer Science and Engineering — *Mar. 2010 – Nov. 2014*
  Advisor: Prof. Mohammed Moshiul Hoque, CGPA: 3.76

Experience

- **Arizona State University** — Tempe, Arizona
  Research Assistant — *Aug. 2018 - Present*
  - **Research Projects**:
    1) Co-optimization of machine learning and join queries based on model decomposition and join push-down.
    2) Designing and implementing a deep-learning and data processing framework for raster imagery and vector datasets.
    3) A benchmark to empirically evaluate raster image labeling approaches while proposing an unsupervised approach.
    4) Evaluating state-of-the-art approaches for the synthesis of SQL queries from natural language questions.
    5) Evaluating machine learning models in predicting the next SQL query in a user-query session.

- **Wherobots Inc.** — Scottsdale, Arizona, USA
  Research and Development Intern — *Jan. 2023 - Aug. 2023*
  - **Responsibilities**: Designing and developing spatial machine learning and deep learning tools, scalable map-matching, and geospatial data analytical algorithms. Integrating the developed tools into the Wherobots cloud platform.

- **Gagagugu PTE LTD** — Dhaka, Bangladesh
  Software Engineer — *Jan. 2017 - Jun. 2018*
  - **Responsibilities**: Developing Android Apps with social networking features such as calling, messaging, and posting.

- **Le Chef Plc** — Dhaka, Bangladesh
  Android Application Developer — *Jan. 2015 - Dec. 2016*
  - **Responsibilities**: Developing Android Apps featuring online order and reservation services for restaurants in the UK.

Publications

- **Kanchan Chowdhury**, Mohamed Sarwat; Deep Learning with Spatiotemporal Data: A Deep Dive into GeotorchAI. *Accepted in 40th International Conference on Data Engineering (ICDE)*, 2024

- Lixi Zhou, Qi Lin, **Kanchan Chowdhury**, Saif Masood, Alexandre Eichenberger, Hong Min, Alexander Sim, Jie Wang, Yida Wang, Kesheng Wu, Binhang Yuan, Jia Zou; Serving Deep Learning Models from Relational Databases. *Accepted in 27th International Conference on Extending Database Technology (EDBT)*, 2024

- **Kanchan Chowdhury**, Mohamed Sarwat; A Demonstration of GeoTorchAI: A Spatiotemporal Deep Learning Framework. *ACM SIGMOD International Conference on Management of Data*, 2023

- **Kanchan Chowdhury**, Vamsi Meduri, Mohamed Sarwat; A Machine Learning-Aware Data Re-partitioning Framework for Spatial Datasets. *38th International Conference on Data Engineering (ICDE)*, 2022

- **Kanchan Chowdhury**, Mohamed Sarwat; GeoTorch: A Spatiotemporal Deep Learning Framework. *30th International Conference on Advances in Geographic Information Systems (SIGSPATIAL '22)*, 2022

- Vamsi Meduri, **Kanchan Chowdhury**, Mohamed Sarwat; Evaluation of Machine Learning Algorithms in Predicting the Next SQL Query From the Future. *ACM Transactions on Database Systems (TODS)*, 2021
- Jia Yu, **Kanchan Chowdhury**, Mohamed Sarwat; Tabula in Action: A Sampling Middleware for Interactive Geospatial Visualization dashboards. *46th International Conference on Very Large Databases (VLDB)*, 2020.
- Vamsi Meduri, **Kanchan Chowdhury**, Mohamed Sarwat; Recurrent Neural Networks for Dynamic User Intent Prediction in Human-Database Interaction. *22nd International Conference on EDBT, 2019*
- **Kanchan Chowdhury**, Lamia Alam, Shyla Sarmin, Safayet Arefin, Mohammed Moshiul Hoque; A Fuzzy Features Based Online Handwritten Bangla Word Recognition Framework. *18th ICCIT*, 2015

Teaching

- **Instructor** — Fall 2022
  Spatial Data Science and Engineering (CSE 594) — *Arizona State University*
  - **Responsibilities**:
    1) Preparing lecture slides and conducting lectures.
    2) Preparing assignments, projects, exam questions, and grading rubrics.
    3) Office hours to help students understand lectures and projects.
- **Teaching Assistant** — Fall 2021, Spring 2021, Fall 2020, and Spring 2020
  Distributed Database Systems (CSE 512) — *Arizona State University*
- **Teaching Assistant** — Fall 2019
  Data Processing at Scale (CSE 511) — *Arizona State University*
- **Teaching Assistant** — Spring 2019
  Object-Oriented Programming & Data Structure (CSE 205) — *Arizona State University*
- **Teaching Assistant** — Spring 2019
  Principles of Programming with C++ (CSE 100) — *Arizona State University*
- **Teaching Assistant** — Fall 2018
  Principles of Programming with Java & Python (CSE 100) — *Arizona State University*

Mentoring & Training

- Mentored a Ph.D. student and an MS student in the Data Systems Lab at Arizona State University.
- Trained more than 75 undergraduate students in Bangladesh on Android Application Development from Feb. 2015 to May 2015 under the ICT Division of Bangladesh.

Additional Services

- **Paper Reviewer**: Served as a reviewer for the journal IEEE TKDE and the conference ICCAD 2023.
- **External Reviewer**: Reviewed papers as an external reviewer for the following conferences and journals - SIGMOD 2020 & 2022, VLDB 2019-2022, ICDE 2020, SIGSPATIAL 2021, VLDB Journal, and TSAS Journal
- **Grant Reviewer**: Served as a travel grant reviewer for the Graduate and Professional Student Association (GPSA) at Arizona State University from May 2022 to August 2023.
- **Conference Volunteer**: Volunteered to organize two conferences - SIGSPATIAL 2022 and SIGMOD 2023.
- **Presentations & Talks**: Five conference presentations - SIGMOD 2023, FOSS4GNA 2023, ICDE 2022, SIGSPATIAL 2022, and ICCIT 2015.
- **Open Source Contribution**: Contributed to Apache Sedona, an open-source geospatial cluster computing framework with 1.6k+ GitHub Stars, by adding support for two new spatial data types.

Participation and Awards

- Recipient of ACM SIGMOD 2023 student travel award to attend the conference and present a paper.
- Recipient of ACM SIGSPATIAL 2022 travel award to attend the conference and present a paper.
- Received ASU Graduate and Professional Student Association (GPSA) travel grant twice - 2022 and 2023
- Recipient of CIDSE Doctoral Fellowship at Arizona State University for the academic year 2018-2019.
- 2nd Runner-up at National Hackathon organized by ICT Division of Bangladesh in 2014. The challenge of the hackathon was to design a project-based solution to solve a national problem of the country.
- 2nd Runner-up at National Mobile Application Code Hub organized by BUET, Bangladesh in 2014.
- Recipient of Honors award from my undergraduate university for maintaining academic excellence.
- Recepient of merit scholarship in all four years of my undergraduate education.
- Recepient of the government merit scholarship in Secondary School Certificate examination, Higher Secondary School Certificate examination, 8th-grade public examination, and 5th-grade public examination.

Technical Skills

- **Programming**: Python, Java, C, C++, Scala, SQL, and HTML
- **Databases**: PostgreSQL, SparkSQL, and MySQL
- **Machine Learning**: PyTorch, Scikit-learn, Keras, TensorFlow, ML & DL Models, and ML Statistics
- **Data Analytics**: Apache Spark, Apache Sedona, PySpark, GeoPandas, Pandas, Matplotlib, and Plotly
- **OS**: Unix/Linux and Windows
- **Others**: Distributed Computing, SDE Design Principles, CI/CD Pipeline, Docker, Jira, and Rest API

Projects

- **GeoTorchAI**: A deep learning and data preprocessing framework for raster imagery and spatiotemporal vector datasets, with **400+ GitHub Stars**. It enables spatiotemporal machine learning practitioners to easily and efficiently implement spatiotemporal deep learning models, besides supporting scalable data preprocessing.
- **Raster Image Labeling**: A benchmark for experimentally evaluating various satellite image labeling approaches. This framework also proposes an unsupervised approach to handle scenarios when manual data labeling is labor intensive and training a supervised model is difficult due to a lack of sufficient labeled data.
- **Map Matching**: A distributed and scalable map matching framework based on the Hidden Markov model and distance computation to map noisy GPS coordinates to road networks.
- **NLIDB-Bench**: A benchmark for evaluating state-of-the-art approaches of SQL query generation from natural language queries. It proposes a set of evaluation metrics and conducts experiments with four datasets.
- **Named Entity Recognition**: This work tunes various steps of state-of-the-art methods for named entity recognition and performs evaluation with two popular datasets: CoNLL-2003 and OntoNotes-5.0.
- **Hotspot Analysis on Apache Sedona**: This work performs spatial queries and range joins between two spatial datasets and calculates the Getis-Ord statistic of NYC Taxi Trip dataset to perform hot-cell analysis.
- **Climate Change Forecasting**: A data science project to perform data cleaning, feature engineering, and data preprocessing operations on raw temperature data and predict temperature trends with the LSTM model.
- **Fake News Detection**: A data science project to perform data cleaning, feature engineering, and data preprocessing operations on news datasets and classifying fake and real news with a Bidirectional LSTM model.

- **Data Visualization in Python**: It performs statistical data analysis using useful visualization patterns with three Python libraries: Matplotlib, Seaborn, and Plotly Express.
- **Sentiment Analysis with BERT**: A PyTorch project to classify emotions using the Twitter smile dataset. It contains ML Engineering steps, such as data preprocessing, tokenization, encoding, training, and evaluation.

Important CourseWorks

- **AI**: Statistical Machine Learning, Fundamentals of Statistical Learning, Data mining, Data Intensive Systems for Machine Learning, Deep Neural Networks with PyTorch, & Artificial Intelligence
- **DBMS**: Distributed Database Systems, Big Data Analysis with Scala and Spark, & Database Management Systems
- **Core**: Discrete Mathematics, Software Engineering, Operating Systems, Data Structure, & Algorithms