Effat Farhana

effat.farhana@vanderbilt.edu
https://effat.github.io/

Research Interests

My research expertise lies in developing Data Mining, Artificial Intelligence (AI), and Machine Learning (ML) methods addressing unique challenges arising in specific application domains, such as education and healthcare domains. Publications in EAAI@AAAI, GECCO, L@S, and EDM.

Education

- **Doctor of Philosophy (Ph.D.) in Computer Science** Aug 2015 - April 2021
  - Dissertation title: "Science Reading Behavior of Middle School Students within a Digital Literacy Platform"
  - Adviser: Dr. Collin F. Lynch
  - Committee: Dr. James Lester, Dr. Noboru Matsuda, and Dr. Teomara Rutherford
  - North Carolina State University, Raleigh, NC, USA
- **Bachelor of Science (B.Sc.) in Computer Science and Engineering** Feb 2011
  - Bangladesh University of Engineering and Technology, Dhaka, Bangladesh

Awards & Honors

- **Rising Star in Data Science- Jan 2021, University of Chicago.**
  For my Ph.D. work. organised by Center for Data and Computing (CDAC), University of Chicago. Link
- **Scholarship by ACM Richard Tapia Celebration of Diversity in Computing, 2020**
  To present poster based on my Ph.D. work in ACM Student Research Competition.
- **Travel award**
  To attend Women in Machine Learning (WiML) workshop, co-located with NeurIPS 2020 (virtual).
- **Travel award**
  For CRA-WP Widening Participation Early and Mid Career Mentoring Workshop, 2020
- **Scholarship by Women in Computer Science (WiCS), NC State**
  To attend Grace Hopper Conference, 2018.
- **Dean's List Award**
  Received this award for achieving academic excellence for six semesters during the Bachelors program in Bangladesh University of Engineering and Technology.
- Attended Computer and Information Science and Engineering (CISE) CAREER Proposal Writing workshop Spring 2022

Selected Research Experience

North Carolina State University

- Topic: **Automated Question Generation for Science Reading** (NLP, ML)
  - Automatically generating questions for K-12 science text and validating with domain experts.
- Topic: **Sequential User Behavior Modelling in an Online Reading Platform** (ML, Data Mining)

- Designed a transformer-based model to predict 12,000 students' question scores incorporating contextual information of questions and learning activities. Proposed model performed better than state-of-the-art approaches.
- Interpretable visualization of the transformer's attention can help teachers to identify students' study habits and performance (EAAI-AAAI 2022).
- Identified reading and meta-cognitive behavior patterns for 12.5K science and 16.2K social study student data. (L@S 2020)

- Topic: **Responsible AI: Interpretable Classification Algorithm Design** (Interpretable ML)
  - Developed a rule-based classification algorithm, BBO-RM, using an evolutionary algorithm, Biogeography- based optimization (BBO). BBO-RM performed significantly better compared to baseline classification algorithms on 14 UCI repository datasets (GECCO 2017).
  - Implemented a parallel version of the BBO-RM algorithm utilizing Julia programming language's data slicing techniques and Julia's parallelization features. (GECCO 2018).

Vanderbilt University

Working in research projects involving people with autism spectrum disorder (ASD). Cross-disciplinary research between departments of computer science, psychology, and medical school at Vanderbilt.

- Topic: **Block Design Test and ASD** (ML, Data Mining)
  - Understanding user behavior patterns in block design test and link behavioral patterns to users' visual thinking abilities (i.e., how well they can reason about visual information). In the block design test, a user has to rearrange a pile of colored blocks to a target design.
- Topic: **Educational Game to Teach Social Skills to ASD Kids** (ML, Data Mining, Psychology)
  - Designing user surveys and applying data mining and machine learning techniques to connect kids' in-game performance with social skills.

Publications

Drafts In-preparation

1. User Behavior Modelling in the Block Design Test and Autism
   Effat Farhana et al.

2. Automated Question Generation for Science Education
   Effat Farhana et al.

Peer-Reviewed Journals and Conference Publications

1. [**EAAI@AAAI 22**] Predictive Student Modelling in an Online Reading Platform
   Effat Farhana, Teomara Rutherford, and Collin F. Lynch.
   Twelfth AAAI Symposium on Educational Advances in Artificial Intelligence (EAAI-AAAI 2022) .

2. [**ACS 22**]Framework for a multi-dimensional test of theory of mind for humans and AI systems
   Caoimhe Harrington Stack, Sarah Myers, **Effat Farhana**, Aviv Roskes, Xinyu Shen, Simeng Zhao, Angela Maliakal, Roxanne Rashedi, Joel Michelson and Maithilee Kunda
   The Tenth Annual Conference on Advances in Cognitive Systems (ACS 2022).

3. [**L@S 20**] Understanding Reading Behaviors of Middle School Students
   Effat Farhana, Teomara Rutherford, and Collin F. Lynch.
   Proceedings of the Seventh ACM Conference on Learning @ Scale (L@S 2020).

4. [**ICLS 20**] Associations Between Self-Regulated Learning Strategies and Science Assignment Score in a Digital Literacy Platform
   Effat Farhana, Teomara Rutherford, and Collin F. Lynch.
   Proceedings of the International Conference of the Learning Sciences (ICLS 2020).

5. [**EDM 20**] Investigating Relations between Self-Regulated Reading Behaviors and Science Question Difficulty
   Effat Farhana, Teomara Rutherford, and Collin F. Lynch.
   Proceedings of the 13th International Conference on Educational Data Mining (EDM 2020).

6. [**ICSE 20**] Gang of eight: A Defect Taxonomy for Infrastructure as Code Scripts
   Akond Rahman, **Effat Farhana**, Chris Parnin, and Laurie Williams.
   Proceedings of the 42nd International Conference on Software Engineering, (ICSE 2020)

7. [**EMSE 20**] The 'as code' Activities: Development Anti-patterns for Infrastructure as Code
   Akond Rahman, **Effat Farhana** and Laurie Williams.
   Empirical Software Engineering. 25, 3430–3467 (EMSE 2020).

8. [**ICSME 19**] Synthesizing Program Execution Time Discrepancies in Julia Used for Scientific Software
   Effat Farhana, Nasif Imtiaz and Akond Rahman,
   IEEE International Conference on Software Maintenance and Evolution (ICSME 2019)

9. [**MSR 19**] Challenges with Responding to Static Analysis Tool Alerts
   Nasif Imtiaz, Akond Rahman, **Effat Farhana** and L. Williams.
   IEEE/ACM 16th International Conference on Mining Software Repositories (MSR 2019)

10. [**GECCO 17**] Biogeography-based Rule Mining for Classification
    Effat Farhana and Steffen Heber.
    Proceedings of the Genetic and Evolutionary Computation Conference (GECCO 2017).

11. [**Inf. Sci. 15**] Constrained sequence analysis algorithms in computational biology
    Effat Farhana, and M. Sohel Rahman.
    Information Sciences 295 (2015).

12. [**Inf. Process. Lett. 12**] Doubly-constrained LCS and hybrid-constrained LCS problems revisited
    Effat Farhana, and M. Sohel Rahman.
    Information Processing Letters 112.13 (2012).

13. [**SPIRE 15**] Finite Automata Based Algorithms for the Generalized Constrained Longest Common Subsequence Problems
    Effat Farhana, Jannatul Ferdous, Tanaeem M. Moosa, M. Sohel Rahman.
    17th International Symposium of String Processing and Information Retrieval, (SPIRE 2010).

Peer-Reviewed Workshop, Poster Publications, and Others

1. [**EDM 18**] Predicting Post-College STEM Enrollment from Middle School Clickstream Data
   Effat Farhana, Maaz Saleem Kapadia, Wenjia Cao, and Collin F. Lynch.
   Workshop on Scientific Findings from the ASSISTments Longitudinal Data Competition: (EDM 2018).

2. [**GECCO 18**] A Parallel Island Model for Biogeography-based Classification Rule Mining in Julia
   Samuel Ebert, **Effat Farhana**, and Steffen Heber.
   Proceedings of the Genetic and Evolutionary Computation Conference Companion (GECCO 2018).

3. [**EDM 21**] Feedback and Self-regulated Learning in Science Reading
   Effat Farhana, Andrew Potter, Teomara Rutherford, and Collin F. Lynch.
   Proceedings of the 14th International Conference on Educational Data Mining (EDM 2021) (Poster).

4. [**EDM 20**] Self-Regulated Learning and Science Reading of Middle-School Students
   Effat Farhana, Teomara Rutherford, and Collin F. Lynch.
   Doctoral Consortium. The 13th International Conference on Educational Data Mining (EDM 2020).

Grant Proposal Writing

MD-ToM: Multi-domain assessment of theory of mind skills for adolescents on the autism spectrum.
PI: Dr. Maithilee Kunda.

- **Role:** Designed and wrote the Data Analysis part under the Research Plan Section.
  Result: Submitted to Institute of Education Science (IES).

Mentoring

North Carolina State University.

- Samuel Ebert, CS undergraduate — Aug 2017- Aug 2018
  Interpretable ML algorithm project at NC State. This work resulted in a student's lead author paper (GECCO 2018).

Vanderbilt University

- Jeannie Jeong, undergraduate in Cognitive Science — Fall 2021 - Summer 2022
  Project on visual spatial skills in the workplace.
  Current position: Stanford Department of Medicine, Healthcare AI Applied Research Team.
- James Foglio, undergraduate in CS — Summer 2022
  Project on Block Design and ASD performance.
- Raymond Yates, MS intern at the First Center of Autism — Summer 2022
  Visualization on ASD project.

Employment History

- **Postdoctoral Research Scholar** — Jul 2021 - Present
  Department of Computer Science, Vanderbilt University — Nashville, TN
- **Graduate Teaching Assistant / Research Assistant** — Aug 2015 -May 2021
  Department of Computer Science, North Carolina State University — Raleigh, NC
- **Lecturer, Department of Computer Science** — Apr 2011 - Dec 2014
  Ahsanullah University of Science and Technology — Dhaka, Bangladesh

Teaching Experience

- **North Carolina State University** — Fall 2016 - Spring 2020

  – Teaching Assistant (TA)
  -TA for graduate level courses: Design and Analysis of Algorithm ($\sim$ 200 students), Database Management Concepts and Systems ($\sim$ 150 students), Software Engineering ($\sim$ 30 students), Artificial Intelligence ($\sim$ 60 students), and undergraduate level Data Structure and Algorithm ($\sim$ 150 students) course.
  - Created assignments and exam questions, held office hours, and graded.

- **Ahsanullah University of Science and Technology** — April 2011 - Dec 2014

  – Lecturer in Computer Science and Engineering
  - Instructor for undergraduate level introductory programming language, design and analysis of algorithm, and compiler courses.

Services

Reviewer

2022	AAAI Symposium on Educational Advances in AI (EAAI) at AAAI (AI in Education Track)
2022	AAAI/ACM Conference on Artificial Intelligence, Ethics, and Society (AIES)
2022	International Conference on Artificial Intelligence in Education (AIED)
2022	ACM Transactions on Computing Education
2022	Advances in Cognitive Systems (ACS)

Sub-reviewer/Shadow PC

2021 International Conference on Educational Data Mining
2021 Mining Software Repository (MSR)

Volunteering

2021 Judge for VandyHacks (Vanderbilt's premier student hackathon)
2020 Ph.D. panel member at Doctoral Recruiting Day, NCSU
2017 NC State International Graduate Student Orientation

Invited Talks

Machine Learning and Data Science for Social Good (Spring 2022) Florida International University

---

Skills

Languages Java, Python, R, C, C++, Julia, Bash

Databases MySQL, Oracle, SQLAlchemy

Frameworks & Libraries Scikit-Learn, NLTK, Keras, TensorFlow, PyTorch

---

Professional Memberships

- Affiliate at The First Center for Autism and Innovation, Vanderbilt University
- ACM Professional Member Membership No.: **4570793**
- AAAI Member