Sumon Biswas

PHD Candidate · Research Assistant · Computer Science · Iowa State University

226 Atanasoff Hall, 2434 Osborn Drive, Ames, IA 50010

(+1) 515-708-6166 | sumon@iastate.edu | https://sumonbis.github.io | sumonbis | sumonb

Summary

I finished my bachelor in 2015 and currently a Ph.D. candidate in computer science at Iowa State University. I am also working as a research assistant in Software Design Lab under the supervision of Professor Hridesh Rajan. My research goal is to innovate and improve state-of-the-arts which have real-world impact. Broadly, I am interested in the intersection of software engineering, programming language, and artificial intelligence. Currently, I am looking for academic position to pursue a research-oriented career and excel in my research area.

Research Interest

Software Engineering	SE for Artificial Intelligence (AI), Empirical Software Engineering, Big Code Mining, Reliability
Data Science	Fairness of Machine Learning (ML) Software, Compositional Fairness, Fairness Verification in DNN
Programming Languages	Program Analysis, Compilers, Verification, Program Evolution and Maintenance

Education

Iowa State University (ISU)   *Ames, IA 50010*

Ph.D. in Computer Science   *August 2016 - Present*

- **Advisor:** Dr. Hridesh Rajan, Coursework [✓], Research Proficiency Exam [✓], Ph.D. Candidacy Exam [✓], Ph.D. Final Defense [–]
- **Courses:** Design & Analysis of Algorithms · Machine Learning · Algorithms for Large Data Sets · Programming Languages · Advanced Topics in Programming Languages · Probabilistic Methods in Computer Engineering · Privacy Preserving Algorithms & Data Security · Distributed Development of Software · Formal Methods in Software Engineering · Theory of Computation · Principles & Internals of Database Systems · Advanced Topics of Modularity in Software Engineering

Jahangirnagar University (JU)   *Dhaka, Bangladesh*

B.Sc. in Information Technology   *January 2011 - January 2015*

- **Advisor:** Dr. Shamim Al Mamun, **Thesis:** Cloud Based Healthcare Application Architecture and Electronic Medical Record Mining [PDF]
- **Courses:** Object Oriented Programming · Mathematics I & II · Data Structures · Discrete Mathematics · Digital Logic Design · Algorithm Analysis · Human Computer Interaction · Signals & Systems · Data Communication & Computer Networks · Operating Systems · Computer Graphics · Artificial Intelligence & Neural Networks · Electronic Devices & Circuits · Embedded System Design · Bioinformatics · Wireless & Mobile Communication · Parallel & Distributed Systems · Statistics & Probability Theory · Object Oriented Programming · Network Security

Professional Experience

Iowa State University   *Ames, IA 50010*

Research Assistant, Laboratory of Software Design   *May 2018 - Present*

- PROJECTS - **ML Fairness:** Study software engineering aspects of ML fairness in real-world models e.g., bias identification, mitigation, impact analysis. Developed causal reasoning in ML pipeline to propose fair data preprocessing. · **DNN Verification:** Verify neural networks for safety properties under certain assumptions and postconditions · Study **Boa:** Worked on *Boa* compiler and infrastructure, especially built Python language support for Boa. I leveraged the framework to conduct large-scale program analysis on ML software from GitHub. · **Data Science Pipeline:** Survey, conduct empirical study, analyze data science programs and Jupyter notebooks, and study evolution of DNN (deep neural network) models written in Python · **Dependable Data Driven Discovery (D^4):** Assisted to write proposal for NSF-TRIPODS grant, work on dependability of data science software and safety critical applications.

Iowa State University   *Ames, IA 50010*

Teaching Assistant, Department of Computer Science   *August 2016 - May 2018*

- **COMS 309 - Software Development Practices:** Weekly lecture · prepare screencasts & recitation on design pattern, version-control, client-server architecture · supervise several 4-member teams developing software projects throughout the semester (technologies include Android SDK, Spring Boot, .Net, Angular JS, MySQL, Java Socket, Client-Server Architecture, MVC, etc.)
- **COMS 327 - Advanced Programming Techniques in C & C++:** Assist students with programming projects & assignments · Prepare and grade tests · Technologies include memory management, concurrent programming, network programming, etc.

Bangladesh University of Business and Technology   *Dhaka, Bangladesh*

Lecturer, Department of Computer Science & Engineering   *January 2016 - July 2016*

- **Responsibilities:** Instruct courses, create tests, conduct programming lab sessions
- **Course instructed:** Computer and Programming Concepts · Structured Programming Language · Pattern Recognition

DataSoft Systems Bangladesh Limited   *Dhaka, Bangladesh*

Software Engineering Intern   *August 2013 - December 2013*

- Develop Android Apps · Analyze mobile app market · Get training on coding conventions and SE practices i.e., agile, scrum, pair programming, etc. · Technologies include Java EE, Android SDK, Web APIs, etc.

Technical Skill

Programming Language	C, C++, Java, Python, C#
Database	MySQL, Microsoft SQL Server, Neo4J (Graph DB), SQLite, Oracle, Hibernate, JDBC
Web Techologies	HTML, CSS, PHP, JavaScript, Angular JS, JSP, Socket
Frameworks, IDEs & Web	Android SDK, AndEngine, .Net, Eclipse, Visual Studio, Netbeans, Matlab
Machine Learning & Deep Learning	Scikit-Learn, TensorFlow, Keras, PyTorch
Version Control & Project Management	Git, Trello, AgileFant, Continuous Integration & Development (CI/CD)
Other Software Skills	Shell, Latex, Hadoop, Spark, AWS, Arduino, Adobe Premiere, CISCO Packet Tracer

Publications

INTERNATIONAL CONFERENCE PAPERS *Google Scholar*

- **Sumon Biswas**, Hridesh Rajan, "Fair Preprocessing: Towards Understanding Compositional Fairness of Data Transformers in Machine Learning Pipeline" ESEC/FSE'2021: *29th ACM Joint European Software Engineering Conference and Symposium on the Foundations of Software Engineering*, Athens, Greece. [ACM Digital Library] [Artifact] [Presentation]
- **Sumon Biswas**, Hridesh Rajan, "Do the Machine Learning Models on a Crowd Sourced Platform Exhibit Bias? An Empirical Study on Model Fairness" ESEC/FSE'2020: *28th ACM Joint European Software Engineering Conference and Symposium on the Foundations of Software Engineering*, Sacramento, California. [ACM Digital Library] [Artifact] [Presentation]
- **Sumon Biswas**, Md Johirul Islam, Yijia Huang, Hridesh Rajan, "Boa Meets Python: A Boa Dataset of Data Science Software in Python Language" *Proceedings of the 16th International Conference on Mining Software Repositories* (MSR 2019). Montreal, Canada. [ACM Digital Library] [Slides]
- **Sumon Biswas**, Mohammad Wardat, Yijia Huang, Hridesh Rajan, "The Art and Practice of Data Science Pipelines" [Under Review]
- **Sumon Biswas**, Anisuzzaman, Tanjina Akhter, M Shamim Kaiser, Shamim Al Mamun, "Cloud Based Healthcare Application Architecture and Electronic Medical Record Mining: An Integrated Approach to Improve Healthcare System" *17th International Conference on Computer and Information Technology* (ICCIT 2014), Dhaka, Bangladesh. [IEEE XPlore]
- **Sumon Biswas**, M Shamim Kaiser, Shamim Al Mamun, "Applying Ant Colony Optimization in Software Testing to Generate Prioritized Optimal Path and Test Data" *2nd International Conference on Electrical Engineering and Information Communication Technology* (ICEEICT 2015), Dhaka, Bangladesh. [IEEE XPlore]
- Manan B T Noor, **Sumon Biswas**, "A Secure Data Security Infrastructure for Small Organization in Cloud Computing" *2nd International Conference on Electrical Engineering and Information Communication Technology* (ICEEICT 2015), Dhaka, Bangladesh. [IEEE XPlore]

Honors & Awards

INTERNATIONAL

2021	**Publication Award,** Awarded by ISU Computer Science for publication in the top-tier venues	*Ames, IA*
2021	**Journal Reviewer @ TSE,** Serving as the reviewer for IEEE Transactions on Software Engineering (TSE)	*Los Alamitos, CA*
2021	**Organizing Committee @ SPLASH'21,** Served as the Accessibility Chair of the ACM SIGPLAN conference on Systems, Programming, Languages, and Applications (SPLASH 2021) with OOPSLA, ECOOP and REBASE	*Chicago, IL*
2021	**Member of Artifact Evaluation Committee @ OOPSLA'21,** Served as the reviewer of Artifact Track in ACM SIGPLAN conference on Object-Oriented Programming, Systems, Languages & Applications (OOPSLA).	*Chicago, IL*
2021	**Web Chair @ Big Data Summer School,** Served as the Webmaster of the Midwest Big Data Summer School, May 17-20, 2021, organized by Iowa State University	*Chicago, IL*
2021	**Shadow PC Member @ MSR'21,** Serving as the shadow program committee (PC) member of the International Conference on Mining Software Repositories (MSR 2021) co-located with ICSE 2021	*Virtual*
2020	**Organizing Committee @ SPLASH'20,** Served as the Accessibility Chair of the ACM SIGPLAN conference on Systems, Programming, Languages, and Applications (SPLASH 2020) with OOPSLA, ECOOP and REBASE	*Chicago, IL*
2020	**Panelist @ ESEC/FSE'20,** Selected as a panelist in the session on Fairness at ESEC/FSE 2020	*Sacramento, CA*
2019	**PLMW Scholar @ PLDI'19,** Awarded by Programming Language Mentoring Workshop (PLMW) at PLDI'19	*Phoenix, Arizona*
2019	**Professional Advancement Grants @ ISU,** Awarded by Graduate and Professional Student Senate at Iowa State University	*Ames, IA*
2019	**ACM Travel Grant,** Awarded to attend IMA Workshop organized by Institute for Mathematics & its Applications (IMA), University of Minnesota	*Minneapolis, MN*

DOMESTIC

2015	**NST Fellowship,** Awarded by Ministry of Science & Technology for Research Excellence	*Dhaka, Bangladesh*
2014	**University Merit Scholarship,** Top 5% in Undergraduate Level, Jahangirnagar University	*Dhaka, Bangladesh*
2014	**Finalist,** National Hackathon organized by ICT Ministry, Bangladesh	*Dhaka, Bangladesh*
2013	**Finalist,** Google Developer Group (GDG) DevFest organized by GDG Bangladesh	*Dhaka, Bangladesh*

Presentations & Activities

2020	**ESEC/FSE'21 Paper Presentation**, Paper presentation in the virtual conference of ESEC/FSE (29th ACM Joint European Software Engineering Conference and Symposium on the Foundations of Software Engineering)	*Athens, Greece*
2020	**ESEC/FSE'20 Paper Presentation**, Paper presentation in the virtual conference of ESEC/FSE (28th ACM Joint European Software Engineering Conference and Symposium on the Foundations of Software Engineering)	*Sacramento, CA*
2020	**TADS Presentation**, Presented my recent research on "Fairness of Machine Learning Models" in front of Theoretical and Applied Data Science (TADS) Group of D4 Institute, Iowa State University	*Ames, IA*
2019-20	**PLDI'19 and PLDI'20**, Attended Programming Language Design and Implementation Conference PLDI 2019 at Phoenix, Arizona and PLDI 2020 (virtual)	*Phoenix, Arizona*
2019-20	**ICSE'19-21**, Attended International Conference on Software Engineering ICSE 2019 at Montreal, Canada and ICSE 2020-21 (virtual)	*Montreal, Canada*
2019	**Summer School**, Completed Professional Development Program in Midwest Big Data Summer School	*Ames, Iowa*
2019	**MSR'19**, Data showcase and paper presentation in Mining Software Repository Conference (MSR), 2019	*Montreal, Canada*
2019	**IMA Workshop**, Attended 4-day workshop on "Recent Themes in Resource Tradeoffs: Privacy, Fairness and Robustness" organized by Institute for Mathematics & its Applications (IMA), University of Minnesota	*Minneapolis, MN*
2019	**Turing Lecture**, Attended Turing Lecture at FCRC 2019 by Geoffrey Hinton and Yann LeCun on "The Deep Learning Revolution"	*Phoenix, AZ*

Academic & Skill Based Software Projects

Supervised Undergraduate Software Projects as TA
Java, Spring, Android, Socket, .NET

Semester-long projects | 3 Semesters | 20 teams | 4-member teams
ISU, Aug. 2017 - Dec. 2018

- *Selected Projects:* NextGen 911, Game X, CyChat, Battle of the Worlds, CyDisk, PrePark, Songusoid, Smart Art, CyBike, Buisiness QR, Twenty One

Uncertainty in DNN Hyperparameter Optimization
Python

Course Project for *Advanced Topics in Programming Languages* Course | 3-Member Team | GitHub
Aug. 2019 - Dec. 2019

- In this project, we provided aid to the deep learning programmers to quantify uncertainty in the model hyperparameters and help them to make informed decision while initializing hyperparameters. We have leveraged a first order type Uncertain<*T*> to represent the uncertainty in the random hyperparameters and choose the best value by performing statistical tests on the distribution.

New Semantics using Lambda Calculus
Coq

Course Project for *Programming Languages* Course | GitHub
Mar. 2019 - May 2019

- I have extended Lambda calculus using Coq proof assistant to implement a core language that includes two Python language features i.e., compound comparison statement and generator function. I have also described the syntax, operational semantics and type system of the features.

Near Duplicate Detection Using Simhash
Java

Course Project for *Probabilistic Methods in Computer Engineering* Course | GitHub
Aug. 2017 - Dec. 2018

- Given millions of documents, for each file, rank the similar documents. I have preprocessed the documents, created weighted word vectors and then implemented Simhash (locality sensitive hasing algorithm to evaluate approximate cosine similarity) to generate 64-bit fingerprint of each document. Finally block permuted hamming search was implemented in the fingerprint space to find the near duplicate.

Naive Bayes Classifier for Text Documents
Java

Semester Project for *Machine Learning* Course | GitHub
Aug. 2017 - Dec. 2018

- Classified text documents from a large pool using machine learning techniques. A Naive Bayes classifier was built to classify approximately 20,000 newsgroup documents. The original dataset is available here.

Performance Benchmarking for Link Prediction Algorithms in Social Networks
Java, MySQL, Neo4J

Research Project for *Database Systems* Course | 3-Member Team | GitHub
Aug. 2017 - Dec. 2018

- From a given snapshot of a social network database, we can predict whether a person can be potentially connected to another person, by analyzing existing links. We take two datasets (Facebook dataset from Stanford Large Network Dataset Collection and bibliography dataset from DBLP) and import that into MySQL, and Neo4J (Graph based DB) to evaluate the metrics for different network topology.

Public Key Encryption (PEKS) with Bloom Filter
Python

Academic Research Project | GitHub
Aug. 2017 - Dec. 2018

- PEKS with Keyword Search is one of the most used method to search keywords over encrypted data. However, semantic security is not preserved and dictionary attack can help attackers to guess keywords. This project solves that issue implementing PEKS and searching keywords over Bloom Filter.

Path ORAM (Oblivious Random Access Memory)
Python

Semester Project for *Privacy Preserving Algorithms and Data Security* Course | GitHub
ISU, Jan. 2017 - Feb. 2017

- While using cloud platform or any other insecure memory, attack can be made using the access pattern. Oblivious RAM is the way to hide the memory access pattern with some extra bandwidth and memory overhead. Path ORAM uses a binary tree to store all memory blocks and changes the location of the blocks repeatedly to access the whole branch for a single block.

Tank Battle (Android Shooting Game)
Java, Android SDK, AndEngine

Semester Project for *Game Development* Course | GitHub
JU, Sep. 2014 - Dec. 2014

- *Tank Battle* is an Android game where you have to shoot at the enemy tanks at a high speed from different angles to destroy them. Enemy tanks comes from the opposite direction and try to hit the gamer's tank. If any enemy tank passes you are defeated and the game is over.

Doctors Window (A solution for Doctors and Patients)
.Net, C#, SQL Server, Android SDK

Undergraduate Project Defense | 2-Member Team | GitHub
JU, Aug. 2014 - Jan. 2015

- This is a complete solution for private practitioner doctors. The doctors can manage patients appointment automatically, generate electronic prescriptions and store patients database. The patient side app helps to schedule appointment, get notification and view medical history.

Societies & Organizations

2019-21	**Student Member**, I am a student member of Association for Computing Machinery (ACM).	*New York, NY*
2020-21	**SIGSOFT & SIGPLAN**, I am a member of the ACM SIGSOFT and ACM SIGPLAN.	*Ames, IA*
2015-16	**Vice Chair**, I served as a vice chair of IEEE Student Branch, Jahangirnagar University	*Dhaka, Bangladesh*