

Subhodip Biswas

AI Verification Engineer, Zoox | Ph.D. in Computer Science, Virginia Tech

Contact— Phone: +1 314 913 0027 | E-mail: subhodip@vt.edu

On **0-1 VISA** | [linkedin.com/in/subhodip-biswas](https://www.linkedin.com/in/subhodip-biswas) | **GOOGLE SCHOLAR**

Experienced professional with deep knowledge of applying Machine Learning, derivative-free Optimization, and spatiotemporal Data Mining in providing data-driven solutions to business problems. Actively seeking **Applied/AI/ML Engineer/ Scientist, ML Researcher** and **Data/Research Scientist** opportunities.

EDUCATION

Ph.D. in Computer Science & Applications 05/2022
Virginia Tech

M.S. in Computer Science & Applications 05/2017
Virginia Tech

Bachelor of Electronics & Telecommunication Engg. 05/2014
Jadavpur University

ACCOLADES/ HONORS

- **Journeyman Fellowship** by US Army Research Lab–Oak Ridge Associated Universities, 2021.
- **Best-paper award** at ACM SIGSPATIAL 2020.
- **Student Travel Award:** ACM SIGSPATIAL 2019, AAAI/EAAI 2020.
- **CS Travel award:** ACM SIGSPATIAL 2019, 2020; INFORMS 2020 Annual Meeting; AAAI 2020.

KEY SKILLS

Applied Machine Learning, Deep Learning, Numerical Optimization, Spatiotemporal Data Mining, NLP

Programming Python, Matlab, R, SQL, Java, C/C++

Libraries BoTorch, GeoPandas, Matplotlib, NetworkX, NumPy, Pandas, PyTorch, Scikit-learn, Shapely

& Tools Gurobipy, OpenAI gym, Spacy, NLTK

Big Data Spark, DataBricks, Elasticsearch, SQL

EXPERIENCE

AI Verification Engineer 5/2022–present
System Design & Mission Assurance (SDMA), Zoox *Foster City, CA, USA*

Member of **Zoox Safety Analysis Framework** team focused on developing Bayesian Optimization, statistical sampling, and semi-supervised learning techniques for AI Verification.

- Created tooling for tracking the probable collision region. This helped in identifying redundancies in the pipeline. Implemented early-termination criterion of the optimization software of the safety analysis stack that improved experiment run-time by **10%**.
- Developed a data-mining strategy for identifying redundant simulation trials and implemented an adaptive pruning strategy to remove them. This reduced the annual simulation budget by **\$2.5M**.

Graduate Research Assistant 8/2021–5/2022
Saghani Center for Artificial Intelligence and Data Analytics, Virginia Tech *Arlington, VA, USA*

- **Machine Learning to Improve Forest Risk Commodity Traceability:** Joint work with World Forest ID aimed at identifying sources of illegal timber. Leveraged a range of geospatial datasets to enhance stable isotope ratio analysis (SIRA) to improve commodity traceability. Set up an ML pipeline to integrate diverse sources of data and pose/answer important questions, e.g., forecasting in unsampled areas. Developed pre-processing, feature selection, model building, and active sampling strategies to support SIRA model selection and evaluation. The work is currently under review in the *Methods in Ecology and Evolution* journal.
- **Ethical school redistricting:** Developed Markov Chain Monte Carlo-based techniques for quantifying how the present practices of designing public school boundaries ignore ethical considerations like students displaced, average commute distance, socioeconomic parity and so on. Proposed alternative metrics for including these factors in the decision-making process in an automated manner. The findings resulted in a publication at AAAI 2023.

Journeyman Fellow¹

Computational and Information Sciences, CCDC Army Research Laboratory

1/2021–8/2021

Adelphi, MD, USA

- Developed Bayesian optimization methods for problems with blackbox objectives and constraints. Applications include uncertainty-aware optimization, AutoML, general algorithmic configuration, etc. The work culminated in an oral paper accepted at IEEE CEC 2021 and ranked 2nd in the single-objective bound-constrained optimization.

Graduate Research Assistant

Discovery Analytics Center, Virginia Tech

5/2017–1/2021

Arlington, VA, USA

Member of core data science team specializing in providing data-driven solutions to urban analytics problems.

- **Spatial optimization techniques for redistricting:** Doctoral dissertation research focused on developing spatial partitioning techniques that work on arbitrary-shaped polygonal geometries with design constraints supplied by domain experts. These methods were used to design the public school boundaries of two counties in Northern Virginia with a 15 – 20% improvement in utilization and proximity. The research resulted in multiple peer-reviewed publications including the best paper award at ACM SIGSPATIAL 2020, the top conference in spatial computing.
- **Hyperparameter optimization in Machine Learning:** Devised a hybrid search algorithm called STEADE, which uses a mixture of surrogate models—Radial Basis Function and Gaussian Processes— for exploring the functional landscape, and then transitions to an Evolutionary Algorithm (EA) for exploitation. STEADE was applied to tuning hyperparameters of ML models and demonstrated how switching from a surrogate model to an EA is an effective search strategy. This work was accepted to the NeurIPS 2020 Blackbox optimization challenge.
- **Deep diffusion-based forecasting of COVID-19:** Developed a deep learning-based time-series model for probabilistic forecasting called Auto-regressive Mixed Density Dynamic Diffusion Network (ARM3Dnet), which considers both people's mobility and disease spread as a diffusion process on a dynamic directed graph. We show that our model, when trained with the best combination of dynamic covariate features and mixture components, can outperform both traditional statistical and deep learning models in forecasting the number of Covid-19 deaths and cases at the county level in the United States. Results were published at ACM/IEEE ASONAM 2021.
- **End-to-end Research Entity and Relationship Extractor:** Developed an information extraction system called E2R that uses weakly-supervised learning to extract technical entities such as source code, dataset, task, method, hardware resources, computing platforms, compute time, programming language, and libraries from full-text scholarly research articles. Besides being lightweight, E2R is scalable to long documents and can distinguish between a salient and non-salient entity in scholarly literature. The work was accepted as a paper at the Workshop on Data-Driven Science of Science, ACM KDD 2022.

Graduate Teaching Assistant

Department of Computer Science, Virginia Tech

8/2014–5/2017

Blacksburg, VA, USA

Performed instructional activities for graduate-level courses: *Data Analytics I*, *Machine Learning*, *Machine Learning with Big Data*, *Ethics and Professionalism in Data Science*.

Visiting Student

Indian Statistical Institute

7/2012–7/2014

Kolkata, WB, India

- **Evolutionary Computation:** Multi-year research work on developing improved variants of Evolutionary Algorithms and Swarm Intelligent algorithms for solving different types of optimization problems, including single-objective, multi-objective, multi-modal, dynamic, constrained, expensive, and so on. These techniques were also applied in solving real-world applications like sleep scheduling in sensors, optimal filter design, etc. The research findings got published in top-tier venues, including *IEEE Transactions on Cybernetics* (Impact Factor: **19.118**) and *IEEE Transactions on Evolutionary Computation* (Impact Factor: **16.497**).

SELECTED PUBLICATIONS

Statistics from [Google Scholar](#)—Citations: **1100+**, h-index: **17**, i10-index: **24**

1. JM Truszkowski, R Maor, RB Yousuf, S Biswas, C Chater, P Gasson, S McQueen, M Norman, J Saunders, J Simeone, N Ramakrishnan, “[A machine learning approach to estimating the geographical origin of timber](#)”, *Methods in Ecology and Evolution*, [In review]: Feb, 2023.
2. S Biswas, F Chen, Z Chen, C-T Lu, and N Ramakrishnan, “[Memetic algorithms for Spatial Partitioning problems](#)”, *ACM Transactions on Spatial Systems and Algorithms*, 9 (1): 1-31, Mar. 2023.
3. F Chen, S Biswas, K Fu, T Ji, C-T Lu, N Ramakrishnan, and Z Chen, “[Exploring Tradeoffs in Automated School Redistricting: Computational and Ethical Perspectives](#)”. In *Proceedings of the 37th AAAI Conference on Artificial Intelligence*, 2023: [Accepted].

¹The fellowship was jointly sponsored by the US Army Research Lab–Oak Ridge Associated Universities.

* Joint first authors. Equal contribution

4. S Biswas,* RB Yousuf,* KK Kaushal, J Dunham, Gelles, S Muthiah, N Self, P Butler, and N Ramakrishnan, “[Lessons from deep learning applied to scholarly information extraction: What works, what doesn’t, and future directions](#)”. In *Proceedings of the 28th ACM SIGKDD international conference on Knowledge Discovery and Data mining (KDD)*, 2022: Workshop on Data-Driven Science Of Science.
5. P Roy,* S Sarkar,* S Biswas,* F Chen, Z Chen, N Ramakrishnan, and C-T Lu, “[Deep diffusion-based forecasting of COVID-19 by incorporating network-level mobility information](#)”. In *Proceedings of the 2021 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*: 168–175.
6. S Biswas, D Saha, S De, AD Cobb, S Das and BA Jalaian, “[Improving Differential Evolution through Bayesian Hyperparameter Optimization](#)”. In *2021 IEEE Congress on Evolutionary Computation (CEC)*: 832-840.
7. S Biswas, AD Cobb, A Sistrunk, N Ramakrishnan and BA Jalaian, “[Better call Surrogates: A hybrid Evolutionary Algorithm for Hyperparameter optimization](#)”. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2020: Blackbox Optimization challenge.
8. S Biswas, F Chen, Z Chen, C-T Lu and N Ramakrishnan, “[Incorporating domain knowledge into Memetic Algorithms for solving Spatial Optimization problems](#)”. In *Proceedings of the 28th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, 2020: 25-35. [**Best paper award**]
9. S Biswas, S Kundu and S Das, “[Inducing niching behavior in Differential Evolution through local information sharing](#)”, *IEEE Transactions on Evolutionary Computation*, 19 (2): 246-263, Apr. 2015. [*Impact Factor: 16.497*]
10. S Das, S Biswas, BK Panigrahi, S Kundu and D Basu, “[An Improved Parent-Centric Mutation With Normalized Neighborhoods for Inducing Niching Behavior in Differential Evolution](#)”, *IEEE Transactions on Cybernetics*, 44 (10): 1726-1737, Oct. 2014. [*Impact Factor: 19.118*]

MEDIA COVERAGE

- Software platform engages communities in school rezoning decisions. [[VTNews](#)]
- Virginia Tech data program boosts local agencies short on analytical resources. [[statescoop](#)]
- UrbComp program team receives Alumni Award for Outreach Excellence. [[Link](#)]
- Subhodip Biswas receives Journeyman Fellowship from Army Research Lab. [[Link](#)]

REFERENCES

- **Naren Ramakrishnan**
Thomas L. Phillips Professor of Engineering
Department of Computer Science, Virginia Tech
Director, [Sanghani Center for AI and Data Analytics](#)
Director, [Amazon–VT Initiative in Efficient & Robust ML](#)
Contact: naren@cs.vt.edu
 - **Chang-Tien Lu**
Professor of Computer Science, Virginia Tech
[ACM Distinguished Scientist](#)
[Faculty Fellow](#), College of Engineering
CS Program Director, National Capital Region
E-mail: ctl@vt.edu
 - **Swagatam Das**
Associate Professor, Indian Statistical Institute–Kolkata, India
Head, Electronics and Communication Sciences Unit
Director, [Machine Learning Research Group](#)
E-mail: swagatam.das@isical.ac.in
-