Subhodip Biswas

AI/ML Researcher + Practitioner on O-1 Visa | M.S. and Ph.D. in Computer Science, Virginia Tech
linkedin.com/in/subhodip-biswas | Contact: subhodip@vt.edu | **RESEARCH** | **GITHUB**

Experienced professional with proven expertise in applied Machine Learning, derivative-free Optimization, and spatiotemporal Data Mining for solving business/research problems in Computer Vision, Natural Language Processing, Time-series Forecasting and Urban Analytics. Actively seeking opportunities—Applied AI/ML Engineer; ML Researcher; Applied/Data/Research Scientist.

ACCOLADES/ HONORS

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

KEY SKILLS

Programming	Python, Matlab, R, SQL, Java, C++, C
Libraries	PyTorch, Tensorflow, OpenCV, Transformers, Spacy, NLTK, BoTorch, NumPy, Pandas, Scikit-learn
& Tools	Gurobipy, GeoPandas,
Application	AWS, Azure Cloud, Docker, Django, Flask, Nginx
Big Data	Spark, DataBricks, Elasticsearch

EXPERIENCE

Artificial Intelligence Engineer   5/2022−present

System Design & Mission Assurance (SDMA), *Zoox*   *Foster City, CA, USA*

Member of **Z**oox **S**afety **A**nalysis **F**ramework team focused on evaluating safety of the AI software by using mix of Bayesian Optimization, statistical sampling, and semi-supervised learning techniques.

- **Adaptive Experiment prioritization:** Developed a data-mining strategy for identifying redundant simulation trials and pruning them based on an adaptive pruning strategy. This reduced the annual simulation budget by **$2.5M**.
- **Few-shot semantic segmentation:** Implemented techniques for segmenting a new object class using only a few annotated examples by transferring meta-knowledge from domains with sufficient training labels to low-resource domains. This helped in labeling obstacles encountered under low probability settings like rain, hail, fog, and so on.

ML Scientist   8/2021−5/2022

Sanghani Center for Artificial Intelligence and Data Analytics, Virginia Tech   *Arlington, VA, USA*

- **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.
- **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 preliminary analysis garnered $230K in research funding for subsequent analysis.

Journeyman Fellow   1/2021−8/2021

Computational and Information Sciences, CCDC Army Research Laboratory   *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.

Data Scientist   5/2017−1/2021

Discovery Analytics Center, Virginia Tech   *Arlington, VA, USA*

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

- **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.
- **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.

- **Traffic forecasting using Deep Learning:** Developed a Deep Kalman Filtering Network (DKFN) model to forecast the network-wide traffic state by modeling the node's self- and neighbor-dependency as two streams, and their predictions are fused using Kalman filters. Experimental results showed that DKFN performs better than existing models (that simply combined LSTMs with Graph Neural Networks) on speed prediction task. The effort culminated into a conference publication at ACM SIGSPATIAL 2020.

Graduate Teaching Assistant · 8/2014−5/2017
Department of Computer Science, Virginia Tech · *Blacksburg, VA, USA*

- Performed instructional activities for graduate-level courses offered by the CS department − Data Analytics I, Machine Learning, Machine Learning with Big Data, Ethics and Professionalism in Data Science.

Visiting Student · 7/2012−7/2014
Indian Statistical Institute · *Kolkata, WB, India*

- Developed 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**).

Summer Intern · 6/2013−8/2013
National University of Singapore · *Singapore*

- **Stochastic optimization in power systems**: Proposed a hybrid optimization model to solve the unit commitment problem in power systems by combining the Genetic Algorithm with the Differential Evolution algorithm. The framework achieved state-of-the-art performance and was later extended to a multi-objective setting to obtain a trade-off between operating costs and emissions. The research resulted in journal publications at *Information Sciences* (Impact Factors: **8.233**) and *Swarm & Evolutionary Computation* (Impact Factors: **10.267**).

EDUCATION

Ph.D. in Computer Science · 05/2022
Virginia Tech
Thesis: Spatial Optimization Techniques for School Redistricting

M.S. in Computer Science · 05/2017
Virginia Tech

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

PUBLICATIONS

Google Scholar−**1200+** citations, h-index: **17**, i10-index: **24**

1. 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. [Special Issue: **Best papers of ACM SIGSPATIAL 2020**]
2. F Chen, S Biswas, Z Chen, S Li, N Ramakrishnan and C-T Lu, "Exploring Tradeoffs in Automated School Redistricting: Computational and Ethical Perspectives". In *Proceedings of the 37th AAAI Conference on Artificial Intelligence,* 2023: [Accepted].
3. 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.
4. 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.
5. 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.
6. 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**]
7. S Biswas, F Chen, A Sistrunk, S Muthiah, Z Chen, N Self, C-T Lu and N Ramakrishnan, "Geospatial clustering for balanced and proximal schools". In *Proceedings of the 34th AAAI Conference on Artificial Intelligence,* 2020, 34 (09): 13358-13365.

---

* Joint first authors/ Equal contribution