SOUMYA BANERJEE

CONTACT INFORMATION
Email: neel.soumya@gmail.com
Website: https://sites.google.com/site/neelsoumya/

EDUCATION
PhD in Computer Science, 2013, University of New Mexico, USA
Bachelor of Engineering (Computer Science) with Distinction, 2003, Nagpur University, India

RECENT EMPLOYMENT
Senior Research Fellow, University of Cambridge, UK, 2022 till date
Project description: I develop explainable AI techniques and apply them to various domains like healthcare.

Postdoctoral Researcher, University of Cambridge, UK, 2019 till 2022
Project description: I use machine learning techniques and data science on electronic healthcare record data to help stratify patients and predict mortality. This has been published in *Nature Partner Journal Schizophrenia*.

Postdoctoral Researcher, University of Oxford, UK, 2016 till 2018
Project description: The project involves modeling negative selection in the immune system. I use multi-scale models to simulate how immune system cells are trained in the thymus. The work resulted in a manuscript published in the *Journal of the Royal Society Interface*.

Researcher, Commonwealth Scientific and Industrial Research Organisation, Australia, 2015 to 2016
Project description: The project involved modeling socio-economic systems by combining open data with machine learning techniques. We used a publicly available dataset from the UCI Machine Learning Repository on crime in cities. Using statistical and machine learning models combined with non-linear dynamical models we inferred novel patterns in crime in cities. This resulted in a manuscript published in the journal *Palgrave Communications*.

Post-doctoral research fellow, Harvard University, Harvard Medical School and Broad Institute of Harvard and MIT, USA, 2014 to 2015
Project description: The work involves investigating the interaction between the immune system and the human microbiome. The human microbiome has been implicated in various disorders like inflammatory bowel disease. I am using computational techniques to predict compounds that are likely to be secreted by microbes in the human gut based on their genome. Compounds that are computationally predicted to be differentially abundant in healthy vs. disease subjects across multiple cohorts are likely to play an important role in inflammatory bowel disease. This project is implemented on python, shell scripts and MATLAB.

Post-doctoral research fellow, Max Planck Institute for Molecular Physiology, Germany, 2013 to 2014
Project description: The Rho GTPase family of proteins influence cell motility and cytoskeleton dynamics across all eukaryotes. I investigated the interaction between three different Rho GTPase proteins and how this shapes cell motility. I used non-linear differential equation models to simulate the interaction between the different Rho GTPase proteins. In order to make more accurate estimates of model parameters despite noisy experimental data, I used hierarchical Bayesian models to combine data from different experiments. All code was implemented in MATLAB and shell

scripts and run on high-performance computing platforms. I also implemented automated image analysis of cell images from microscopes using tools like CellProfiler and ImageJ.

I was also trained in various wet-lab skills like plating and splitting cells, staining with primary and secondary antibodies, transfection and microscopy.

The work resulted in a publication in the journal *Angewandte Chemie* (Impact Factor = 13.1) and the *Journal of Cell Biology* (Impact Factor = 9.7). One more manuscript is in preparation.

Research Intern, Los Alamos National Laboratories, Theoretical Biology and Biophysics Group, USA, 2009 to 2012

Project description: The project (PhD thesis) involved modeling emerging pathogens (like West Nile virus and avian influenza) that have sparse experimental data and considerable uncertainty in parameters that go into models. I simulated the complex interaction between the virus, host cells and the host immune system using non-linear differential equations. I built mathematical models that are able to predict biologically relevant quantities like the basic reproductive number and infectious virion burst size with a lot of precision, despite noisy data and uncertainty in inputs to our models. Our novelty lay in:

1) A combination of knockout experimental data and mathematical models, and

2) Hierarchical Bayesian models that incorporate hierarchy in data and biologically motivated priors to generate more precise estimates than otherwise possible from sparse experimental data.

The work resulted in a manuscript accepted in the *Journal of the Royal Society Interface*.

I also presented the work in three international venues. The MATLAB programs to solve these complex statistical problems was made accessible using easy to use graphical user interfaces and made openly available. Based on the number of times these tools have been downloaded, I have been ranked within the [Top 200 worldwide as a MATLAB open-source code contributor](#).

Research Assistant, University of New Mexico, Department of Computer Science, USA, 2007 to 2013

Project description: The project involved work in computational immunology: using techniques from computer science to solve problems in immunology and also taking inspiration from the immune system to solve problems in computer science. Specifically, my work suggested how the physical architecture of the immune system might lead to nearly scale-invariant immune search and response. Using the immune system as an inspiration I proposed architectures for human-engineered distributed systems (like peer-to-peer networks and mobile ad-hoc networks) with faster search and response characteristics.

The work resulted in 1 journal publication (*Swarm Intelligence*) and 4 peer-reviewed conference publications.

Senior Software Engineer and Technical Lead, Cognizant Technology Solutions Ltd., India, 2004 to 2007

Project description: I worked on projects in the financial services sector with various Fortune 500 clients. The work involved interacting with the client, defining project requirements, implementing solutions according to requirements, coordinating a team of developers, and rigorous testing. I worked on the UNIX platform and used C, C++, shell scripts and Sybase to implement solutions. I also worked on developing web-based applications with ASP.NET and database development with MS SQL Server.

Software Engineer, Automation Engineers, India, 2003 to 2004

PUBLICATIONS

A link to preprints of my most relevant papers is here:
https://sites.google.com/site/neelsoumya/publications/PAPERS.zip?attredirects=0&d=1

NOTE: In my field (Computer Science), people submit to conferences and conferences are peer-reviewed

Refereed Journal Papers
1) dsSurvival 2.0: Privacy enhancing survival curves for survival models in the federated DataSHIELD analysis system, S.Banerjee, T. Bishop, *BMC Research Notes* 16, 98, 2023
2) Patient and public involvement to build trust in artificial intelligence: a framework, tools and case studies, S.Banerjee, P. Alsop, L. Jones, R. Cardinal, *Patterns*, 3(6):100506, 2022
(Cell Press publishing group)
3) Software Application Profile: ShinyDataSHIELD—an R Shiny application to perform federated non-disclosive data analysis in multicohort studies, X. Escribà-Montagut, Y. Marcon, D. Avraam, S.Banerjee, T. Bishop, P. Burton, J. González, *International Journal of Epidemiology,* dyac201, 2022
(*Oxford University Publishing Group, Impact Factor 9.8*)
4) dsSynthetic: Synthetic data generation for the DataSHIELD federated analysis system, S.Banerjee, T. Bishop, *BMC Research Notes*, 15(1):230, 2022
5) dsSurvival: Privacy preserving survival models for federated meta-analysis of clinical data, S.Banerjee, et al, *BMC Research Notes*, 15(1):197, 2022
6) A class-contrastive human-interpretable machine learning approach to predict mortality in severe mental illness, S. Banerjee, P. Lio, P. Jones, R. Cardinal, *Nature Partner Journal Schizophrenia*, 7, 60, 2021
(*Nature publishing group, Impact Factor = 6.3*)
7) Optogenetic tuning reveals Rho amplification-dependent dynamics of a cell contraction signal network, D. Kamps, J. Koch, V. Juma, E. Campillo, M. Graessl, S. Banerjee, T. Mazel, X. Chen, Y. Wu, S. Portet, A. Madzvamuse, P. Nalbant, L. Dehmelt, *Cell Reports*, 33(9):108467, 2020
(*Cell Press publishing group, Impact Factor = 8.1*)
8) Deconvolution of monocyte responses in inflammatory bowel disease reveals an IL-1 cytokine network that regulates IL-23 in genetic and acquired IL-10 resistance, D. Aschenbrenner, M. Quaranta, S. Banerjee, et al., *Gut,* 2020
(*British Medical Journal publishing group, Impact Factor = 19.8*)
9) Hydroxychloroquine: balancing the needs of LMICs during the COVID-19 pandemic, S. Banerjee, *Lancet Rheumatology*, 2(7):385-386, 2020
(*Lancet publishing group*)
10) The early impact of COVID-19 on mental health and community physical health services and their patients' mortality in Cambridgeshire and Peterborough, UK, S. Chen, P. Jones, B. Underwood, A. Moore, E. Bullmore, S. Banerjee, et al., *Journal of Psychiatric Research*, 131, 244-254, 2020
(*Impact factor = 4.4*)
11) Predictive metabolomic profiling of microbial communities using amplicon or metagenomic sequences, H. Mallick, E. Franzosa, L. McIver, S. Banerjee, A. Sirota-Madi, A. Kostic, C. Clish, H. Vlamakis, R. Xavier, C. Huttenhower, *Nature Communications*, 10(1):3136, 2019
(*Nature publishing group, Impact Factor = 12.1*)
12) Influence of correlated antigen presentation on T cell negative selection in the thymus, S. Banerjee, S.J. Chapman, *Journal of the Royal Society Interface*, 15(148), 20180311, 2018
(*Royal Society publishing group, Impact Factor = 4.3*)

13) Modelling the effects of phylogeny and body size on within-host pathogen replication and immune response, S. Banerjee, A. Perelson, M. Moses, *Journal of the Royal Society Interface*, 14(136), 20170479, 2017
(*Royal Society publishing group, Impact Factor = 4.3*)
14) An excitable Rho GTPase signaling network generates dynamic subcellular contraction patterns, M. Graessl, J. Koch, A. Calderon, S. Banerjee, T. Mazel, N. Schulze, J. Jungkurth, A. Koseska, L. Dehmelt, P. Nalbant, *Journal of Cell Biology*, 216(12), 4271-4285, 2017
(*Cell publishing group, Impact Factor = 9.7*)
15) Towards a Quantitative Understanding of Within Host Dynamics of West Nile Virus Infection, S. Banerjee, J. Guedj, R. Ribeiro, M. Moses & A. Perelson, *Journal of the Royal Society Interface*, 13(117), 20160130, 2016
(*Royal Society publishing group, Impact Factor = 4.3*)
16) A spatial model of the efficiency of T cell search in the influenza infected lung, D. Levin, S. Forrest, S. Banerjee, M. Moses & F. Koster, *Journal of Theoretical Biology*, 398(7), 52-63, 2016
(*Impact Factor = 2.2*)
17) Competitive dynamics between criminals and law enforcement explains the super-linear scaling of crime in societies, S. Banerjee, Hentenryck, P.V. & Cebrian, M., *Palgrave Communications*, 1(1), 15022, 2015
(*Nature publishing group*)
18) A bioorthogonal small-molecule switch system for controlling protein function in cells, P. Liu, A. Calderon, G. Konstantinidis, J. Hou, S. Voss, X. Chen, F. Li, S. Banerjee, J. Hoffmann, C. Theiss, L. Dehmelt & Y. Wu, *Angewandte Chemie*, 53(38), 10049-10055, 2014
(*Impact Factor = 13.7*)
19) Science and technology consortia in US biomedical research: A paradigm shift in response to unsustainable academic growth, C. Balch, H. Arias-Pulido, S. Banerjee, A. Lancaster, K. Clark, M. Perilstein, B. Hawkins, J. Rhodes, P. Sliz, J. Wilkins and T. Chittenden, *BioEssays*, 37(2), 119-122, 2014
(*Impact Factor = 5.4*)
20) Scale Invariance of Immune System Response Rates and Times: Perspectives on Immune System Architecture and Implications for Artificial Immune Systems, S. Banerjee & M. Moses, *Swarm Intelligence*, 4(4), 301-308, 2010
(*Impact Factor = 2.1*)

Refereed Conference Papers
1) Analysis of a Planetary Scale Scientific Collaboration Dataset Reveals Novel Patterns, S. Banerjee, *Proceedings of the Complex Systems Digital Campus 2015 – World e-Conference, Conference on Complex Systems,* 2016 (peer-reviewed conference)
2) A Multi-Agent System Approach to Load-Balancing and Resource Allocation for Distributed Computing, S. Banerjee and J. Hecker, *Proceedings of the Complex Systems Digital Campus 2015 – World e-Conference, Conference on Complex Systems*, 2016 (peer-reviewed conference)
3) Computationally Simulating Intermodal Terminal Attractiveness and Demand, S. Banerjee et al., *Proceedings of the 23rd World Congress on Intelligent Transport Systems*, 2016
4) Analysis of Demand and Operations of Intermodal Terminals, R. Garcia-Flores, S. Banerjee et al., *Proceedings of the 24th National Conference of the Australian Society for Operations Research*, 2016
5) Forecasting in the era of Big Data: Lessons and Pitfalls, Y. Tyshetskiy, S. Banerjee et al., *Proceedings of the Annual Conference of the International Association of Maritime Economists*, 2016
6) The Value of Inflammatory Signals in Adaptive Immune Responses, S. Banerjee, D. Levin, M. Moses, F. Koster & S. Forrest, *The 10th International Conference on Artificial Immune Systems (ICARIS)*, Lecture Notes in Computer Science, Volume 6825/2011, 1-14, 2011
7) Biologically Inspired Design Principles for Scalable, Robust, Adaptive, Decentralized Search and

Automated Response (RADAR), M. Moses & <u>S. Banerjee</u>, *Proceedings of the 2011 IEEE Conference on Artificial Life*, 30-37, 2011

8) Modular RADAR: An Immune System Inspired Search and Response Strategy for Distributed Systems, <u>S. Banerjee</u> & M. Moses, *The 9th International Conference on Artificial Immune Systems (ICARIS)*, Lecture Notes in Computer Science, Volume 6209/2010, 116-129, 2010

9) A Hybrid Agent Based and Differential Equation Model of Body Size Effects on Pathogen Replication and Immune System Response, <u>S. Banerjee</u> & M. Moses, *The 8th International Conference on Artificial Immune Systems (ICARIS)*, Volume 5666-014, 2009

Book Chapters

Using Optimisation and Machine Learning to Validate the Value of Infrastructure Investments, R. Garcia-Flores, S. Banerjee, G. Mathews (2016) Book chapter in Infrastructure Investments: Politics, Barriers and Economic Consequences

HONORS AND AWARDS

1) 2010 Student Award for Innovation in Informatics (the award is given to a University of New Mexico graduate or undergraduate student for the best paper describing innovation or research in the field of biomedical informatics)

2) Distinction and 5th rank in Nagpur University (out of 2000 examinees) in Bachelor of Engineering program

GRANTS AWARDED

1) SCAP (Student Conference Award Program) grant from the University of New Mexico to attend ICARIS 2009 (July 2009): $ 600

2) Bursary from conference organizers to attend ICARIS 2009 (July 2009): $300

3) RPT (Research Project Travel) grant from the University of New Mexico to attend ICARIS conference 2009 (June 2009): $ 400

4) Travel grant from University of New Mexico, Department of Biology to attend ICARIS conference 2009 (May 2009): $ 500

5) RPT (Research Project Travel) grant from the University of New Mexico to attend Yale University Summer School in Computational Immunology (July 2008): $ 170

6) Travel grant from University of New Mexico, Department of Biology to attend Yale University Summer School in Computational Immunology (May 2008): $ 500

7) Travel grant from University of New Mexico, Department of Biology to present research work at Gordon Research Conference, Maine (Jan 2008): $ 500

INVITED TALKS

1) Computational Immunology, Microsoft Research, Cambridge, UK, March 2017

2) Modeling Emerging Pathogens under Uncertainty and Sparse Experimental Data, Harvard Medical School, Boston, USA, September 2014

3) Modeling Emerging Pathogens under Uncertainty and Sparse Experimental Data, IBM Research, India, August 2014

4) Computational Screens for Novel Gut Microbial Bioactive Compounds, Novartis Institutes for Biomedical Research, Boston, USA, July 2014

5) Scaling in the Immune System, Commonwealth Scientific and Industrial Research Organisation, Australia, March 2013

6) A Mathematical Model of Body Size Effects on Pathogen Replication and Immune System Response, International Network of Theoretical Immunology, Los Alamos National Laboratories, USA, August 2010

7) A Hybrid Agent Based and Differential Equation Model of Body Size Effects on Pathogen Replication and Immune System Response, School of Health Sciences, University of New Mexico, USA, April 2010

CONTRIBUTED TALKS

1) Modular RADAR: An Immune System Inspired Search and Response Strategy for Distributed Systems, 9th International Conference on Artificial Immune Systems (ICARIS), Edinburgh, UK, 2010

2) Immune System Inspired Strategies for Distributed Systems, 6th Annual Computer Science at UNM Student Conference (CSUSC) 2010

3) A Hybrid Agent Based and Differential Equation Model of Body Size Effects on Pathogen Replication and Immune System Response, 8th International Conference on Artificial Immune Systems (ICARIS), York, UK, 2009

4) A Hybrid Agent Based and Differential Equation Model of Body Size Effects on Pathogen Replication and Immune System Response, Western Region COBRE-INBRE Scientific Conference, Montana, September 2009

5) A Metabolic Theory of Immunology, Biology Research Day, University of New Mexico Dept. of Biology, April 2009

TEACHING EXPERIENCE

I am teaching machine learning at the University of Cambridge Department of Computer Science and Bioinformatics Training Centre. I have also taught students in Uganda through the University of Cambridge Africa programme. I taught data science at the University of Oxford Complex Networks Summer School.

I have taught Unconventional approaches to Artificial Intelligence. This is a special topics course I developed. The teaching material is available online here:

https://github.com/neelsoumya/special_topics_unconventional_AI

I have also developed materials for and taught a course on reproducible research in R:

https://github.com/neelsoumya/teaching_reproducible_science_R

I have taught machine learning, complex systems, and basics of programming. All my teaching material are available here:

https://sites.google.com/site/neelsoumya/teaching

Other teaching activities are outlined below:

1) Cambridge-Africa Programme, University of Cambridge, 2020 (lectures)

2) Oxford Summer School in Economic Networks, Mathematical Institute, University of Oxford, 2017 (lectures and tutorials)

3) Trained in teaching at the University of Oxford Doctoral Training Center, 2017

4) Lectured at the Complex Adaptive Systems course in Department of Computer Science, University of New Mexico, USA, 2009-2012 (lectures and tutorials on complex systems and computer science)

5) Co-designed teaching resources with educators to make my teaching broadly accessible (samples of my teaching material are available online https://www.simiode.org/resources/3206 and https://osf.io/25gnz/)

SUPERVISIONS

I have supervised 5 MPhil/MSc students.

ACADEMIC SERVICE

1) Served on recruitment panels for PhD and MPhil/MSc students
2) Program committee member of International Conference on Artificial Immune Systems
3) Review Editor for Frontiers in Virology
4) Reviewer for IEEE Transactions on Systems, Man, and Cybernetics and IEEE Transactions on Emerging Topics in Computational Intelligence
5) Reviewer for SIMIODE (Systemic Initiative for Modeling Investigations and Opportunities with Differential Equations)

COMPUTER LANGUAGES

1) Programming - Python, R, MATLAB, UNIX shell scripting, C, C++, Perl, ASP.NET, Haskell
2) Databases - MS SQL Server, Sybase
3) Image analysis and automated cell tracking tools – ImageJ, CellProfiler
4) Distinctions – Ranked within the Top 250 worldwide as an open-source code contributor in the MATLAB Central code repository
https://sites.google.com/site/neelsoumya/software
5) R packages I have written
https://github.com/neelsoumya/dsSurvival
https://github.com/neelsoumya/dsSurvivalClient

ATTENDED WORKSHOPS AND SHORT COURSES

1) Summer School on Data Science, Northwestern University, April 2020.
2) Workshop on Spatial Statistics, Harvard University, July 2014
3) Santa Fe Institute Complex Systems Summer School, June 2nd to June 27th, 2008
4) 3rd Annual Summer School in Computational Immunology, Yale University, August 11th to August 16th, 2008
5) Gordon Research Conference in Metabolic Basis of Ecology, Maine, July 2008
6) Scaling in Biological and Social Networks Workshop, Santa Fe Institute, July 2007